

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université A. MIRA de Bejaia  
Faculté des Sciences Exactes  
Département de Mathématiques  
Laboratoire de Mathématiques Appliquées



## MÉMOIRE

Présenté par :

Wissem Boubaya

Pour l'obtention du diplôme de

### MASTER

Filière : Mathématiques

Option : Probabilités Statistiques Appliquées

### Thème

## Sur quelques tests statistiques et applications

Soutenu le 13/07/2022, devant le jury composé de :

Présidente	Mme Timeridjine K.	Professeur	Univ. A.Mira de Bejaia
Rapporteur	M. Gharout H.	M.C.B	Univ. A.Mira de Bejaia
Co-Rapporteur	M. Bouraine M.	M.C.B	Univ. A.Mira de Bejaia
Examineur	M. Soufit M.	M.C.A	Univ. A.Mira de Bejaia

Année universitaire 2021/2022

## *\* Remerciements \**

*Au terme de ce modeste travail, je tiens à remercier, MM. GHAROUT H. et BOURAINÉ M. pour avoir accepté de m'encadrer, pour leurs précieux conseils et orientations les plus précieuses. Je garderais un très bon souvenir de leurs qualités profondément humaines.*

*J'adresse mes plus vifs remerciements à Mme TIMERIDJINE K. en acceptant de présider le jury et à M. SOUFIT M. pour avoir accepté d'examiner ce mémoire.*

*Mes sincères remerciements à tous les membres de la faculté des sciences exactes et en particulier, le département de mathématiques.*

*Je remercie, également, toute les personnes qui ont contribué de près ou de loin à la réalisation de ce mémoire.*

*W.BOUBAYA*

※ *Dédicaces* ※

Je dédie ce modeste travail :

**A ma très chère mère :**

Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi.

**A mon très cher père :**

Aucune dédicace ne pourrai exprimer l'amour, l'estime, le dévouement et le respect que j'ai toujours eu pour toi.

**A mon cher mari :**

Je t'adresse ces mots pour t'exprimer toute ma reconnaissance. Sans demander, tu as su m'aider au moment où j'en avais le plus besoin. Merci d'être tout simplement toi.

**A mes deux chers frères Khaled et Sami et mes chères soeurs Lamia, Noura et Basma :**

Je vous souhaite un avenir plein de joie, de réussite et de sérénité. Je vous exprime à travers ce travail mes sentiments de fraternité et d'amour.

**A ma meilleure amie Chahra :**

J'ai cherché longtemps comment je pouvais te remercier pour ta présence à ce moment de ma vie, pour tes témoignages, tes gestes et ton soutien.

**A toutes mes amies : Rania, Khadidja, Hasna, Ilhem B, Ilhem W, Sara.....**

Je vous n'oublierai jamais merci pour tout.

Wissem BOUBAYA

# Table des matières

<b>Table des matières</b>	<b>1</b>
<b>Table des figures</b>	<b>2</b>
<b>Introduction</b>	<b>4</b>
<b>1 Généralités sur les variables aléatoires et lois usuelles de probabilités</b>	<b>6</b>
1.1 Variable aléatoire . . . . .	6
1.2 Variable aléatoire discrète . . . . .	7
1.2.1 Loi de probabilité (distribution de probabilité) . . . . .	7
1.2.2 Fonction de répartition . . . . .	8
1.2.3 Espérance et variance d'une variable aléatoire discrète . . . . .	8
1.3 Variable aléatoire continue . . . . .	11
1.3.1 Fonction de répartition . . . . .	11
1.3.2 Densité de probabilité . . . . .	12
1.3.3 Espérance et variance d'une variable aléatoire continue . . . . .	13
1.4 Lois usuelles de probabilités . . . . .	13
1.4.1 Lois discrètes . . . . .	13
1.4.2 Lois usuelles continues . . . . .	19
1.4.3 Autres lois utiles pour les tests . . . . .	24
<b>2 Tests Paramétriques</b>	<b>30</b>
2.1 Généralités sur les tests statistiques . . . . .	30
2.1.1 L'hypothèse nulle et l'hypothèse alternative . . . . .	31
2.1.2 Les erreurs et les risques . . . . .	31
2.1.3 Variable de décision . . . . .	32
2.1.4 Région critique (région de rejet) . . . . .	33
2.1.5 Puissance d'un test . . . . .	33

---

2.1.6	Notions de p-valeur . . . . .	33
2.2	Principaux tests paramétriques usuels . . . . .	34
2.2.1	Comparaison des moyennes (Test de Student) . . . . .	34
2.2.2	Comparaison des variances (Test de Fisher) . . . . .	41
2.2.3	Tests sur les proportions . . . . .	45
<b>3</b>	<b>Tests non paramétriques</b>	<b>47</b>
3.1	Introduction . . . . .	47
3.2	Test non paramétrique . . . . .	47
3.3	Test de conformité à une loi donnée . . . . .	48
3.3.1	Test de Kolmogorov Smirnov . . . . .	48
3.3.2	Test de Cramer-Von Mises . . . . .	48
3.3.3	Test de Khi-deux (Test de Pearson) . . . . .	50
3.4	Test de comparaison pour deux échantillons . . . . .	51
3.4.1	Test de Wilcoxon . . . . .	51
3.4.2	Test de signes . . . . .	54
3.4.3	Test d'indépendance . . . . .	55
<b>4</b>	<b>Application des tests statistiques avec R</b>	<b>58</b>
4.1	Répartition des pathologies selon l'année . . . . .	58
4.2	Répartition des pathologies selon le sexe . . . . .	61
4.3	Poids du cerveau selon le sexe . . . . .	62
4.4	Efficacité du vaccin antigripale . . . . .	64
	<b>Conclusion Générale</b>	<b>65</b>
	<b>Bibliographie</b>	<b>66</b>

# Table des figures

1.1	Lois de probabilité pour la loi Binomiale. . . . .	15
1.2	La loi normale centré réduit. . . . .	21
1.3	La loi normale de moyenne $m$ et de variance $\sigma^2$ . . . . .	23
1.4	La loi de Khi deux pour $n$ degré de liberté. . . . .	25
1.5	La densité de la loi de Student. . . . .	27
1.6	La loi de Fisher. . . . .	28
2.1	L'erreur de première et deuxième espèce. . . . .	32
2.2	La relation entre les deux risques d'erreur. . . . .	33
2.3	Région d'acceptation et de rejet du test bilatéral de la moyenne ( $\sigma$ connue) au niveau de signification $\alpha$ . . . . .	36
2.4	Région d'acceptation et de rejet du test bilatéral de la moyenne ( $\sigma$ inconnue) au niveau de signification $\alpha$ . . . . .	38
2.5	Région d'acceptation et de rejet du test bilatéral de la variance ( $\mu$ connue) au niveau de signification $\alpha$ . . . . .	42
2.6	Région d'acceptation et de rejet du test bilatéral de la variance ( $\mu$ inconnue) au niveau de signification $\alpha$ . . . . .	42

# Introduction

L'inférence statistique est l'ensemble des techniques permettant d'introduire les caractéristiques d'un groupe général (la population) à partir de celles d'un groupe particulier (l'échantillon).

Les méthodes d'inférence statistiques ont connu deux grandes phases de développement. La première commence à la fin du XIXe siècle, avec les travaux de K. Pearson, R. Fisher, Jerzy Neyman, Egon Pearson et Abraham Wald qui dégagent les notions fondamentales de vraisemblance, de puissance des tests d'hypothèse et d'intervalle de confiance. La seconde période, qui perdure aujourd'hui, a été rendue possible grâce à la puissance de calcul des ordinateurs et à la banalisation de l'outil informatique à partir de la fin des années 1940 [19]. Ces calculateurs ont permis de dépasser les hypothèses traditionnelles d'indépendance et de normalité, commodes du point de vue mathématique mais souvent simplistes, pour donner toute leur fécondité à des concepts même anciens comme l'hypothèse bayésienne. L'informatique a permis aussi l'explosion des techniques de simulation par application des techniques de ré-échantillonnage : méthode de Monte Carlo, bootstrap, jackknife etc. imaginées par John von Neumann, Stanislas Ulam, Bradley Efron, Richard von Mises [2].

Dans tous les domaines, de l'expérimentation scientifique à la vie quotidienne, on est amené à prendre des décisions sur une activité risquée au vu de résultats d'expériences ou d'observation de phénomènes dans un contexte incertain, par exemple : Informatique : au vu des résultats des tests d'un nouveau système informatique, on doit décider si ce système est suffisamment fiable et performant pour être mis en vente. Finance : au vu du marché, décider si on doit ou pas se lancer dans une opération financière donnée. Dans chaque cas, le problème de décision consiste à trancher, au vu d'observations, entre une hypothèse appelée hypothèse nulle, et une autre hypothèse dite hypothèse alternative [19]. En général, une et une seule de ces deux hypothèses est vraie. La théorie des tests statistique, qui est un outil très puissant de l'inférence statistique, permet de répondre à ce type de problèmes [4].

De nombreux chercheurs ont étudié des tests d'hypothèses statistiques. Un exemple de test d'hypothèse statistique figure dans un mémoire de Pierre Simon de Laplace (1749-1827) publié en 1773 et réédité en 1891. Une étude heuristique des conditions sous lesquelles l'hypothèse  $H_0$  devrait être rejetée (c'est-à-dire, en langage moderne, une étude des régions critiques) est donnée dans le livre *Le Hasard* (1914) d'Émile Borel (1871-1956) . À cette époque les idées concernant les tests statistiques étaient encore floues. Borel insistait sur l'existence d'un test statistique " remarquable " dont les propriétés pourraient être démontrées rigoureusement. Il disait aussi que le choix du test devait être fait avant toute expérience [11].

Les tests paramétriques fonctionnent en supposant que les données que l'on a à disposition suivent un type de loi de distribution connu (en général la loi normale). Pour calculer le risque alpha du test statistique, il suffit de calculer la moyenne et l'écart-type de l'échantillon afin d'accéder à la loi de distribution de l'échantillon. La loi de distribution étant ainsi parfaitement connue, on peut calculer le risque alpha en se basant sur les calculs théoriques de la loi gaussienne. Ces tests sont en général très fins, mais ils nécessitent que les données suivent effectivement la loi de distribution supposée. Ils sont en particulier très sensibles aux valeurs aberrantes et ne sont pas conseillés si des valeurs aberrantes sont détectées [19].

Les tests non paramétriques ne font aucune hypothèse sur le type de loi de distribution des données. Ils se basent uniquement sur les informations numériques apportées par les échantillons. De plus, ils sont très peu sensibles aux valeurs aberrantes et sont donc conseillés dans ce cas [19].

Nous nous intéressons dans ce mémoire à la théorie des tests statistiques. Un intérêt particulier a été donné aux tests non paramétriques en mettant en évidence leurs applications dans différents domaines de la vie quotidienne, notamment en médecine.

Ce mémoire est constitué de quatre chapitres. Le premier chapitre s'intéresse à la notion de variables aléatoires, ainsi qu'à quelques lois usuelles. Le deuxième chapitre porte sur la théorie des tests paramétriques. Les tests non paramétriques font l'objet du chapitre trois. Le chapitre quatre contient quelques applications de ces tests dans le domaine de la médecine. Ces dernières ont été réalisées à l'aide du logiciel R. Ce mémoire se termine par une conclusion générale et quelques références bibliographiques.



# Chapitre 1

## Généralités sur les variables aléatoires et lois usuelles de probabilités

### Introduction

La théorie des probabilités est l'étude mathématique des phénomènes caractérisés par le hasard. Elle est utilisée dans les domaines d'application, tel que les sciences physique, l'économie sociale et humaine, etc. Actuellement, c'est une théorie qui s'applique le plus souvent dans l'étude des phénomènes aléatoires et l'analyse des expériences déterministes. L'étude des phénomènes aléatoires a commencé avec l'étude des jeux de hasard ; jeux de dés, tirage dans des urnes et jeux de pile ou face ont été des motivations pour comprendre et prévoir les expériences aléatoires.

Ce chapitre introduit les notions de variables aléatoires discrètes et continues, ainsi que quelques lois usuelles de probabilité.

### 1.1 Variable aléatoire

#### Définition 1.1.

Une variable aléatoire est une fonction définie sur l'ensemble des résultats possibles d'une expérience aléatoire.

Soit  $(\Omega, A, P)$  un espace de probabilité, on appelle variable aléatoire sur  $(\Omega, A, P)$  à valeurs dans l'ensemble  $E$  (fini ou infini dénombrable) toute application  $X$  de  $\Omega$  dans  $E$  telle que

[10] :

$$\begin{aligned} X : \Omega &\longrightarrow E \\ \omega &\longrightarrow X(\omega) = x \end{aligned}$$

$E$  est muni d'une tribu  $B$  telle que  $X^{-1}(A) \in \mathcal{A}, \forall A \in B$

Il existe deux types de variables aléatoires, à savoir les variables discrètes et les variables continues.

## 1.2 Variable aléatoire discrète

### Définition 1.2.

Une variable aléatoire  $X$  est dite discrète si elle ne prend qu'un nombre fini ou infini dénombrable de valeurs.

Nous rencontrons ce type de variables aléatoires dans différents domaines des sciences expérimentales, en l'occurrence [18] :

- Le nombre de petits par portée pour une espèce animale donnée (chien, vache, ...).
- Le nombre de bactéries dans un milieu de culture biologique.
- le nombre de mutations dans une séquence d'ADN, ...

### Exemple 1.1.

On lance deux fois une pièce de monnaie équilibrée. L'ensemble des résultats possible est :

$$\Omega = \{PP, PF, FP, FF\}$$

Chacun des événements de  $\Omega$  a une probabilité égale à  $1/4$  de se produire.

Considérons la variable aléatoire  $X$  représentant le nombre de "pile" obtenus.

donc :  $X(\Omega) = \{0, 1, 2\}$ .

### 1.2.1 Loi de probabilité (distribution de probabilité)

Soit  $X$  variable aléatoire discrète à valeurs dans  $E = x_1, \dots, x_n$ . La loi de probabilité d'une variable aléatoire discrète permet de connaître les chances d'apparition des différentes valeurs de cette variable, notée  $P_X$ , telle que :

- $P_X = P(X = x_i), i \in \{1, 2, \dots, n\}$ ,
- $P_X \in [0, 1]$  et  $\sum_{n \geq 0} P(X = x_n) = 1$ .

### 1.2.2 Fonction de répartition

#### Définition 1.3.

On appelle fonction de répartition d'une variable aléatoire discrète  $X$  la fonction  $F_X$  définie sur  $\mathbb{R}$  dans  $[0, 1]$  par :

$$F_X(x) = P(X \leq x) = \sum_{x_i \in E} P(X = x_i).$$

#### Propriété 1.1.

1.  $\forall x \in \mathbb{R}, 0 \leq F_X(x) \leq 1,$
2. Si  $a \leq b, P(a \leq X \leq b) = F_X(b) - F_X(a),$
3.  $F_X$  est une fonction croissante et en escalier sur  $\mathbb{R}.$

#### Remarque 1.1.

La représentation graphique de la fonction de répartition d'une variable aléatoire discrète est la courbe en escalier.

#### Exemple 1.2.

On jette un dé. Soit  $X$  la variable aléatoire égale aux résultats obtenus ( $X(\Omega) = \{1, 2, 3, 4, 5, 6\}$ ). La fonction de répartition de  $X$  est définie par :

$$F(x) = \begin{cases} 0 & \text{si } x < 1; \\ \frac{1}{6} & \text{si } x \in [1, 2[; \\ \frac{2}{6} & \text{si } x \in [2, 3[; \\ \frac{3}{6} & \text{si } x \in [3, 4[; \\ \frac{4}{6} & \text{si } x \in [4, 5[; \\ \frac{5}{6} & \text{si } x \in [5, 6[; \\ 1 & \text{si } x \geq 6. \end{cases}$$

### 1.2.3 Espérance et variance d'une variable aléatoire discrète

#### Espérance mathématique

L'espérance mathématique  $E(X)$  d'une variable aléatoire discrète  $X$  à valeurs dans  $E = \{x_1, \dots, x_n\}$  est la moyenne arithmétique des valeurs possibles pondérées par les probabilités correspondantes, elle est donnée par :

$$E(X) = \sum_{i=1}^n x_i P(X = x_i)$$

Si  $X$  est à valeurs dans  $E$ ,  $E$  dénombrable infinie.

**Propriété 1.2.**

- Si  $a$  est une constante alors :
  - $E(a) = a$ .
  - $E(aX) = aE(X)$ .
  - $E(X + a) = E(X) + a$ .
- L'espérance d'une somme de variables aléatoires (indépendantes ou non) est égale à la somme de leurs espérances.

$$E(X + Y) = E(X) + E(Y)$$

- Indépendance des variables :  
 $X$  et  $Y$  indépendantes  $\Rightarrow E(XY) = E(X) E(Y)$ .

**Preuve :** voir [10]

**Variance et l'écart type d'une variable aléatoire discrète**

1. La variance d'une variable aléatoire discrète, notée  $V(X)$ , est un paramètre de dispersion qui correspond au moment centré d'ordre 2 de la variable aléatoire  $X$ , alors la variance égale à :

$$V(X) = E[(X - E(X))^2] = E(X^2) - [E(X)]^2 = \left( \sum_{i=1}^n (x_i^2 \cdot P(X = x_i)) \right) - (E(X))^2.$$

2. L'écart type de la variable aléatoire  $X$ , que l'on note  $\sigma(X)$ , est la racine de  $V(X)$ .

$$\sigma(X) = \sqrt{V(X)}$$

**Covariance d'un variables aléatoire**

La covariance est un nombre permettant d'évaluer le sens de variation de deux variables, et ainsi de qualifier l'indépendance de ces variables (voir propriété 1.3). Si deux variables sont indépendantes leur covariance est nulle, mais la réciproque n'est pas toujours vraie.

**Théorème 1.1. (Théorème de Koenig[2])**

$$V(X) = E(X^2) - E(X)^2 = \left( \sum_{i \geq 1} (x_i^2 \cdot P(X = x_i)) \right) - (E(X))^2$$

**Démonstration**

On a :

$$V(X) = E[(X - E(X))^2]$$

$$V(X) = \sum_{i=1}^n ((x_i - E(X))^2) P(X = x_i)$$

$$V(X) = \sum_{i=1}^n (x_i^2 - 2x_i E(X) + E(X)^2) \cdot P(X = x_i)$$

En décomposant, trois expressions apparaissent :

$$V(X) = \sum_{i=1}^n x_i^2 \cdot P(X = x_i) - \sum_{i=1}^n 2x_i E(X) \cdot P(X = x_i) + \sum_{i=1}^n E(X)^2 \cdot P(X = x_i)$$

$$V(X) = \sum_{i=1}^n x_i^2 \cdot P(X = x_i) - 2E(X)E(X) + (E(X)^2) \times 1$$

$$V(X) = E(X^2) - 2E(X)^2 + E(X)^2$$

Finalement,

$$V(X) = E(X^2) - E(X)^2.$$

**Propriété 1.3.**

- Si  $a$  est une constante alors :

$$- V(X + a) = V(X).$$

$$- V(aX) = a^2 V(X).$$

- $V(X + Y) = V(X) + V(Y) + 2 \text{COV}(X, Y)$
- $\text{COV}(X, Y) = E(XY) - E(X) E(Y) = E[(X - E(X))(Y - E(Y))]$
- Si  $X$  et  $Y$  sont indépendantes, alors  $V(X, Y) = V(X) + V(Y)$ .

**Exemple 1.3.**

Dans un jeu, un joueur doit choisir entre deux questions, une question facile et une question difficile. S'il répond juste à la première question, il peut tenter de répondre à l'autre question. La question facile rapporte au joueur 1000DA et la question difficile lui rapporte 3000DA. Les questions sont indépendantes, et on estime avoir 30% de chances de bien répondre à la question difficile, et 60% de chances de répondre à la question facile.

Soit  $X$  la variable aléatoire égale au gain du jeu.

1. La loi de la variable aléatoire  $X$  dans le cas où le joueur choisit de répondre à la question facile en premier :

$$X(\Omega) = \{0, 1000, 4000\}.$$

$x_i$	0	1000	4000
$P(X = x_i)$	0.4	0.42	0.18

Le gain moyen du joueur est :

$$\begin{aligned} E(X) &= \sum_{i=1}^3 x_i P(X = x_i) \\ &= 0 \times (0.4) + 1000 \times (0.42) + 4000 \times (0.18) \\ &= 1140DA. \end{aligned}$$

2. La loi de la variable aléatoire  $X$  dans le cas où le joueur choisit de répondre à la question difficile en premier :

$$X(\Omega) = \{0, 3000, 4000\}.$$

$x_i$	0	3000	4000
$P(X = x_i)$	0.7	0.12	0.18

Le gain moyen du joueur est alors :

$$\begin{aligned} E(X) &= \sum_{i=1}^3 x_i P(X = x_i) \\ &= 0 \times (0.7) + 3000 \times (0.12) + 4000 \times (0.18) \\ &= 1080DA. \end{aligned}$$

## 1.3 Variable aléatoire continue

### Définition 1.4.

Une variable aléatoire continue est une variable qui peut prendre une infinité non dénombrable de valeurs, par exemple dans un intervalle ou sur tout  $\mathbb{R}$ .

### Quelques exemples :

- temps d'attente pour avoir le bus.
- taille des individus d'une population.

### 1.3.1 Fonction de répartition

#### Définition 1.5.

On appelle fonction de répartition de la variable aléatoire  $X$  la fonction  $F_X$  définie pour tout  $x \in \mathbb{R}$  par :

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(x) dx.$$

Autrement dit,  $F_X(x)$  est la probabilité de l'évènement "la valeur de  $X$  est inférieure ou égale à  $x$ ",  $f$  c'est la densité de probabilité.

**Remarque 1.2.**

1. Pour une variable aléatoire continue  $X$ , pour tout  $a \in \mathbb{R} : P(X = a) = 0$  on a donc :
  - $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$ ,
  - $P(a < X) = P(a \leq X < b)$ ,
  - $P(X > b) = P(X \geq b)$ .
2. La représentation graphique  $F_X$  prend la forme d'une courbe cumulative.

**1.3.2 Densité de probabilité**

Une variable aléatoire est dite absolument continue si  $\exists f > 0$ , telle que :

1.  $\forall x \in \mathbb{R}, f(x) \geq 0$  et  $F(x) = \int_{-\infty}^x f(t)dt$ .
2.  $\int_{-\infty}^{+\infty} f(x)dx = 1$ .
3.  $P(a < X < b) = F(b) - F(a) = \int_a^b f(x)dx$

**Exemple 1.4.**

Soit la fonction  $f$  définie par :

$$f(x) = \begin{cases} x & \text{si } x \in [0, 1] \\ -x + 2 & \text{si } x \in [1, 2] \\ 0 & \text{si } x < 0 \text{ ou } x \geq 2 \end{cases}$$

Il s'agit bien d'une densité de probabilité :

$$\begin{aligned} \int_{-\infty}^{+\infty} f(x) dx &= \int_0^1 x dx + \int_1^2 (-x + 2) dx \\ &= \left[ \frac{x^2}{2} \right]_0^1 + \left[ -\frac{x^2}{2} + 2x \right]_1^2 \\ &= 1. \end{aligned}$$

La fonction de répartition de la variable aléatoire associée à cette densité de probabilité :

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \int_0^x t dt = \frac{x^2}{2} & \text{si } x \in [0, 1[ \\ \int_0^1 t dt + \int_1^x (-t + 2) dt = -\frac{x^2}{2} + 2x - 1 & \text{si } x \in [1, 2[ \\ 1 & \text{si } x \geq 2 \end{cases}$$

### 1.3.3 Espérance et variance d'une variable aléatoire continue

#### Espérance mathématique

Soit  $X$  une variable aléatoire continue de densité  $f$ , son espérance si elle existe est définie par :

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx$$

#### Variance mathématique

La variance d'une variable aléatoire continue  $X$  si elle existe est le nombre positif définie par :

$$V(X) = \int_{-\infty}^{+\infty} (x - E(X))^2 f(x) dx = \int_{-\infty}^{+\infty} x^2 f(x) dx - E(X)^2.$$

## 1.4 Lois usuelles de probabilités

### 1.4.1 Lois discrètes

#### Loi de Bernoulli

Une expérience aléatoire ayant deux résultats possibles : succès ( auquel on associe la valeur 1) et échec ( auquel on associe la valeur 0) est appelé expérience de Bernoulli [13]. La v.a  $X$  suit une loi de probabilité appelée "loi de Bernoulli", notée  $\mathcal{B}(1, p)$ . Elle est définie par :

$$P : \{0, 1\} \rightarrow [0, 1]$$

$$P(X = 1) = p \quad \text{et} \quad P(X = 0) = 1 - p = q.$$

L'expérience de Bernoulli consiste à jeter une pièce de monnaie (forcement non équilibrée) pour laquelle la probabilité de succès est égale à  $p$  (soit par exemple obtenue face).

Dans ce cas la variable aléatoire  $X$  définie par

$$X = \begin{cases} 1, & \text{si on obtient un succès (face).} \\ 0, & \text{si on obtient un échec (pile).} \end{cases}$$

#### Cas particulier



Si la pièce est équilibrée alors  $p = \frac{1}{2}$

$$P(X = 1) = P(X = 0) = \frac{1}{2}$$

### Remarque 1.3.

La fonction indicatrice d'un événement  $A \in \Omega$  est définie par

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A. \\ 0 & \text{si } \omega \notin A. \end{cases}$$

### Propriété 1.4.

Soit  $X$  une v.a. de loi de Bernoulli notée  $X \rightsquigarrow \mathcal{B}(p)$ , alors :

1.  $X(\Omega) = \{0, 1\}$ .
2.  $\forall k \in \{0, 1\}, P(X = k) = p^k(1 - p)^{1-k}$ .
3.  $E(X) = p$  et  $Var(X) = p(1 - p)$

### Loi Binomiale

Soit une expérience de Bernoulli répétée  $n$  fois dans les mêmes conditions et de manières indépendantes  $(X_1, X_2, \dots, X_n)$ . La loi binomiale, notée  $\mathcal{B}(n, p)$ , modélise le nombre de succès obtenus lors de répétitions indépendantes de plusieurs expériences aléatoires identiques (avec  $p$  la probabilité du succès et  $q = 1 - p$  la probabilité de l'échec) [13].

La variable aléatoire  $X = \sum_{i=1}^n X_i = X_1 + X_2 + \dots + X_n$  correspond au nombre de succès, si on a  $k$  succès on aura  $(n - k)$  échecs et la probabilité d'avoir  $k$  succès dans une expérience aléatoire répétée  $n$  fois, est :

$$P(X = k) = C_n^k \cdot p^k \cdot q^{(n-k)} \quad k \in \{0, 1, \dots, n\}.$$

$$\text{avec } C_n^k = \frac{n!}{(n-k)!k!}.$$

Il est facile de démontrer que l'on a bien une loi de probabilité, car :

$$\sum_{k=0}^n P(X = k) = \sum_{k=0}^n C_n^k \cdot p^k \cdot q^{(n-k)} = (p + q)^n = 1, \text{ car } p + q = 1.$$

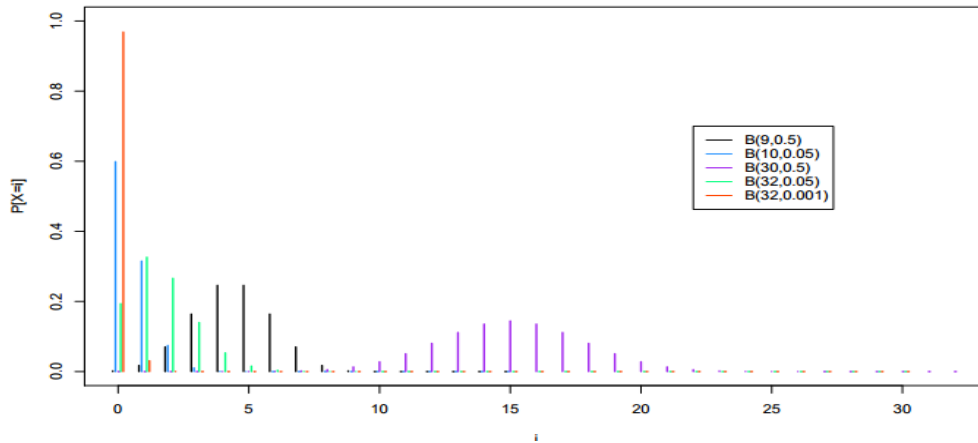


FIGURE 1.1 – Lois de probabilité pour la loi Binomiale.

**Remarque 1.4.**

Le développement du binôme de Newton  $(p + q)^n$  permet d'obtenir l'ensemble des probabilités pour une distribution binomiale  $\mathcal{B}(n, p)$  avec  $n$  et  $p$  des valeurs données.

**Espérance mathématique**

L'espérance mathématique de la distribution binomiale  $\mathcal{B}(n, p)$  est :

$$E(X) = n \cdot p$$

En effet,

$$\begin{aligned} E(X) &= E\left(\sum_{i=1}^n X_i\right) \text{ où chaque } X_i \text{ est une v.a. de Bernoulli} \\ &= E(X_1 + X_2 + \cdots + X_n) \\ &= E(X_1) + E(X_2) + \cdots + E(X_n) \\ &= \sum_{i=1}^n E(X_i) = \sum_{i=1}^n p = n \cdot p \end{aligned}$$

**Variance**

La variance de cette variable aléatoire qui suit une loi binomiale  $\mathcal{B}(n, p)$  est :

$$V(X) = n \cdot p \cdot q$$

En effet,

$$\begin{aligned}
 V(X) &= V\left(\sum_{i=1}^n X_i\right) \text{ où chaque } X_i \text{ est une v.a. de Bernoulli} \\
 &= V(X_1 + X_2 + \cdots + X_n) \\
 &= V(X_1) + V(X_2) + \cdots + V(X_n) \\
 &= \sum_{i=1}^n V(X_i) = \sum_{i=1}^n p \cdot q = n \cdot p \cdot q
 \end{aligned}$$

### Écart type

L'écart type est

$$\sigma(X) = \sqrt{V(X)} = \sqrt{n \cdot p \cdot q}$$

### Exemple 1.5.

On considère un dé cubique dont les faces sont numérotées de 1 à 6, et elles ont toutes la même probabilité d'apparaître.

On lance le dé 9 fois de suite. On considère que l'on obtient un succès si la réponse obtenue est supérieure ou égale à 5. Soit  $X$  la variable aléatoire associée au nombre du succès obtenus sur les 9 jets.

Sur un jet la probabilité de succès est  $p = \frac{2}{6} = \frac{1}{3}$ , la probabilité de l'échec est donc  $q = (1 - p) = \frac{4}{6} = \frac{2}{3}$ .

$$X(\Omega) = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$P(X = 0) = (2/3)^9, \quad P(X = 9) = (1/3)^9 \text{ et } P(X = 4) = C_9^4 (1/3)^4 (2/3)^5.$$

En effet, on choisit la place des 4 succès parmi les 9 jets.

$$\forall k \in \{0, 1, \dots, 9\} \quad P(X = k) = C_9^k (1/3)^k (2/3)^{9-k}$$

On dit que  $X$  suit la loi binomiale de taille 9 et de paramètre  $1/3$ , notée  $\mathcal{B}(9, 1/3)$ .

### Théorème 1.2.

Si  $X \rightsquigarrow \mathcal{B}(n, p)$  et  $Y \rightsquigarrow \mathcal{B}(m, p)$  sont deux variables aléatoires indépendantes de même probabilité  $p$ , alors leur somme  $X + Y$  est une variable aléatoire qui suit une loi binomiale [2] :

$$X + Y \rightsquigarrow \mathcal{B}(n + m, p).$$

### Loi géométrique

La loi géométrique est une loi de probabilité qui modélise l'observation du nombre d'épreuves de Bernoulli identiques et indépendantes devant se succéder pour espérer un

premier succès.

On dit qu'une variable aléatoire  $X$  suit une loi géométrique de paramètre  $p$ , notée  $\mathcal{G}(p)$ , si :

1.  $X(\Omega) = \mathbb{N}^*$ .
2.  $\forall k \in X(\Omega), P(X = k) = p(1 - p)^{k-1}$ .

Elle admet pour moments :

$$E(X) = \frac{1}{p} \text{ et } V(X) = \frac{1-p}{p^2}.$$

### Loi de poisson

On dit qu'une variable aléatoire  $X$  suit une loi de poisson (appelée aussi loi des événements rares ou de petits nombres) de paramètre réel  $\lambda$ , notée  $\mathcal{P}(\lambda)$ , si elle prend des valeurs entières dont les probabilités de réalisation sont [13] :

$$\forall k \in \mathbb{N}, P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}.$$

Elle admet pour moments :

$$E(X) = V(X) = \lambda.$$

### Propriété 1.5.

Si  $X \rightsquigarrow \mathcal{P}(\lambda_1)$  et  $Y \rightsquigarrow \mathcal{P}(\lambda_2)$ , les variables aléatoires  $X$  et  $Y$  étant indépendantes, alors

$$X + Y \rightsquigarrow \mathcal{P}(\lambda_1 + \lambda_2).$$

### Exemple 1.6.

Une centrale téléphonique reçoit en moyenne 300 appels par heure. Quelle est la probabilité que durant une minute, la centrale reçoit exactement deux appels ?

On a : les appels dans cette centrale suivent une loi de poisson de paramètre  $\lambda = \frac{300}{60} = 5$  appels par minutes en moyenne.

$$\begin{aligned} P(X = 2) &= e^{-\lambda} \frac{\lambda^k}{k!} = e^{-5} \frac{5^2}{2!} \\ &= 0,08422. \end{aligned}$$

### L'approximation de la loi Binomiale par la loi de Poisson

La loi Binomiale dépend de deux paramètres  $n$  et  $p$ , alors que la loi de Poisson ne dépend que d'un seul paramètre  $\lambda$ . Pour qu'une loi Binomiale soit au plus proche d'une loi de Poisson, on doit au moins souhaiter que ces deux lois aient la même espérance. L'espérance de la loi Binomiale étant  $np$  et celle de la loi de Poisson étant  $\lambda$ , il faut que  $\lambda = np$ . Cette condition nécessaire n'est pas suffisante pour réaliser une telle approximation, théoriquement l'approximation est parfaite lorsque [8] :

$$\begin{cases} n & \rightarrow +\infty \\ p & \rightarrow 0 \\ np & = \text{constante.} \end{cases}$$

En pratique, la condition :

$$\begin{cases} n & > 30 \\ np & < 5 \end{cases}$$

Où :

$$\begin{cases} n & > 50 \\ p & < 0.1 \end{cases}$$

est suffisante pour envisager l'approximation.

**Preuve : Voir [17].**

#### Exemple 1.7.

Des observations ont montré que la probabilité qu'un homme soit atteint d'une maladie M est  $p = 0.1$ . En considérant 40 hommes pris au hasard, soit  $k$  le nombre d'hommes touchés par la maladie et  $X$  la variable aléatoire qui compte le nombre d'hommes malades.

1.  $X$  suit une loi binomiale  $\mathcal{B}(40; 0.1)$  :

$$P(X = k) = C_{40}^k \cdot \left(\frac{1}{10}\right)^k \cdot \left(\frac{9}{10}\right)^{(40-k)}$$

2. Calcul de  $E(X)$ ,  $V(X)$  et  $\sigma(X)$  :

$$\begin{aligned} E(X) &= np = 40 \times \frac{1}{10} = 4. \\ V(X) &= npq = 40 \times \frac{1}{10} \times \frac{9}{10} = 3,6. \\ \sigma(X) &= \sqrt{V(X)} = 1,89. \end{aligned}$$

3. Cette loi binomiale  $\mathcal{B}(40; 0.1)$  est approchée par une loi de Poisson  $\mathcal{P}(\lambda)$  :

$(n = 40 > 25 \text{ et } np = 4 < 5) \Rightarrow \text{Binomiale } \mathcal{B}(40; 0.1) \rightsquigarrow \text{Poisson } \mathcal{P}(\lambda), \lambda = np = 4.$

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} = e^{-4} \frac{4^k}{k!}.$$

4. Calcul de  $P(X = 2)$  et  $P(X = 5)$  :

$$P(X = 2) = e^{-4} \frac{4^2}{2!} = 0,14653.$$

$$P(X = 5) = e^{-4} \frac{4^5}{5!} = 0,15629.$$

## 1.4.2 Lois usuelles continues

### Loi uniforme

Une variable aléatoire  $X$  suit une loi uniforme sur l'intervalle  $[a, b]$ , notée  $X \rightsquigarrow U_{[a,b]}$ , si sa densité de probabilité  $f$  est définie par :

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{si } a \leq X \leq b, \\ 0, & \text{sinon.} \end{cases}$$

Sa fonction de répartition :

$$F(x) = \begin{cases} 0, & \text{si } x < a, \\ \frac{x-a}{b-a}, & \text{si } a \leq x \leq b, \\ 1, & \text{si } x > b. \end{cases}$$

Elle admet pour moments :

$$E(X) = \int_a^b x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a} \left[ \frac{x^2}{2} \right]_a^b = \frac{a+b}{2}.$$

$$V(X) = \frac{(b-a)^2}{12}.$$

### Loi exponentielle

On dit que  $X$  suit la loi exponentielle de paramètre  $\lambda > 0$ , notée  $exp(\lambda)$ , si elle admet la densité :

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0, x \in \mathbb{R}, \\ 0, & \text{sinon.} \end{cases}$$

Sa fonction de répartition est :

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{si } x \geq 0, \\ 0, & \text{sinon.} \end{cases}$$

Elle admet pour moments :

$$E(X) = \frac{1}{\lambda} \quad \text{et} \quad V(X) = \frac{1}{\lambda^2}$$

### Remarque 1.5.

La loi exponentielle est une loi sans mémoire, c'est à dire la probabilité qu'un phénomène dure au moins  $s + t$  heures (ou n'importe quelle autre unité de temps), sachant qu'il a déjà duré  $s$  heures sera la même que la probabilité de durer  $t$  heures. On a alors :

$$P(T > t + s | T > s) = P(T > t).$$

### Loi gamma

Une variable aléatoire suit la loi gamma de paramètre  $a$  et  $b$ , si sa fonction densité est donnée par la fonction gamma, qui est définie pour  $x > 0$  par :

$$f(x) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & , x \geq 0 \quad , a > 0 \quad , b > 0. \\ 0 & \text{Sinon.} \end{cases}$$

$$\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt.$$

Elle admet pour moments :

$$E(X) = \frac{a}{b} \quad \text{et} \quad V(X) = \frac{a}{b^2}$$

### Propriété 1.6.

- pour  $n \in \mathbb{N}$ ,  $\Gamma(n) = (n - 1)!$  et  $\Gamma(1) = 1$ .
- $\Gamma(x + 1) = x \Gamma(x)$ .
- $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ .
- $\Gamma(n + \frac{1}{2}) = \frac{2n!}{2^{2n} n!} \sqrt{\pi}$ .
- $\beta(p, q) = \frac{\Gamma(p) \Gamma(q)}{\Gamma(p + q)}$ .

### Loi normale

La loi normale est une loi très importante en statistique, plusieurs phénomènes ont une loi de probabilité très proche de cette loi. Elle est souvent utilisée pour faire des approximations dans le domaine de l'estimation et des tests d'hypothèses [17].

#### Loi normale centrée réduite( ou loi gaussienne)

Une variable aléatoire continue  $X$  suit la loi normale centrée réduite, notée  $X \rightsquigarrow \mathcal{N}(0, 1)$ , si sa densité de probabilité est définie par :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad x \in \mathbb{R} .$$

Nous admettrons ici que  $f$  est bien une densité de probabilité

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$$

$f$  est une fonction symétrique en forme de cloche.

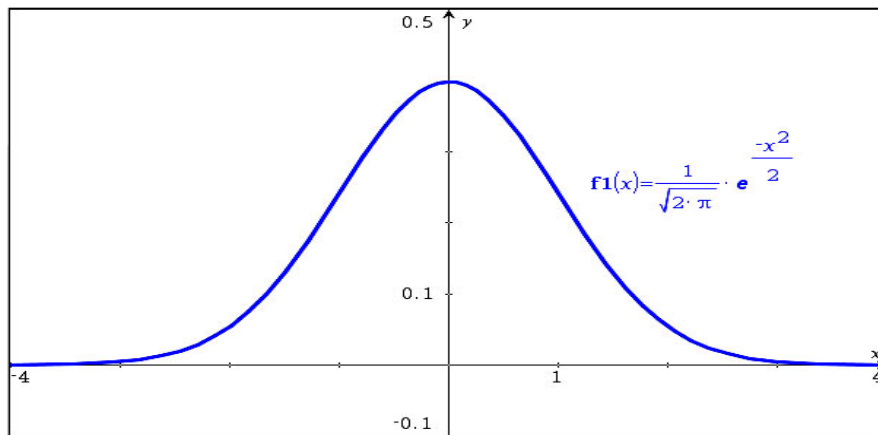


FIGURE 1.2 – La loi normale centré réduit.

Sa fonction de répartition  $F(x)$  s'écrit :

$$\phi = P(X < x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$



**Propriété 1.7.**

Si  $X \rightsquigarrow \mathcal{N}(0, 1)$  alors  $E(X) = 0$  et  $V(X) = 1$ .

**Remarque 1.6.**

Notons  $\phi$  la fonction de répartition associée à la variables  $X$  de loi  $\mathcal{N}(0, 1)$ , on a :

- Pour  $x < 0$ , on utilise la formule  $\phi(x) + \phi(-x) = 1$ .
- Pour  $x > 0$ , on dispose aussi la relation  $P(-x \leq X \leq x) = 2\phi(x) - 1$ .
- Pour tous  $a, b \in \mathbb{R}$ , avec  $a < b$  :  $P(a < X < b) = F(b) - F(a)$ .

**Exemple 1.8.**

On suppose que  $X \rightsquigarrow \mathcal{N}(0, 1)$ .

$$P(X < 1.25) = \phi(1.25) \simeq 0.8944$$

$$\phi(1.0669) = 0.85 + 0.007 = 0.857.$$

$$P(X < 0) = 0.5.$$

$$P(X > 2.34) = 1 - P(X < 2.34) = 0.0096.$$

$$P(X < -0.47) = \phi(-0.47) = 1 - \phi(0.47).$$

$$P(X \in [-1.5, 2.3]) = \phi(2.3) - \phi(-1.5) = \phi(2.3) - (1 - \phi(1.5)) = 0.9225.$$

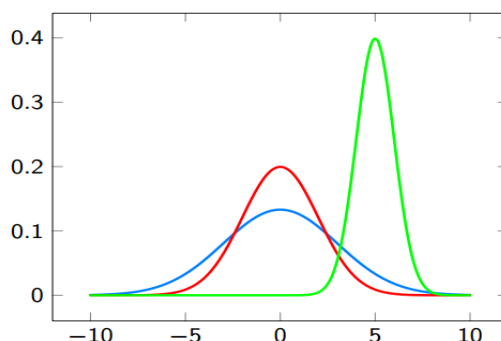
$$P(X \in [-1, 1]) = \phi(1) - \phi(-1) = \phi(1) - (1 - \phi(1)) = 2\phi(1) - 1 = 0.6826$$

**La loi normale de moyenne  $m$  et de variance  $\sigma^2$**

Une variable aléatoire continue  $X$  suit une loi normale de moyenne  $m$  et de variance  $\sigma^2$ , si sa fonction densité de probabilité est définie par :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \quad x \in \mathbb{R}$$

Dans ce cas  $E(X) = m$  et  $V(X) = \sigma^2$ .

$\mathcal{N}(0, 3), \mathcal{N}(0, 2), \mathcal{N}(5, 1)$ 

 FIGURE 1.3 – La loi normale de moyenne  $m$  et de variance  $\sigma^2$ .

**Remarque 1.7.**

Si  $X \rightsquigarrow \mathcal{N}(m, \sigma^2)$  alors  $\frac{X-m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$ .

Pour calculer la fonction de répartition d'une variable aléatoire normale  $\mathcal{N}(m, \sigma^2)$ , il faut transformer cette variable en variable aléatoire centrée et réduite, puis on utilise la table de la loi normale  $\mathcal{N}(0, 1)$  pour la lecture de sa valeur.

$$P(a < X < b) = P\left(\frac{a-m}{\sigma} < \frac{X-m}{\sigma} < \frac{b-m}{\sigma}\right) = \Phi\left(\frac{b-m}{\sigma}\right) - \Phi\left(\frac{a-m}{\sigma}\right).$$

**Exemple 1.9.**

Considérons  $X \rightsquigarrow \mathcal{N}(10, 6)$  alors on cherche  $P(5 < X < 11)$  :

$$\begin{aligned} P(5 < X < 11) &= P\left(\frac{5-10}{\sqrt{6}} < \frac{X-10}{\sqrt{6}} < \frac{11-10}{\sqrt{6}}\right) \\ &= \Phi\left(\frac{11-10}{\sqrt{6}}\right) - \Phi\left(\frac{5-10}{\sqrt{6}}\right) \\ &= \Phi(0.41) - \Phi(-2.04) \\ &= 0.63635. \end{aligned}$$

**Approximation par une loi normale**
**Loi Binomiale**

Une loi binomiale  $B(n, p)$  peut être approximée par une loi normale  $\mathcal{N}(m, \sigma)$ , où  $m = n p$  et  $\sigma = \sqrt{n p (1-p)}$  lorsque :

$$\begin{cases} n & \geq 30, \\ n p & \geq 5, \\ n p (1-p) & \geq 5. \end{cases}$$

Autrement dit,  $n$  doit être assez grand et  $p$  est proche de 0.5 [8].

#### Loi de poisson

Une loi de poisson de paramètre  $\lambda$  tend vers une loi normale de moyenne  $\lambda$  et l'écart type  $\sqrt{\lambda}$ , lorsque  $\lambda$  est assez grand ( $\lambda > 20$ ) [23].

### 1.4.3 Autres lois utiles pour les tests

On va s'intéresser maintenant à plusieurs lois utiles pour les tests statistiques et construites à partir de variables gaussiennes.

#### Loi de Khi-deux

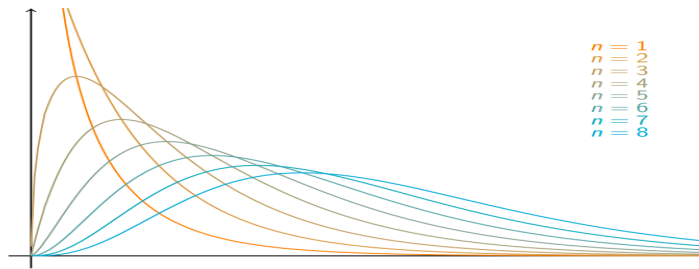
Une variable aléatoire est dite suit la loi de khi-deux à  $n$  degrés de liberté et on note  $X \rightsquigarrow \chi_n^2$  si sa fonction de densité est suit la loi gamma de paramètre  $(\frac{n}{2}, \frac{1}{2})$ , elle est donnée par [9] :

$$f(x) = \frac{(\frac{1}{2})^{\frac{n}{2}}}{\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{1}{2}x}, \quad x \in \mathbb{R}^+.$$

Soient  $X_1, \dots, X_n$  un échantillon de  $n$  variables aléatoires indépendantes et identiquement distribuées de loi  $\mathcal{N}(0, 1)$ , alors la variable aléatoire  $Y = X_1^2 + X_2^2 + \dots + X_n^2$  suit la loi de  $\chi_n^2$ .

Elle admet pour moments :

$$E(Y) = n \quad \text{et} \quad V(Y) = 2n.$$

FIGURE 1.4 – La loi de Khi deux pour  $n$  degré de liberté.**Remarque 1.8.**

Cette loi est principalement utilisée dans le test du  $\chi^2$  en statistique. Il permet de tester l'adéquation entre des données et une famille de lois.

**Propriété 1.8.**

1. Soient  $Z \rightsquigarrow \mathcal{N}(0, 1)$  et  $X = Z^2$ , alors  $X \rightsquigarrow \chi_1^2$ .
2. Soit  $(X_1, X_2, \dots, X_n)$  un échantillon issu d'une loi normale  $\mathcal{N}(\mu, \sigma^2)$ .

Posons :

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

On a  $((X_i - \mu)/\sigma)^2 \rightsquigarrow \chi_1^2$  pour  $i = 1, 2, \dots, n$ . On déduit que :

$$n \frac{S^2}{\sigma^2} = \sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \rightsquigarrow \chi_n^2.$$

**Théorème 1.3.**

Soient  $U$  et  $V$  deux variables aléatoires indépendantes telle que  $U \rightsquigarrow \chi_u^2$  et  $V \rightsquigarrow \chi_v^2$ . Alors, on a :  $W = U + V \rightsquigarrow \chi_{u+v}^2$ .

**La loi de Student**

Une variable aléatoire  $X$  suit la loi de Student si sa fonction de densité est donnée par :

$$f(t) = \frac{1}{\sqrt{\pi}} \frac{1}{\beta(\frac{1}{2}, \frac{n}{2})} \frac{1}{1 + (\frac{t}{\sqrt{n}})^{\frac{n+1}{2}}}, \quad t > 0.$$

On note  $T \rightsquigarrow t_n$  (on lit  $T$  suit la loi de Student à  $n$  degrés de liberté).

La loi de Student est symétrique, de moyenne nulle, sa variance est égale à  $\left(\frac{n}{n-2}\right)$  et dépend d'un paramètre  $n$  appelé nombre de degrés de liberté.

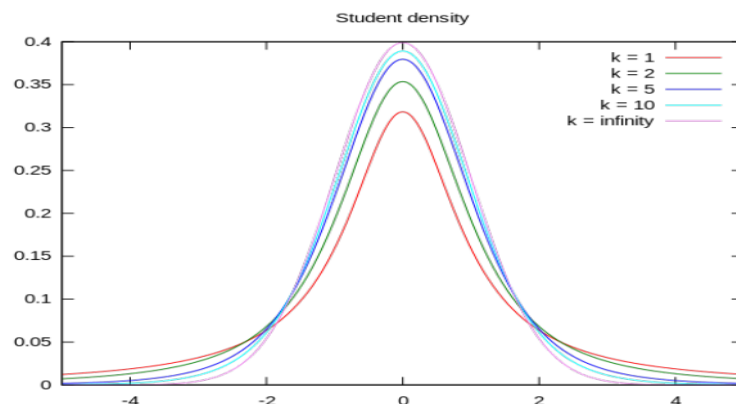


FIGURE 1.5 – La densité de la loi de Student.

**Propriété 1.9.**

Soit  $X$  une variable aléatoire,  $X \rightsquigarrow N(0, 1)$ , et soit  $Y$  une variable aléatoire,  $Y \rightsquigarrow X_n^2$ , avec  $X$  et  $Y$  indépendantes alors la variable aléatoire  $T = \frac{X}{\sqrt{\frac{Y}{n}}} \rightsquigarrow t_n$ .

**Remarque 1.9.**

La loi de Student est très utilisée dans les tests statistiques, elle permet également de construire un intervalle de confiance pour l'estimation de la moyenne. Elle est utilisée pour les petits échantillons.

Soit  $X_1, X_2, \dots, X_n$  un échantillon gaussien d'une variable aléatoire  $X \rightsquigarrow N(\mu, \sigma)$ .

On considère  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ .

1. Si  $\mu$  est connu, alors

$$\bar{X} \rightsquigarrow N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

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

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \Rightarrow \frac{n S_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2 \rightsquigarrow X_n^2$$

Comme  $\bar{X}$  et  $S_n^2$  sont indépendantes alors la variable aléatoire :

$$T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{\sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma}\right)^2}{n}}} = \frac{\bar{X} - \mu}{\sqrt{\sum_{i=1}^n (X_i - \mu)^2}} \rightsquigarrow t_n.$$

2. Si  $\mu$  est inconnu, alors

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

$$\frac{n S_n'^2}{\sigma^2} \rightsquigarrow X_{n-1}^2$$

$$T = \frac{\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{n S_n'^2}{\sigma^2(n-1)}}} = \frac{\bar{X} - \mu}{\frac{S_n'^2}{\sigma^2}} \sqrt{n-1} \rightsquigarrow t_{n-1}$$

### La loi de Fisher

La loi de Fisher joue un rôle essentiel pour la comparaison des variances de variables dont les réalisations sont issues d'échantillons extraits de deux populations. Ce type de test est particulièrement utile à l'analyse de variance (ANOVA) et à l'analyse de la régression [25, 9].

Soient  $X \rightsquigarrow X_n^2$  et  $Y \rightsquigarrow X_m^2$  deux variables aléatoires indépendantes. Alors  $Z = \frac{X/n}{Y/m}$  suit une loi de Fisher de paramètres  $n$  et  $m$ . On note  $Z \rightsquigarrow F(n, m)$ , sa densité est :

$$f(z) = \frac{\Gamma((n+m)/2)(n/m)^{n/2} z^{n/2-1}}{\Gamma(n/2)\Gamma(m/2)[(n+m)z+1]^{(n+m)/2}}, \quad z > 0$$

où  $\Gamma$  désigne la fonction gamma.

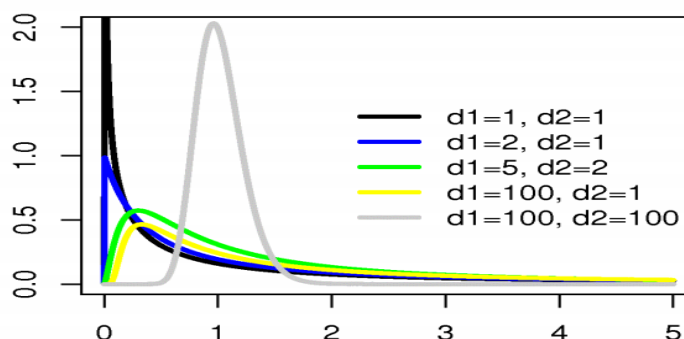


FIGURE 1.6 – La loi de Fisher.

**Propriété 1.10.**

Si  $X \rightsquigarrow F(n, m)$ , alors  $E(X) = \frac{m}{m-2}$  pour  $m > 2$  et  $V(X) = \frac{2m^2(n+m-2)}{n(m-4)(m-2)^2}$  pour  $m > 4$ .

**Remarque 1.10.**

Si  $X$  suit une loi de Student à  $n$  degrés de liberté

$$X \rightsquigarrow t_n$$

Alors  $X^2$  suit une loi de Fisher de paramètres 1 et  $n$

$$X^2 \rightsquigarrow \mathcal{F}_{1,n}.$$

**Propriété 1.11.**

Soient  $(X_1, X_2, \dots, X_n)$  et  $(Y_1, Y_2, \dots, Y_m)$  deux échantillons issus respectivement de lois  $\mathcal{N}(\mu_X, \sigma_X^2)$  et  $\mathcal{N}(\mu_Y, \sigma_Y^2)$ . On suppose que les moyennes théoriques  $\mu_X$  et  $\mu_Y$  sont connues.

On a :

$$\frac{n S_X^2}{\sigma_X^2} \rightsquigarrow \chi_n^2 \quad \text{et} \quad \frac{m S_Y^2}{\sigma_Y^2} \rightsquigarrow \chi_m^2.$$

Les échantillons étant indépendants, on a donc :

$$(\sigma_X^2)/(\sigma_Y^2) \rightsquigarrow F_{n,m},$$

Cette propriété est utilisée pour trouver un intervalle de confiance au niveau  $1 - \alpha$  pour le rapport de variances  $\sigma_X^2/\sigma_Y^2$ , lorsque les moyennes théoriques  $\mu_X$  et  $\mu_Y$  sont connues.



# Chapitre 2

## Tests Paramétriques

### Introduction

En statistique, un test (ou test d'hypothèse) est une procédure de décision entre deux hypothèses faites sur une population, il s'agit d'accepter ou de rejeter une hypothèse statistique. On distingue deux catégories de tests :

1. tests paramétriques ;
2. tests non paramétriques.

Nous présentons dans ce chapitre les notions de base sur les tests statistiques, ainsi que les principaux tests paramétriques.

### 2.1 Généralités sur les tests statistiques

Le principe des tests statistiques est de poser une hypothèse de travail et de prédire les conséquences de cette hypothèse pour la population ou l'échantillon. Les différentes étapes qui doivent être suivies pour tester une hypothèse :

1. Le choix de  $H_0$  et  $H_1$ .
2. Détermination de la variable de décision.
3. Calcul de la région critique ( $RC$ ) en fonction d'un seuil  $\alpha$  fixé à l'avance.
4. Calcul de la puissance  $1 - \beta$ .
5. Calcul de la valeur expérimentale de la variable de décision.
6. Conclusion, rejet ou non rejet de  $H_0$ .

### 2.1.1 L'hypothèse nulle et l'hypothèse alternative

L'objectif d'un test statistique est toujours de trancher entre deux hypothèses antagonistes. La première est appelée hypothèse nulle et notée  $H_0$  et la deuxième est appelée hypothèse alternative et notée par  $H_1$ .

#### Exemple 2.1.

On souhaite déterminer si une pièce de monnaie est bien équilibrée. La probabilité d'obtenir "pile" en lançant cette pièce suit une loi de Bernoulli de paramètre  $\theta$ . Le problème posé est le choix d'une hypothèse parmi :

$$\begin{aligned} H_0 & : \theta = \frac{1}{2}; \\ H_1 & : \theta \neq \frac{1}{2}. \end{aligned}$$

#### Remarque 2.1.

On dit que le test est :

- Bilatéral si elle est la forme : " $\theta \neq \theta_0$ ".
- Unilatéral à gauche si elle est la forme : " $\theta < \theta_0$ ".
- Unilatéral à droite si elle est la forme : " $\theta > \theta_0$ ".

### 2.1.2 Les erreurs et les risques

Le résultat d'un test statistique est le rejet ou le non rejet d'une hypothèse nulle  $H_0$ .

#### Définition 2.1.

L'erreur pour un test statistique est de rejeter l'hypothèse nulle alors qu'elle est vraie ou de l'acceptée alors qu'elle est fausse.

On distingue alors deux types d'erreurs :

1. Une erreur dite de première espèce qu'est la décision de rejeter  $H_0$  alors que  $H_0$  est vraie. La probabilité de commettre cette erreur est  $\alpha = P(H_1|H_0)$ , appelée risque de première espèce ou seuil du test.
2. Une erreur de seconde espèce qu'est la décision d'accepter  $H_0$  alors que  $H_1$  est vraie. La probabilité de commettre cette erreur est  $\beta = P(H_0|H_1)$ , appelée risque de deuxième espèce.

On peut résumer la situation dans le tableau suivant [6] :

Décision	$H_0$ est vraie	$H_1$ est vraie
non rejet de $H_0$	$1 - \alpha$	erreur de deuxième espèce $\beta$
rejeter $H_0$	erreur de première espèce $\alpha$	$1 - \beta$

TABLE 2.1 – Erreurs de première et deuxième espèces

$$\alpha = P(H_1|H_0 \text{ vraie}) = P(\overline{H_0}|H_0);$$

$$\beta = P(H_0|H_1 \text{ vraie}) = P(H_0|\overline{H_0}).$$

$\alpha$  est le risque de première espèce et  $\beta$  le risque de deuxième espèce.

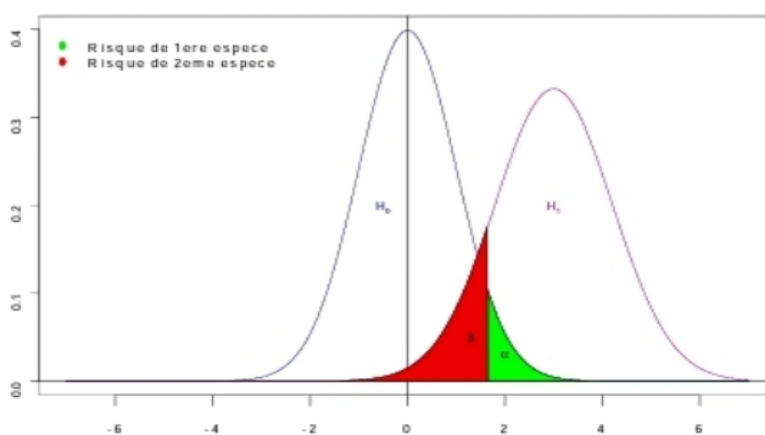


FIGURE 2.1 – L'erreur de première et deuxième espèce.

### Remarque 2.2.

Les seuils de signification les plus utilisés sont

$$\alpha = 0.1, 0.05, \text{ et } 0.01$$

### 2.1.3 Variable de décision

#### Définition 2.2.

C'est une variable statistique qui doit apporter le maximum d'informations sur le problème posé et dont la loi sera différente sous  $H_0$ . Il faut que sa loi soit entièrement connue au moins sous  $H_0$  (i.e. lorsque  $H_0$  est vraie).

### 2.1.4 Région critique (région de rejet)

#### Définition 2.3.

On appelle région critique  $RC$  du test toute région, contenant l'ensemble des valeurs de la variable de décision qui conduisent à rejeter  $H_0$ .

La détermination d'une région critique  $RC$  dépend de  $\alpha$  et vérifie  $P(RC|H_0) = \alpha$ .

Le complémentaire de  $RC$  est la région d'acceptation, noté par  $\overline{RC}$ , elle est déterminée par  $P(\overline{RC}|H_1) = 1 - \alpha$ .

### 2.1.5 Puissance d'un test

#### Définition 2.4.

On appelle puissance d'un test, la probabilité de rejeter  $H_0$  et d'accepter  $H_1$  alors que  $H_1$  est vraie, sa valeur est  $1 - \beta$  [22].

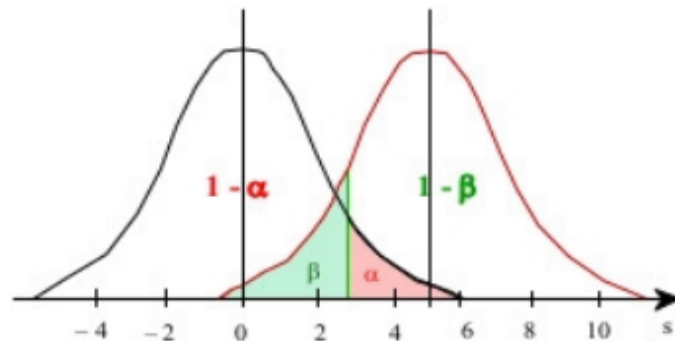


FIGURE 2.2 – La relation entre les deux risques d'erreur.

### 2.1.6 Notions de p-valeur

Les logiciels de calcul statistique expriment souvent le résultat d'un test en fournissant une grandeur appelée  $p$  – *value* en anglais. C'est une probabilité critique qu'on note par  $\alpha_{obs}$ . Aussi pour un seuil de significativité  $\alpha$  donnée, on compare  $p$  – *valeur* et  $\alpha$ , afin d'accepter ou rejeter  $H_0$  [5].

- Si  $p$  – *valeur*  $\leq \alpha$ , on rejete l'hypothèse  $H_0$ .
- Si  $p$  – *valeur*  $> \alpha$ , on accepte l'hypothèse  $H_0$ .

## 2.2 Principaux tests paramétriques usuels

Un test paramétrique est un test dont la forme fonctionnelle de la distribution de l'échantillon (ou des échantillons) à tester est connue mais dépend d'un paramètre inconnu. Le test porte sur le paramètre inconnu.[7, 11]

Les tests paramétriques sont de deux types

1. Les tests de conformité à une norme portent sur un seul paramètre. Ils sont construits sur la base d'une valeur connue que doit ou devrait avoir ce paramètre. On teste, par exemple, si la durée moyenne des grossesses est égale à 41 semaines.
2. Les tests de comparaison (ou d'homogénéité) tranchent entre deux hypothèses concernant le même paramètre mesuré dans deux populations.

Pour les tests paramétriques de conformité, le paramètre testé est noté  $\theta$  et la norme qui sert de référence est notée  $\theta_0$ . Pour les tests paramétriques de comparaison, les valeurs du paramètre testé sont notées  $\theta_A$  et  $\theta_B$ .

Les hypothèses d'un test statistique sont simples ou composites. Une hypothèse simple s'exprime sous la forme d'une égalité soit :

- Test de conformité :  $\theta = \theta_0$ .
- Test de comparaison :  $\theta_A = \theta_B$ .

Une hypothèse composite se traduit par une ou plusieurs inégalités :

- Test de conformité :  $\theta > \theta_0$  ou  $\theta < \theta_0$  ou  $\theta \neq \theta_0$ .
- Test de comparaison :  $\theta_A > \theta_B$  ou  $\theta_A < \theta_B$  ou  $\theta_A \neq \theta_B$ .

### 2.2.1 Comparaison des moyennes (Test de Student)

Comparaison d'une moyenne à une valeur donnée

On suppose dans ce chapitre que les échantillons sont issus d'une loi normale ou d'une loi approximée par une loi normale.

Soit  $X_1, X_2, \dots, X_n$  un échantillon gaussien de moyenne  $\mu$  et de variance  $\sigma^2$ .

Test :

$$H_0 : \text{"}\mu = \mu_0\text{"} \quad \text{vs} \quad H_1 : \text{"}\mu \neq \mu_0\text{"}.$$

On sait qu'une estimation ponctuelle de  $\mu$  est la moyenne empirique donnée par :

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

d'où

$$\begin{aligned}\bar{X}_n &= \frac{1}{n} \sum_{i=1}^n X_i \rightsquigarrow \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right) \\ \Rightarrow \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} &\rightsquigarrow \mathcal{N}(0, 1).\end{aligned}$$

Sous  $H_0 : \mu = \mu_0$ , la statistique de décision est  $U = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

- On accepte l'hypothèse nulle si la valeur  $\mu_0$  appartient à l'intervalle de confiance de niveau de confiance  $1 - \alpha$
- On rejette l'hypothèse nulle sinon.

1.  $\sigma$  connue :

La région critique de la statistique est de la forme  $RC = [|U| > k_\alpha]$ , avec  $k_\alpha$  est donnée par la table de la loi normale centrée réduite  $\mathcal{N}(0, 1)$ , telle que :

$$\begin{aligned}\alpha &= P(|U| > k_\alpha) = P(U < -k_\alpha) + P(U > k_\alpha) \\ &= 2P(U > k_\alpha) = 2(1 - P(U \leq k_\alpha)).\end{aligned}$$

On a  $\alpha = 2(1 - P(U \leq k_\alpha))$  alors  $P(U \leq k_\alpha) = 1 - \frac{\alpha}{2}$ .

Par conséquent,  $P(U < k_\alpha) = P(U \leq k_\alpha) = 1 - \frac{\alpha}{2}$ .

En utilisant la table de  $\mathcal{N}(0, 1)$  on déduit  $k_\alpha : k_\alpha = F_{\mathcal{N}(0,1)}^{-1}$ .

L'intervalle de confiance de moyenne  $\mu$  au niveau de confiance  $1 - \frac{\alpha}{2}$  est :

$$\left[ \bar{X}_n - U_\alpha \frac{\sigma}{\sqrt{n}}; \bar{X}_n + U_\alpha \frac{\sigma}{\sqrt{n}} \right]$$

**Exemple 2.2.**

D'après une société conseil, le salaire annuel moyen d'administrateurs de banques de données serait de 49738 Euros. Une enquête effectuée auprès d'un organisme sur un échantillon aléatoire de 36 entreprises de ce secteur donne les résultats suivants concernant la rémunération des administrateurs de banque de données :

Salaire moyen : 50200 Euros.

Écart type : 1560 Euros.

Est-ce que les résultats de cette enquête permettraient de supporter l'affirmation de cette société conseil ? Utiliser un seuil de signification de  $\alpha = 0.05$ .

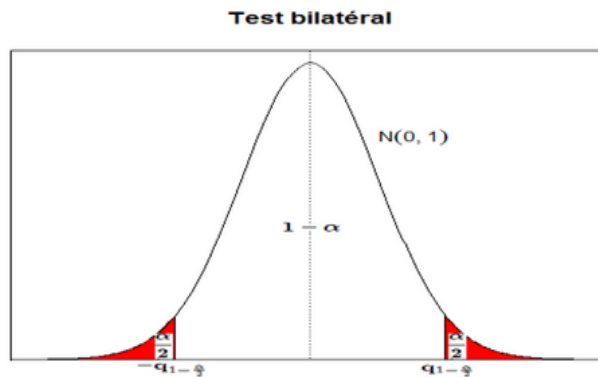


FIGURE 2.3 – Région d'acceptation et de rejet du test bilatéral de la moyenne ( $\sigma$  connue) au niveau de signification  $\alpha$ .

(a) Hypothèse Statistique :

$$H_0 : \mu = 49738 \quad \text{vs} \quad H_1 : \mu \neq 49738.$$

(b) Seuil de signification :

$$\alpha = 0.05$$

(c) Condition d'application du test :

Grand échantillon ( $n \geq 30$ ) provenant d'une population de variance connue.

(d) La statistique :

La statistique qui convient pour le test est  $\bar{X}_n$ , l'écart réduit est  $Z = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}}$  où  $\mu_0 = 49738$  et  $Z$  suit la loi  $N(0, 1)$ .

(e) Règle de décision :

D'après  $H_1$  et au seuil  $\alpha = 0.05$  les valeurs critiques de l'écart réduit sont  $Z_{\frac{\alpha}{2}}$ .

On adoptera la règle de décision suivante :

Rejeter  $H_0$  si  $Z > 1.96$ , sinon ne pas rejeter  $H_0$ .

(f) Calcul de l'écart réduit :

Puisque  $\bar{X}_n = 50200$ ,  $\sigma = 1560$  et  $n = 36$ , donc :

$$Z = \sqrt{36} \frac{50200 - 49738}{1560} = \frac{462}{260} = 1.77.$$

(g) Décision et conclusion :

La valeur de  $Z = 1.77$  se situe dans la région de non rejet de  $H_0$  donc on ne peut rejeter l'affirmation de la société conseil. L'écart observé entre  $\bar{X}_n$  et  $\mu_0$ , soit  $50200 - 49738 = 462$ , n'est pas statistiquement significatif au seuil  $\alpha = 0.05$ .

2.  $\sigma$  inconnue :

Sous  $H_0$ , la variable aléatoire  $\bar{X}_n$  suit une loi normale  $\mathcal{N}(\mu, \frac{\sigma^2}{n})$  comme la variance est inconnu on l'estime par la variance empirique donnée par

$$S'^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2.$$

On a :

$$\frac{(n-1)S_n'^2}{\sigma^2} = \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma} \right)^2 \rightsquigarrow \chi_{n-1}^2$$

et

$$T_{n-1} = \frac{U}{\sqrt{\frac{(n-1)S_n'^2}{(n-1)\sigma^2}}} = \frac{\frac{\bar{X} - \mu_0 \sqrt{n}}{\sigma}}{\frac{\sqrt{n-1}S_n'}{\sqrt{n-1}\sigma}} = \sqrt{n-1} \frac{\bar{X} - \mu_0}{S_n'} \rightsquigarrow t_{n-1}.$$

La région critique  $W = [|T_{n-1}| > k_\alpha]$ .

Pour calculer l'intervalle de confiance de la moyenne  $\mu$ , on a remplacer  $\sigma$  par son estimateur ponctuelle  $S$  fournie par l'estimation  $S^2$  (ce nombre n'est autre que l'écart type) calculé sur l'échantillon de taille  $n$  avec  $n-1$  degrés de liberté (ddl). On utilise le procédé dit de studentisation qui consiste à remplacer la variable  $U = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$  par la variable  $T = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n}}}$  qui suit une loi de Student à  $n-1$  ddl. La table de Student nous permet de déterminer  $t_{n-1, \alpha}$ .

On obtiendra alors l'intervalle de confiance suivant :

$$\left[ \bar{X} - t_{n-1, \alpha} \frac{S}{\sqrt{n}}, \bar{X} + t_{n-1, \alpha} \frac{S}{\sqrt{n}} \right]$$

**Exemple 2.3.**

Un échantillon aléatoire de  $n = 16$  prélèvements indépendants est tiré dans une population normale (d'écart-type inconnu). Les valeurs sont :

487.5, 519.8, 485.2, 503.1, 505.1, 473.3, 500.1, 470.3

499.4, 504.9, 494.7, 480.3, 500.2, 499.7, 480.7, 488.3.

Faire le test de comparaison à la valeur donnée  $\mu_0 = 500$ , en adoptant le risque de première espèce  $\alpha = 0.05$ .

On a :

Effectif  $n = 16$ . Somme des valeurs observées  $\sum_{i=1}^{16} X_i = 789.6$



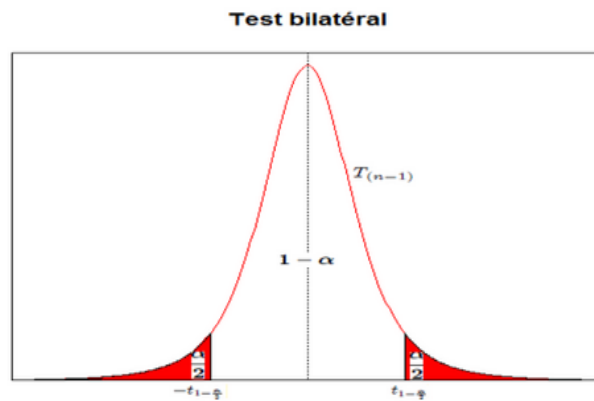


FIGURE 2.4 – Région d'acceptation et de rejet du test bilatéral de la moyenne ( $\sigma$  inconnue) au niveau de signification  $\alpha$ .

Somme des carrés des valeurs  $\sum_{i=1}^{16} X_i^2 = 3895948$

Somme des carrés centrés  $\sum_{i=1}^{16} (X_i - \bar{X}_n)^2 = 2627.0775$

### Éléments de calcul

Pour  $\alpha = 0.05$ , estimation de la moyenne :  $\hat{\mu} = \bar{X}_n = 493.288$

estimation de la variance  $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X}_n)^2 = 175.1385$ ,  $S = 13.234$ .

**Test :**

$$H_0 : \mu = 500 \quad vs \quad H_1 : \mu \neq 500$$

$$t_{1-\frac{\alpha}{2}}(15) = 2.13.$$

d'où les limites de l'intervalle d'acceptation

$$500 - 2.13 \frac{13.234}{\sqrt{16}} = 492.357 \quad , \quad 500 + 2.13 \frac{13.234}{\sqrt{16}} = 507.643$$

On a  $\bar{X} = 493.288 \in [492.357, 507.643]$ .

Donc on accepte l'hypothèse  $H_0$ .

### Comparaison de deux moyennes

Soient  $X_1, X_2, \dots, X_n$  un  $n$  échantillon d'une variable aléatoire  $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$  et  $Y_1, Y_2, \dots, Y_m$  un  $m$  échantillon d'une variable aléatoire  $Y \rightsquigarrow \mathcal{N}(\mu', \sigma')$ .

**Remarque 2.3.**

Pour comparer deux échantillons, on compare en premier les variances; et si le test est positif  $\sigma = \sigma'$  on va comparer les moyennes, sinon on dit que les deux échantillons n'appartiennent pas à la même population.

On suppose que  $\sigma = \sigma'$

On teste :

$$H_0 : \mu = \mu' \quad vs \quad H_1 : \mu \neq \mu'$$

On a :

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightsquigarrow \mathcal{N}(0, 1) \quad \text{et} \quad \frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{m}}} \rightsquigarrow \mathcal{N}(0, 1).$$

Sous

$$H_0 : \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \rightsquigarrow \mathcal{N}(0, 1)$$

$$\bar{X} - \bar{Y} \rightsquigarrow \mathcal{N}\left(0, \sqrt{\text{var}(\bar{X} - \bar{Y})}\right),$$

avec

$$\text{var}(\bar{X} - \bar{Y}) = \text{var}(\bar{X}) + \text{var}(\bar{Y})$$

$$= \frac{\sigma^2}{n} + \frac{\sigma^2}{m}.$$

Donc

$$\bar{X} - \bar{Y} \rightsquigarrow \mathcal{N}\left(0, \sigma \sqrt{\frac{1}{n} + \frac{1}{m}}\right)$$

$$\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \rightsquigarrow \mathcal{N}(0, 1).$$

1.  $\sigma$  est connu :

La région critique  $W = \left[ \left| \frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}} \right| > k_\alpha \right]$

2.  $\sigma$  est inconnu :

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

$$S_Y^2 = \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y})^2 = \frac{m S_Y^2}{\sigma^2} \rightsquigarrow \chi_{m-1}^2$$

$$\sigma^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{i=1}^m (Y_i - \bar{Y})^2}{n + m} = \frac{n S_X^2 + m S_Y^2}{n + m}$$

et on a :  $\frac{n S_X^2}{\sigma^2} + \frac{m S_Y^2}{\sigma^2} \rightsquigarrow \chi_{n+m-2}^2$

$$T = \frac{\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n} + \frac{1}{m}}}}{\sqrt{\frac{n S_X^2 + m S_Y^2}{\sigma^2 (n+m-2)}}} = \frac{\bar{X} - \bar{Y}}{\sqrt{n S_X^2 + m S_Y^2}} \frac{\sqrt{n+m-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \rightsquigarrow t_{n+m-2}.$$

La région critique est donnée par  $[|T| > k_\alpha]$ .

**Remarque 2.4.** Si la variable aléatoire parente  $X$  ne suit pas une loi normale, les tests précédents s'appliquent encore dès que  $n$  est grand ( $n > 30$ ) en raison du théorème centrale limite.

**Exemple 2.4.**

Deux échantillons aléatoires sont tirés de façon indépendante dans deux populations d'où sont prélevées respectivement  $n = 16$  et  $m = 14$  unités indépendantes. Les variances des deux populations sont supposées connues,  $\sigma_1^2 = 400$  et  $\sigma_2^2$ . Les valeurs sont :

**Echantillon n° 1 :** 487.5, 519.8, 485.2, 503.1, 505.1, 473.3, 500.1, 470.3,  
499.4, 504.9, 494.7, 480.3, 500.2, 499.7, 480.7, 488.3

**Echantillon n° 2 :** 480.3, 494, 486.1, 534.5, 521.8, 479.8, 496.9  
528.6, 540.3, 532.6, 558.2, 547.2, 553.8, 486.1

Faire le test de comparaison des deux moyennes, en adoptant le risque de première espèce  $\alpha = 0.05$ .

**Données relatives aux échantillons**

	Echantillon n°1	Echantillon n°2
Taille de l'échantillon	16	14
Somme des valeurs	$\sum X_i = 7892.6$	7240.2
Somme des carrés des valeurs	$\sum X_i^2 = 3895948$	3755107.98
Somme des carrés centrés	$\sum (X_i - \bar{X})^2 = 2627.0775$	10786.8343.

**Eléments de calcul**

$\sigma = 0.05$

Estimation des moyennes :  $\hat{\mu}_1 = \bar{X} = 493.288$  et  $\hat{\mu}_2 = \bar{Y} = 517.157$

Variance de la différence  $(\bar{X} - \bar{Y})$  :  $\sigma_d^2 = \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}$  d'ou  $\sigma_d = 9.45$ .

**Test**

$$H_0 : \mu_1 = \mu_2 \quad vs \quad H_1 : \mu_1 \neq \mu_2$$

$$\bar{X} - \bar{Y} = -23.869, \quad U_{1-\frac{\alpha}{2}} \sigma_d = 1.96 \times 9.45 = 18.522.$$

$$\bar{X} - \bar{Y} = -23.869 < -U_{1-\frac{\alpha}{2}} \sigma_d = -18.522, \text{ alors on rejette } H_0.$$

## 2.2.2 Comparaison des variances (Test de Fisher)

Comparaison de variance d'une variable aléatoire gaussienne à une constante donnée

Soit  $X_1, X_2, \dots, X_n$  un échantillon issu d'une variable aléatoire  $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$  et soit  $\sigma_0$  une constante positive. On teste :

$$H_0 : \sigma = \sigma_0 \quad \text{vs} \quad H_1 : \sigma \neq \sigma_0.$$

Deux cas se présentent :  $\mu$  connu ou  $\mu$  inconnu.

1.  $\mu$  connue :

$$\text{Soit } S_n^2 = \frac{1}{n} \sum (X_i - \mu)^2.$$

La statistique  $\frac{n S_n^2}{\sigma_0^2} = \sum \left( \frac{X_i - \mu}{\sigma_0} \right)^2$  suit une loi de Khi-deux à  $n$  degrés de liberté.

La région critique du test est donnée par :

$$RC = [ |S_n^2 - \sigma_0^2| > k_\alpha ], \text{ avec } P_{H_0}(W) = \alpha.$$

Pour déterminer la région critique (déterminer  $k_\alpha$ ), on doit connaître la loi de la statistique de décision.

On a :

$$\frac{n S_n^2}{\sigma_0^2} \rightsquigarrow \chi_n^2.$$

$$|S_n^2 - \sigma_0^2| > k_\alpha \Leftrightarrow S_n^2 - \sigma_0^2 > k_\alpha \quad \text{ou} \quad S_n^2 - \sigma_0^2 < -k_\alpha.$$

$$\begin{aligned} P_{H_0} (|S_n^2 - \sigma_0^2| > k_\alpha) &= P(S_n^2 > k_\alpha + \sigma_0^2) + P(S_n^2 < -k_\alpha + \sigma_0^2). \\ &= P\left(\frac{n S_n^2}{\sigma_0^2} > \frac{n(k_\alpha + \sigma_0^2)}{\sigma_0^2}\right) + P\left(\frac{n S_n^2}{\sigma_0^2} < \frac{-n(k_\alpha + \sigma_0^2)}{\sigma_0^2}\right) \\ &= 1 - F_{\chi_n^2}\left(\frac{n(k_\alpha + \sigma_0^2)}{\sigma_0^2}\right) + F_{\chi_n^2}\left(\frac{n(-k_\alpha + \sigma_0^2)}{\sigma_0^2}\right) = \alpha \\ &\Leftrightarrow F_{\chi_n^2}\left(\frac{n(k_\alpha + \sigma_0^2)}{\sigma_0^2}\right) = \frac{1-\alpha}{2}. \end{aligned}$$

La valeur de  $\frac{n(k_\alpha + \sigma_0^2)}{\sigma_0^2}$  est lue sur la table de Khi-deux .

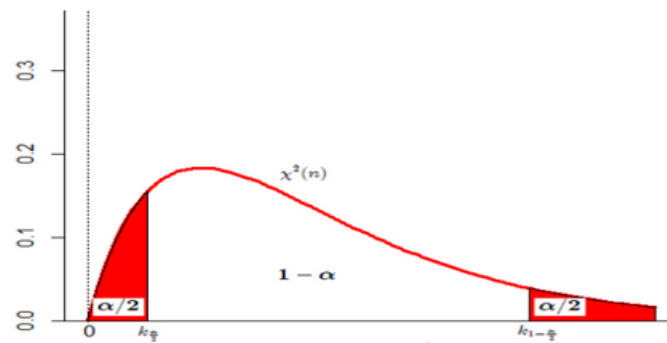


FIGURE 2.5 – Région d’acceptation et de rejet du test bilatéral de la variance ( $\mu$  connue) au niveau de signification  $\alpha$ .

2.  $\mu$  est inconnue :

On utilise la statistique  $\frac{(n-1) S_n^2}{\sigma_0^2}$ .

En effet, sous ( $H_0$ ) cette statistique suit une loi de chi-deux à  $n - 1$  degré de liberté.

On a :

$$T_n = \frac{(n-1) S_n^2}{\sigma_0^2} = \sum_{i=1}^n \left( \frac{X_i - \bar{X}}{\sigma_0} \right)^2 \rightsquigarrow \chi_{n-1}^2.$$

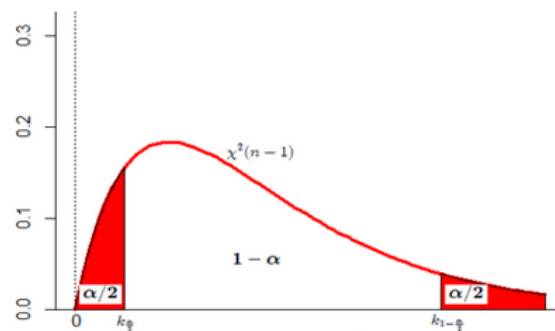


FIGURE 2.6 – Région d’acceptation et de rejet du test bilatéral de la variance ( $\mu$  inconnue) au niveau de signification  $\alpha$ .

**Exemple 2.5.**

La responsable d’une entreprise a voulu tester la dextérité de ses employés. La variabilité des résultats aux tests de dextérité n’excède pas 144 ( $\sigma = 12$ ). Les résultats de ce test ont été obtenus sur un échantillon aléatoire de vingt employés

donnent une somme de carrés des écarts par rapport à la moyenne de 2952. On suppose que l'échantillon aléatoire provient d'une population normale.

L'hypothèse selon laquelle  $\sigma^2$  n'excède pas 144 est-elle acceptable au seuil de signification  $\alpha = 0.05$  ?

1. **Hypothèse Statistique :**

$$H_0 : \sigma = 12 \quad vs \quad H_1 : \sigma > 12$$

2. **Seuil de signification :**

$$\alpha = 0.05$$

3. **La statistique :**

La statistique qui convient pour le test en supposant  $H_0$  vraie est  $T_n = \frac{n S^2}{\sigma_0^2}$  où  $\sigma_0^2 = 144$ . La quantité  $T_n$  est distribuée selon la loi de chi-deux à 20 degrés de liberté.

4. **Règle de décision :**

D'après  $H_1$  et au seuil  $\alpha = 0.05$  la valeur critique de  $\chi_{0.05,20}^2 = 30.1435$ .

On adoptera la règle de décision suivante :

Rejeter  $H_0$  si  $T_n > 30.1435$ , sinon on accepte  $H_0$ .

5. **Calcul de l'écart réduit :**

On a  $\sum (X_i - \bar{X}_n)^2 = 2952$  d'où  $T_n = 20.5$ .

6. **Décision et conclusion :**

Puisque  $T_n = 20.5 < 30.1435$ , on ne peut pas rejeter  $H_0$ .

### Comparaison de deux variances gaussiennes

Soit  $X_1, X_2, \dots, X_n$  un  $n$  échantillon de  $X \rightsquigarrow \mathcal{N}(\mu_1, \sigma_1)$ , soit  $Y_1, Y_2, \dots, Y_n$  un  $m$  échantillon de  $Y \rightsquigarrow \mathcal{N}(\mu_2, \sigma_2)$ .

**Remarque 2.5.** Pour comparer deux échantillons, on teste d'abord l'égalité des variances.

On veut tester l'hypothèse :

$$H_0 : \text{''}\sigma_1 = \sigma_2\text{''} \quad vs \quad H_1 : \text{''}\sigma_1 \neq \sigma_2\text{''}.$$

La règle de décision consiste à rejeter  $H_0$  si :

$$|S_1^2 - S_2^2| > k_\alpha \Leftrightarrow \frac{S_1^2}{S_2^2} > k_\alpha.$$

$$S_1^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2, \quad \frac{n S_1^2}{\sigma_1^2} \rightsquigarrow \chi_{n-1}^2$$

$$S_2^2 = \frac{1}{m} \sum_{i=1}^m (Y_i - \bar{Y})^2, \quad \frac{m S_2^2}{\sigma_2^2} \rightsquigarrow \chi_{m-1}^2$$

La variable de décision est la statistique :

$$Z = \frac{\frac{n S_1^2}{\sigma_1^2}}{\frac{m S_2^2}{\sigma_2^2}} \rightsquigarrow \mathcal{F}(n-1, m-1).$$

Sous  $H_0 : \sigma_1 = \sigma_2$

Donc

$$\frac{n S_1^2}{m S_2^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sum_{i=1}^m (Y_i - \bar{Y})^2} \rightsquigarrow \mathcal{F}(n-1, m-1).$$

On a :

$$\alpha = P\left(\frac{n S_1^2}{m S_2^2} > k_\alpha\right) + P\left(\frac{n S_1^2}{m S_2^2} < -k_\alpha\right)$$

La valeur de  $k_\alpha$  est lue sur la table de Fisher :  $k_\alpha = F_{\mathcal{F}(n-1, m-1)}^{-1}\left(\frac{1-\alpha}{2}\right)$ .

**Exemple 2.6.**

Deux échantillons aléatoires sont tirés de façon indépendante dans deux populations normales d'où sont prélevées  $n = m = 20$  unités indépendantes. Les valeurs sont :

**Echantillon n° 1 :** 8, 7.8, 8, 8, 8, 7.8, 8, 7.9, 8, 8.1,

8, 8.1, 7.7, 7.5, 8, 8, 8.1, 7.9, 7.8, 8.2

**Echantillon n° 2 :** 7.7, 7.8, 8.4, 8.2, 8, 7.9, 8.2, 7.8, 8.1, 8.1

7.9, 8.1, 8, 8.2, 8.4, 8.1, 8.4, 8.1, 7.8, 8.2

Faire le test de comparaison des deux variances en adoptant le risque de première espèce  $\alpha = 0.05$ .

	Echantillon n°1	Echantillon n°2
Taille de l'échantillon	20	20
Somme des valeurs	$\sum X_i = 158.9$	$\sum Y_i = 161.4$
Somme des carrés des valeurs	$\sum X_i^2 = 1262.95$	$\sum Y_i^2 = 1303.32$
Somme des carrés centrés	$\sum (X_i - \bar{X})^2 = 0.4895$	$\sum (Y_i - \bar{Y})^2 = 0.822$

**Élément de calcul**

$\alpha = 0.05$	estimations des variances	$\sigma_1^2 = 2.576 \times 10^{-2}$	$\sigma_2^2 = 4.326 \times 10^{-2}$
	degré de liberté	$v_1 = 19$	$v_2 = 19$

Test

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad vs \quad H_1 : \sigma_1^2 \neq \sigma_2^2$$

En inverse l'ordre des échantillons pour disposer d'un rapport de variances supérieur à 1,

on a :

$$\frac{S_2^2}{S_1^2} = \frac{0.822/19}{0.4895/19} = 1.679 < 2.53 = \mathcal{F}_{0.975}(19, 19), \text{ alors on ne rejette pas l'hypothèse } H_0.$$

### 2.2.3 Tests sur les proportions

Soit  $p$  la proportion d'une population possédant le caractère considéré. Soit  $X$  une variable aléatoire observée sur cette population et soit  $(X_1, X_2, \dots, X_n)$  un  $n$  échantillon extrait de cette population. Soit  $K_n$  le nombre d'individus qui vérifient un certain caractère. On veut savoir si la fréquence observée (empirique)  $F = \frac{K}{n}$  estimateur de  $p$ , appartient à une population de fréquence  $p_0$  (sous  $H_0$ ) ou à une autre population inconnu de fréquence  $p$  (sous  $H_0$ ).

$$H_0 : p = p_0 \quad vs \quad H_1 : p \neq p_0$$

On sait que  $\frac{K}{n} = F \rightsquigarrow \mathcal{N}\left(p_0, \sqrt{\frac{p_0(1-p_0)}{n}}\right)$  sous  $H_0$ .

$$W = \left\{ x \in \mathbb{R}, \left| F - \frac{K}{n} - p_0 \right| > k \right\}.$$

$$P\left(\left| F - \frac{K}{n} - p_0 \right| > k\right) = \alpha \Rightarrow P\left(\frac{\left| F - \frac{K}{n} - p_0 \right|}{\sqrt{\frac{p_0(1-p_0)}{n}}} > \frac{k}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = \alpha.$$

$$k = \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) \sqrt{\frac{p_0(1-p_0)}{n}} = \phi_{\mathcal{N}(0,1)}^{-1}(1 - \alpha) \sqrt{p_0 \frac{1-p_0}{n}}.$$

#### Exemple 2.7.

Sur un échantillon de 200 étudiants 45% sont favorable qu'il y ait synthèse à la fin de l'année. Ceci contredit-il l'hypothèse qu'un étudiant sur deux y est favorable ?

$$\begin{cases} H_0 : p = \frac{1}{2}; \\ H_1 : p \neq \frac{1}{2}. \end{cases}$$

$\alpha = 0.05$  ,  $u = 1.96$ .

$$W = \left\{ x \in \mathbb{R}, \left| F - \frac{1}{2} \right| > 1.96 \sqrt{\frac{(1/2)^2}{200}} \right\}$$

Comme  $|F - 0.5| = |0.45 - 0.5| = 0.05$  donc on ne rejette pas  $H_0$  au seuil  $\alpha = 0.05$ .



## Conclusion

Nous avons abordé dans ce chapitre les tests paramétriques. Néanmoins, ces derniers ne peuvent pas s'appliquer dans beaucoup de situations, par exemple en présence d'échantillons de petites tailles.

A cet effet, des tests non paramétriques ont été développés. Ces derniers font l'objet du prochain chapitre.

# Chapitre 3

## Tests non paramétriques

### 3.1 Introduction

Dans ce chapitre, nous allons nous intéresser à quelques tests très utilisés en statistique non paramétrique. Ces tests sont basés sur les résultats d'expériences qui sont des réalisations de variables aléatoires dont la loi est complètement inconnue, contrairement, aux tests paramétriques où on suppose que la forme fonctionnelle de la densité est connue et dépend d'un paramètre.

### 3.2 Test non paramétrique

#### Définition 3.1.

Un test non paramétrique est un test dont la forme fonctionnelle de la distribution de l'échantillon (ou des échantillons) à étudier est complètement inconnue. On utilise uniquement l'information apportée par les résultats de l'expérience [7, 11].

Les tests non paramétriques les plus répandus sont les tests d'ajustement, les tests d'indépendance et les tests de comparaison de distributions.

**Les tests non paramétriques présentent les avantages suivants [16] :**

- Leur application est relativement facile et rapide,
- Ils peuvent s'appliquer à de petits échantillons,
- Ils traitent des données qualitatives exprimées soit en variables nominales soit par la comparaison des rangs,
- Ils peuvent s'appliquer à des données incomplètes ou imprécises.

### 3.3 Test de conformité à une loi donnée

Le test consiste à vérifier si un échantillon suit une loi donnée.

#### 3.3.1 Test de Kolmogorov Smirnov

Le test de Kolmogorov Smirnov (KS) est un test d'hypothèse basé sur la comparaison de la fonction de répartition empirique (construite à partir de l'échantillon) à la fonction de répartition théorique  $F_0(x)$  connue sous  $H_0$ .

Soit  $X_1, X_2, \dots, X_n$  un échantillon issue d'une variable aléatoire  $X$  de fonction de répartition inconnue et  $F_0$  une fonction de répartition connue [21].

On teste :

$$H_0 : "F(x) = F_0(x)" \quad vs \quad H_1 : \exists x \text{ tq} : "F(x) \neq F_0(x)" .$$

On a :  $F_n^*$  la fonction de répartition empirique de  $X_1, X_2, \dots, X_n$  :

$$F_n^*(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(X_i < x)}$$

La statistique de décision du test de KS est définie par :

$$D_n = \sup |F_n^*(x) - F_0(x)|$$

La région critique du test est :

$$RC = [D_n > k_\alpha].$$

Au seuil  $\alpha$  fixé,  $P_{H_0}(D_n > k_\alpha) = \alpha$ .

On rejette  $H_0$  si  $D_n > k_\alpha$  où  $k_\alpha$  est la quantile théorique lu à partir de la table de Kolmogorov Smirnov.

#### 3.3.2 Test de Cramer-Von Mises

Le test de Cramer-Von Mises est une véritable alternative au test de Kolmogorov-Smirnov. Il permet également de tester toute forme de différentiation entre les distributions.

Le test d'hypothèse s'écrit [20] :

$$H_0 : F = F_0 \quad vs \quad H_1 : \exists x, F \neq F_0.$$

$F$  : la fonction de répartition empirique.

$F_0$  : la fonction de répartition théorique.

Ce test s'intéresse à l'ensemble des écarts quadratiques entre  $F$  et  $F_0$  :

$$\int_{\mathbb{R}} |F(x) - F_0(x)|^2 dx.$$

La fonction de répartition empirique de  $X_1, X_2, \dots, X_n$  est :  $F_n^*(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{]-\infty, x]}(x_i)$ ,  
où  $x_i$  est une réalisation de  $X$ .

La statistique de décision est donnée par :

$$I_n = \frac{1}{12n} + \sum_{i=1}^n \left[ \frac{2i-1}{2n} - F_n^*(x_i) \right]^2.$$

Région critique du test :

$$W = [I_n > k_\alpha]$$

Pour un niveau  $\alpha$  donnée la valeur de  $k_\alpha$  calculée à partir de la table de Gramer-Von Mises.

### Exemple 3.1.

On dispose d'un échantillon de  $n$  matériels identiques et on note les durées de vie en heures  $x_1, x_2, \dots, x_n$ . On a  $n = 5$  et :

$$X_1 = 8, X_2 = 58, X_3 = 122, X_4 = 133, X_5 = 169.$$

On test :

$H_0$  :  $X$  suit une loi exponentielle vs  $H_1$  :  $X$  ne suit pas une loi exponentielle.

Le moyenne est estimée par  $\bar{x} = \frac{1}{5} \sum_{i=1}^5 x_i = 98$ , la fonction de répartition estimée est :

$$F(x) = 1 - \exp\left(-\frac{x}{98}\right),$$

d'où le tableau :

$x_i$	8	58	122	133	169
$F(x_i)$	0.079	0.447	0.711	0.743	0.821

La statistique de Cramer-Von-Mises vaut :

$$I_5 = \sum_{i=1}^5 \left( \frac{2i-1}{10} - F(x_i) \right)^2 + \frac{1}{60} = 0.09133$$

La valeur critique,  $K_\alpha$ , pour le test de Cramer Von Mises pour un échantillon de taille  $n = 5$  et un niveau de signification  $\alpha = 0.05$  vaut 0.199.

$I_5$  étant inférieure à  $K_{0.05}$ , alors on ne rejette pas l'hypothèse nulle. D'où  $X$  est de loi exponentielle.

### 3.3.3 Test de Khi-deux (Test de Pearson)

Soit  $X_1, X_2, \dots, X_n$  un  $n$  échantillon issu de  $X$  qui suit une loi donnée notée  $P$ . On considère une partition en classes notées  $C_1, C_2, \dots, C_k$ . On note  $N_i$  la variable aléatoire qui compte le nombre d'observations de l'échantillon appartenant à la classe  $C_i$  [21] :

$$N_i = \sum_{j=1}^n \mathbb{1}_{C_i}(x_j).$$

Soit  $p_i = P_0(X \in C_i)$  les probabilités théoriques, pour une loi  $P$ , on note  $p_i = P(X \in C_i)$  et on teste :

$$H_0 : \forall i = 1, \dots, k, p_i = p_i^0$$

$$H_1 : \exists i \text{ tq } p_i \neq p_i^0$$

Si on retient  $H_0$ , on conclura que la loi de  $X$  est  $P_0$ .

La statistique de décision du test de  $\chi^2$  est donnée par :

$$D^2 = \sum_{i=1}^k \frac{(N_i - np_i^0)^2}{np_i^0}$$

$$N_i \rightsquigarrow B(n, p_i^0)$$

Pour  $n$  assez grand  $N_i \rightsquigarrow \mathcal{N}(np_i^0, \sqrt{np_i^0(1-p_i^0)})$ .

D'où :

$$\frac{(N_i - np_i^0)^2}{np_i^0} \rightsquigarrow \chi_1^2.$$

Et  $D^2 \rightsquigarrow \chi_{k-1}^2$

On a la région critique :  $RC = [D^2 > k_\alpha]$ .

pour un risque  $\alpha$  donne :  $P_{H_0}(D^2 > k_\alpha) = \alpha$ .

On rejette  $H_0$  si  $D^2$  constaté est supérieur à la valeur théorique  $k_\alpha$  lue dans la table du Khi-deux à  $k - 1$  degré de liberté.

**Exemple 3.2.**

Un pisciculteur possède un bassin qui contient trois variétés de truites : communes, saumonées et arc-en ciel. Il voudrait savoir s'il peut considérer que son bassin contient autant de truites de chaque variété. Pour cela, il effectue, au hasard 339 prélèvements avec remise et obtient les résultats suivants :

Variétés	commune	saumonée	arc-en-ciel
Effectifs	145	188	136

TABLE 3.1 – Résultats de chaque variété de truite.

On cherche à savoir s'il y a équirépartition des truites entre chaque espèce c'est à dire on suppose que  $P_0$  est la loi uniforme.

$n = 399$ , une probabilité de  $1/3$  pour chaque classe (soit  $C_i = 399 \times \frac{1}{3}$ ).

Variétés	commune	saumonée	arc-en-ciel
Effectifs $N_i$	145	118	136
Effectifs $C_i$	133	133	133

TABLE 3.2 – Effectifs théoriques de chaque variété de truite.

On obtient :

$$D^2 = \frac{(145 - 133)^2}{133} + \frac{(118 - 133)^2}{133} + \frac{(136 - 133)^2}{133} \simeq 2.84.$$

La valeur théorique lue dans la table du  $\chi^2$  au risque de 5% avec  $k = 3 - 1 = 2$  degrés de liberté vaut 5.99. On ne peut pas rejeter l'hypothèse que son bassin contient autant de truites de chaque variété car ( $D^2 < k_\alpha$ ).

## 3.4 Test de comparaison pour deux échantillons

### 3.4.1 Test de Wilcoxon

Le test de Wilcoxon est un test de comparaison de moyennes de deux échantillons indépendants ou appariés. Il est une alternative non paramétrique au test de Student [1].

## Cas d'échantillons indépendants

Soit  $(X_1, X_2, \dots, X_{n_1})$  un  $n_1$  échantillon de  $X$  et  $(Y_1, Y_2, \dots, Y_{n_2})$  un  $n_2$  échantillon de  $Y$ . Pour comparer les distributions de  $X$  et  $Y$ , on va tester l'hypothèses suivantes :

- $H_0$  : les deux échantillons appartiennent à la même population.
- $H_1$  : les deux échantillons différents.

Ce test est basé sur le rangement en ordre croissant des observations  $X$  et  $Y$ . La statistique correspondante est la somme des rangs de  $X$  (en générale celui de plus petite taille, on pose  $n_1 =$  taille du plus petit échantillon) :

$$W_{n_1, n_2} = \sum_{i=1}^{n_1} R(X_i),$$

où  $R(X_i)$  est le rang de la  $i^{\text{ème}}$  observation de  $X$  dans l'échantillon complet.

On a :

$$E(W_{n_1, n_2}) = \frac{n_1(n_1 + n_2 + 1)}{2} \quad \text{et} \quad V(W_{n_1, n_2}) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

**Critère de décision :**

Le rejet ou non de l'hypothèse nulle  $H_0$ , au niveau  $\alpha$ , est basé sur la valeur de  $p$  obtenue à partir de la statistique  $W_{n_1, n_2}$ .

Pour  $n_1 + n_2$  assez grand, la statistique de Wilcoxon est définie par :

$$Z = \frac{W_{n_1, n_2} - E(W_{n_1, n_2})}{\sqrt{V(W_{n_1, n_2})}} \rightsquigarrow \mathbb{N}(0, 1), \text{ quand } n \rightarrow \infty$$

La région critique du test au niveau de signification  $\alpha$  est  $|Z| > k_{1-\frac{\alpha}{2}}$ , où  $k_{1-\frac{\alpha}{2}}$  est le quantile d'ordre  $1 - \frac{\alpha}{2}$  de loi normale centrée réduite.

**Exemple 3.3.**

La taille des feuilles de ronces ont été mesuré pour savoir s'il y a une différence entre la taille des feuilles des plants qui poussent en plein soleil et des qui poussent à l'ombre. Les résultats sont les suivants (largeur de feuilles en cm) :

Soleil	6	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

On réordonne les 16 observations par ordre croissant. Les résultats des plants au soleil sont soulignés :

*Observations* : 4.1 4.5 4.8 5.1 5.1 5.3 5.5 5.5 5.5 : 5.5 5.9 6 6.3 6.5 6.8 7.2.  
*Rangs* : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16.

La somme des rangs des individus du Soleil est :

$$W_X = 1 + 2 + 3 + 4.5 + 4.5 + 6 + 8.5 + 12 = 41.5.$$

Si  $H_0$  était vraie :

$$E(W_X) = \frac{8(8 + 8 + 1)}{2} = 68; \text{Var}(W_X) = \frac{8 \times 8(8 + 8 + 1)}{12} = 90.66 = (9.525)^2.$$

comme  $\frac{41.5 - 68}{9.525} = -2.7821$ , et  $|-2.7821| \geq k_{0.975} = 1.96$  ( $\alpha = 5\%$ ), alors, on rejette  $H_0$ .

La différence entre la taille des feuilles à l'ombre et au soleil est donc significative au risque ( $\alpha = 5\%$ ).

### Cas d'échantillons appariés (Test des rangs signés de Wilcoxon)

Les échantillons appariés sont des échantillons identiques, c'est à dire des échantillons composés d'individus possédant les mêmes caractéristiques.

Pour calculer la statistique de décision de ce test, on range par ordre croissant les valeurs absolues des différences de deux variables statistiques  $X$  et  $Y$  :  $|Z_1| = |X_1 - Y_1|, \dots, |Z_n| = |X_n - Y_n|$ , on note  $r_i$  le rang de  $|Z_i|$  et on désigne par  $T^+$  la somme des rangs des différences positives et  $T^-$  la somme des rangs des différences négatives [5].

La région critique du test, au risque  $\alpha$ , s'écrit :

$$T \leq T_\alpha(n).$$

où  $T_\alpha(n)$  est la valeur tabulée.

#### Remarque 3.1.

1. On a :

$$T^+ + T^- = \frac{n(n+1)}{2}.$$

2. Les variables  $T^+$  et  $T^-$  suivent la même loi de probabilité.



**Proposition 3.1.**

On a :

$$E(T^+) = \frac{n(n+1)}{4} \quad \text{et} \quad V(T^+) = \frac{n(n+1)(2n+1)}{24}.$$

De plus :

$$Z = \frac{T^+ - E(T^+)}{\sqrt{V(T^+)}} \rightsquigarrow \mathbb{N}(0, 1) \quad \text{quand } n \rightarrow \infty.$$

En pratique, on applique ce résultat lorsque  $n > 30$ , si  $n \leq 30$  la statistique du test ne suit pas une loi usuelle.

**3.4.2 Test de signes**

Le test de signes est une méthode statistique qui permet de tester les différences constantes entre les paires d'observations issues de deux variables statistiques  $X$  et  $Y$ . On utilise les signes (+) et (-) au lieu de données quantitatives.

L'hypothèse nulle peut s'écrire  $p_- = p_+ = \frac{1}{2}$  avec  $p_+$  la probabilité d'observer une différence positive et  $p_-$  celle d'observer une différence négative [5].

On pose :

$$S^+ = \sum_{i=1}^n \mathbb{1}_{]0, +\infty[}(Y_i - X_i) \quad \text{et} \quad S^- = \sum_{i=1}^n \mathbb{1}_{] -\infty, 0[}(Y_i - X_i)$$

Le nombre de différences positives (ou négatives) est une variable aléatoire binomiale  $\mathcal{N} \rightsquigarrow \mathcal{B}(n, \frac{1}{2})$  puisque ce test permet de compter le nombre observé de signes.

Il est possible d'obtenir directement la probabilité critique ( $p$ -value) du test bilatéral en calculant la quantité

$$p = 2P \left[ \mathcal{B} \left( n, \frac{1}{2} \right) \geq \max(S, n - S) \right].$$

La probabilité cumulée d'une loi binomiale est

$$\begin{aligned} P[\mathcal{B}(n, \pi) \geq s] &= \sum_{j=s}^n \binom{n}{j} \pi^j (1 - \pi)^{n-j} \\ &= \sum_{j=s}^n \frac{n!}{j!(n-j)!} \pi^j (1 - \pi)^{n-j}. \end{aligned}$$

### 3.4.3 Test d'indépendance

C'est un test entre deux variables aléatoires basées sur la statistique de décision de la loi de Khi-deux.

Soient  $X$  et  $Y$  deux variables aléatoires définies sur la même population  $\Omega$  mesurant deux caractères [15, 17].

$X : \Omega \rightarrow M, M$  étant un ensemble de modalités divisé en  $k$  classes  $C_1, C_2, \dots, C_k$ .

$Y : \Omega \rightarrow M', M'$  étant un ensemble de modalités divisé en  $l$  classes  $D_1, D_2, \dots, D_l$ .

On veut savoir s'il existe une liaison significative entre  $X$  et  $Y$ .

- On veut tester l'hypothèse

$H_0 : X$  et  $Y$  sont indépendantes vs  $H_1 : X$  et  $Y$  ne sont pas indépendantes

- Il s'agit de déterminer la variable de décision.

Pour cela, on dispose d'un échantillon de  $X$  et d'un échantillon de  $Y$  dont les résultats peuvent se mettre sous la forme du tableau de contingence suivant :

	$D_1$	$D_2$	$\dots$	$D_l$	Effectifs des $C_i$
$C_1$	$n_{11}$	$n_{12}$		$n_{1l}$	$n_{1.}$
$C_2$	$n_{21}$	$n_{22}$		$n_{2l}$	$n_{2.}$
$C_k$	$n_{k1}$	$n_{k2}$		$n_{kl}$	$n_{k.}$
Effectif des $D_j$	$n_{.1}$	$n_{.2}$		$n_{.l}$	$n$

avec  $n_{.j} = \sum_{i=1}^k n_{ij}$  et  $n_{i.} = \sum_{j=1}^l n_{ij}$  et  $n = \sum_{i=1}^k \sum_{j=1}^l n_{ij}$ .

Si  $H_0$  est vraie, alors  $P((X \in C_i) \cap (Y \in D_j)) = P(X \in C_i) \times P(Y \in D_j), \forall i, j$ .

Comme on ne connaît pas les probabilités théoriques de  $X$  et  $Y$ , on peut traduire cette propriété par  $f_{ij} = f_{i.} \times f_{.j}, \forall i, j$  avec  $f_{ij} = \frac{n_{ij}}{n}, f_{i.} = \frac{n_{i.}}{n}$  et  $f_{.j} = \frac{n_{.j}}{n}$

$$\begin{aligned}
 \text{On a } X \text{ et } Y \text{ indépendantes} &\iff f_{ij} = f_{i.} \times f_{.j} \\
 &\iff \frac{n_{ij}}{n} = \frac{n_{i.}}{n} \times \frac{n_{.j}}{n} \\
 &\iff n_{ij} = \frac{n_{i.} \times n_{.j}}{n}
 \end{aligned}$$

Posons  $T_{ij} = \frac{n_{i.} \times n_{.j}}{n}$ , la statistique de décision est donnée par :

$$Z = \sum_{i=1}^k \sum_{j=1}^l \frac{(n_{ij} - T_{ij})^2}{T_{ij}} \rightsquigarrow \chi_{(k-1)(l-1)}^2$$

La région critique du test est :  $RC = [Z > k_\alpha]$  avec  $P_{H_0}(RC) = \alpha$  un seuil fixé.  
 $k_\alpha$  est lu sur la table de Khi-deux.

**Exemple 3.4.**

Un chercheur veut vérifier si deux universités ont un même barème pour l'attribution des côtes. Pour ce faire il choisit un échantillon de 21000 étudiants provenant des deux universités et il regarde les côtes attribuées aux étudiants de 2001 :

Côte	A	B	C	D	E
Université 1	605	1400	1789	300	70
Université 2	2014	4178	8032	2005	607

En fait, on cherche à vérifier si la répartition des côtes est dépendantes des universités, c'est à dire si les variables " université" et " côte" sont des variables aléatoires indépendantes au niveau 5%.

Les hypothèses statistiques sont

$H_0$  : les variables aléatoires sont indépendantes.

$H_1$  : les variables aléatoires ne sont pas indépendantes .

On obtient le tableau de contingence suivant :

Côte	A	B	C	D	E	Total
Université 1	605	1400	1789	300	70	4164
Université 2	2014	4178	8032	2005	607	16836
Total	2619	5578	9821	2305	677	21000

Les effectifs théoriques sont donnés par :

Côte	A	B	C	D	E	Total
Université 1	519.310	1106.038	1947.364	457.049	134.239	4164
Université 2	2099.690	4471.962	7873.636	1847.951	542.761	16836
Total	2619	5578	9821	2305	677	21000

La statistique observée est

$$\sum_{i=1}^2 \sum_{j=1}^5 \frac{(n_{ij} - T_{ij})^2}{T_{ij}} = \frac{(605 - 519.310)^2}{519.310} + \dots + \frac{(607 - 542.761)^2}{542.761} = 236.808$$

On a :  $\chi_{(k-1)(l-1); \alpha}^2 = \chi_{4; 0.05}^2 = 9.4877$ , d'où :

$$236.808 > \chi_{4; 0.05}^2 = 9.4877$$

Donc on rejette  $H_0$  au niveau 5% et on peut dire qu'il y a dépendance. C'est à dire que la répartition des côtes dépend de l'université.

# Chapitre 4

## Application des tests statistiques avec R

### Introduction

Nous avons vu dans les chapitres précédents les différents tests statistiques paramétriques et non paramétriques. Dans ce chapitre, nous allons appliquer quelques uns de ces tests dans les domaines de l'expérimentation scientifique et à la vie quotidienne. Les calculs sont fait à l'aide du logiciel de statistique R.

### 4.1 Répartition des pathologies selon l'année

L'étude porte sur certaines pathologies dans le service chirurgie de l'hôpital d'Amizour [3]. On voudrait savoir si le nombre de maladies varie d'une année à l'autre ou bien c'est le même pour les années 2013 et 2014.

Les différentes pathologies concernées par cette étude sont :

- ccl : Lithiase vésiculaire
- aa : Appendicite aigue
- gmm : Goitre
- even : Eventration
- neo : Cancer
- fa : Fissure Anale
- hig : Hernie
- etb : Etrapie vésiculaire

Le tableau suivant donne le nombre de malades pour les années 2013 et 2014 respectivement.

Pathologie   Année	2013 (X)	2014 (Y)
ccl	349	307
aa	182	187
gmm	193	143
even	26	46
neo	41	20
fa	38	56
hig	152	162
etb	26	24

TABLE 4.1 – Variation des pathologies en fonction du temps.

On teste au seuil  $\alpha = 0.05$ ,  $H_0$  : le nombre de malades pour l'année 2013 est supérieur à celui de l'année 2014 vs  $H_1$  : le nombre de malades pour l'année 2013 est inférieur à celui de l'année 2014.

On est en présence d'échantillons appariés. On applique dans ce cas le test de Wilcoxon pour deux échantillons appariés.

**Remarque 4.1.** On peut aussi appliquer le test du signe pour deux échantillons appariés.

Tout d'abord, calculons les différences des nombres de malades de 2013 et 2014

$Z = X - Y$ , on obtient les résultats suivants :

X	Y	X-Y	$ Z  =  X - Y $	rangs de $ X - Y $	rangs signé
1	349	307	42	45	7
2	182	187	-5	5	2
3	193	143	50	50	8
4	26	46	-20	20	5
5	41	20	21	21	6
6	38	56	-18	18	4
7	152	162	-10	10	3
8	26	24	2	2	1

TABLE 4.2 – Valeurs des rangs signés

On résume ces résultats dans le tableau suivant :

	nombre	somme des rangs
rangs négatifs	4	14
rangs positifs	4	24
égalité	0	
Total non nul	8	

TABLE 4.3 – Récapitulatif des résultats

La statistique des rangs positif est  $T^+ = 7 + 8 + 6 + 1 = 24$  et la somme des rangs négatifs est  $T^- = 2 + 5 + 4 + 3 = 14$

Le code R de cet exemple est le suivant :

```
> X <- c(349, 182, 193, 26, 41, 38, 152, 26)
> Y <- c(307, 187, 143, 46, 20, 56, 162, 24)
> wilcox.test(X, Y, paired = T, correct = FALSE, exact = FALSE)
```

*wilcoxonsignedranktest*

data : X and Y

$V = 22, p - value = 0.5754$

On note que  $p - value = 0.5754 > \alpha = 0.05$ , alors on décide de ne pas rejeter l'hypothèse  $H_0$ .

## 4.2 Répartition des pathologies selon le sexe

Reprenons le problème précédent [3]. Cette fois-ci, on souhaite comparer la distribution des pathologies entre les femmes et les hommes pendant l'année 2015 au niveau  $\alpha = 0.05$ .

On teste alors,  $H_0$  : le nombre de malades femmes est inférieur au nombre de malades hommes pendant l'année 2015 vs  $H_1$  : le nombre de malades femmes est supérieur au nombre de malades hommes pendant l'année 2015.

Pour cela, on applique le test du signe pour échantillons appariés.



maladie	femme	homme	signe(écart)
ccl	105	269	-
aa	113	146	-
gmm	81	114	-
even	19	32	-
néo	2	16	-
fa	28	24	+
hig	152	67	+
etb	35	22	+

TABLE 4.4 – Répartition des données pour l'année 2015

Le tableau suivant présente les données pour l'année 2015. On a donc 5 différences négatives et 3 différences positives.

Le code de  $\mathcal{R}$  pour ce problème est le suivant :

```
> Femme <- -c(269, 146, 114, 32, 16, 24, 67, 24)
> Homme <- -c(105, 113, 81, 19, 2, 28, 152, 35)
> diff <- -c(Femme, Homme)
> nmoins <- -sum(diff < 0)
> nplus <- -sum(diff > 0)
> binom.test(min(nplus, nmoins), nplus + nmoins) p-value = 3.052
```

On a  $p\text{-value} = 3.052 > 0.05 = \alpha$ , ce qui nous conduit à ne pas rejeter l'hypothèse  $H_0$ .

### 4.3 Poids du cerveau selon le sexe

On veut savoir si le poids du cerveau (en grammes) est différent selon le sexe, alors les poids des cerveaux de 10 hommes et 10 femmes sont données dans le tableau ci-dessous : On applique à ce problème le test de student, puis

Hommes ( $X$ )	1393	1431	1325	1455	1416	1365	1248	1258	1359	1381
Femmes ( $Y$ )	1212	1229	1197	1134	1204	1252	1120	1155	1305	1045

TABLE 4.5 – Le poids du cerveau de 10 hommes et 10 femmes.

le test

## Test de Student (comparaison de deux moyennes)

On suppose que  $\sigma_X = \sigma_Y$  et on test, au niveau  $\alpha = 0.05$  :

$$H_0 : \mu_X = \mu_Y \text{ vs } H_1 : \mu_X \neq \mu_Y.$$

on utilise la fonction `t.test()`.

Le logiciel R nous donne les résultats suivants :

```
> Hommes = c(1393, 1431, 1325, 1455, 1416, 1365, 1248, 1258, 1359, 1381)
> Femmes = c(1212, 1229, 1197, 1134, 1204, 1252, 1120, 1155, 1305, 1045)
> t.test(Hommes, Femmes, var.eq = T)
```

*TwoSamplet - test*

data : Hommes and Femmes

t= 5.5371, df=18, p-value= 2.948e - 05

alternative hypothesis : true difference in means is not equal to 0.

95 percent confidence interval :

110.0892 244.7108

sample estimates :

mean of x mean of y

1363.1...1185.7

On a  $p - \text{value} = 2.948e - 05 < \alpha = 0.05$ , ce nous conduit à rejeter l'hypothèse d'égalité des poids des cerveaux des deux sexes  $H_0$ .

## Test de Wilcoxon, cas d'échantillons indépendants

On peut appliquer aussi, au problème précédent, le test non paramétrique de Wilcoxon dans le cas d'échantillons indépendants. On test au niveau  $\alpha = 0.05$  :

- $H_0$  : Les deux échantillons appartiennent à la même population,
- $H_1$  : Les deux échantillons différents.

Le logiciel R nous donne les résultats suivants

```
> Hommes = c(1393, 1431, 1325, 1455, 1416, 1365, 1248, 1258, 1359, 1381)
> Femmes = c(1212, 1229, 1197, 1134, 1208, 1252, 1120, 1155, 1305, 1045)
> wilcox.test(Hommes, Femmes)
```

*wilcoxonranksumtest*

data : Hommes and Femmes.

$w = 97, p - \text{value} = 7.578e - 05$ .

alternative hypothesis : true location shift is not equal to 0.

On a  $p - value = 7.578e - 05 < \alpha = 0.05$ , ce nous conduit à rejeter l'hypothèse d'égalité des poids des cerveaux des deux sexes  $H_0$ .

## 4.4 Efficacité du vaccin antigripale

Lord d'une épidémie, la grippe touche 30% d'une population [24]. Pour tester l'efficacité d'un vaccin antigripale, on vaccine 300 personnes. A la fin de la saison grippale, on dénombre 50 ont atteintes par la grippe.

On désire tester l'efficacité du vaccin au niveau de signification  $\alpha = 0.05$ . on teste alors :

$$H_0 : p = 0.3 \text{ contre } H_1 : p < 0.3.$$

Comme  $\frac{F-0.3}{0.026} = -5.04 < -1.64 = -q_{1-\alpha}$  (lue dans la table de gauss), donc on rejette l'hypothèse  $H_0$ , c'est à dire le vaccin est efficace.

Ce résultat est confirmé par le test adéquat avec le logiciel R. En effet :

```
prop.test(50,300,p=0.3, alternative = "less", correc= FALSE)
```

```
data : 50 out of 300, null probability 0.3
```

```
X-squared = 25.397, df=1, p-value= 2.333e-07.
```

On a  $p - value = 2.333 < \alpha = 0.05$ , l'hypothèse  $H_0$  est rejetée au risque  $\alpha$  fixé à 5%, cela signifié que le vaccin est efficace.

## Conclusion

Les applications abordées dans ce chapitre nous permettent de voir l'intérêt des tests statistiques dans la prise de décision dans un domaine très sensible qui est la médecine. Ceci est valable dans tous les domaines de la vie quotidienne. Nous mesurons aussi l'intérêt d'appliquer des logiciels de statistique pour le traitement des données, dans notre cas le logiciel R.

# Conclusion Générale

Ce travail est consacré à l'étude des tests statistiques. Mais avant d'aborder ces aspects, nous avons rappelé un certain nombre de notions sur les variables aléatoires et les lois de probabilités les plus utilisées dans les tests statistiques.

Nous avons traité les deux grands types de tests. Notons que les tests paramétriques nécessitent le respect des hypothèses de base faites lors de leur conception, et application lorsque les populations étudiées sont de distributions normales ou lorsque les échantillons sont de grandes tailles.

Néanmoins, lorsque les conditions de ces tests ne sont pas vérifiées, il convient d'utiliser d'autres tests qui permettent de s'affranchir de ces conditions. Il s'agit des tests non paramétriques qui sont utilisés pour les petit échantillons ou lorsque les distributions des variables ne sont pas gaussienne dans les échantillons.

Quelques tests paramétriques et non paramétriques ont été abordés, ainsi que des applications sur des problèmes réels ont été effectuées à l'aide du logiciel R.

# Bibliographie

- [1] F. ATMANI. *Sur quelques tests non paramétriques et applications*. Mémoire de master, Département des Mathématiques, Université de Biskra, Algérie, Septembre 2020.
- [2] F. AUBIN and R. SIGNORET. *L'essentiel des Probabilités et Statistiques*. Ellipses, 2016.
- [3] H. BENHAMMA and L. DJERRAHI. *Etude des pathologies de l'hôpital d'Amizour*. Mémoire de licence, Département de Mathématiques, Université de Béjaia, Algérie, 2016.
- [4] A. BOUBECHE. *Tests d'Hypothèse Statistique*. Mémoire de master, Département de Mathématiques, Université de Biskra, Algérie, Septembre 2020.
- [5] A. BOUDEN. *Tests de comparaison de deux distributions*. Mémoire de master, Département de Mathématiques, Université de Biskra, Algérie, Septembre 2020.
- [6] N. BOUKHARI and I. ZOUAOUI. *Les méthodes statistiques et applications*. Mémoire master, Département des Mathématiques, Université de Bordj Bou Arréridj, Algérie, Juillet 2021.
- [7] S. DEJEAN. *Quelques Mots sur les Tests Non Paramétriques*. UTIn Capitole, Institut de Mathématiques de Toulouse, France, 13 Mai 2014.
- [8] Z.S. DELHOUM. *Cours et exercices corrigés en probabilités*. Ecole Supérieure d'Economie, Oran, Algérie, 2020.
- [9] J. DELON. *Probabilités continues*. Cours de master, Université PARIS DESDARTES.
- [10] P. DUSART. *Cours de Probabilités*. Licence 2-S3 SI- MASS, 2013.
- [11] M. GENIN. *Tests non paramétriques*. EA2964- Santé Publique. Université de Lille2, France.

- 
- [12] H. GHAROUT. *Transformation ponctuelle de type polynomiale de dimension trois*. Thèse de doctorat, Département des Mathématiques, Université de Bejaia, Algérie, Décembre 2018.
- [13] H. GHAROUT. *Cours de Biostatistique au profil des étudiants des tronc communs sciences de la nature et de la vie*. Editions universitaires Européennes, 2022.
- [14] H. GHAROUT, N. Akroune, A.K. Taha, and D. Fournier-Prunaret. Chaotic dynamics of a three-dimensional endomorphism. *Journal of Siberian Federal University Mathematics and Physics*, 12(1) :36–50, 2019.
- [15] L. GRAMMONT. *Cours de statistiques inférentielles, Licence d'économie et de gestion*. PhD thesis, 19 Septembre 2003.
- [16] L. JEAN. *AIDE MEMOIRE STATISTIQUE*. CISIA. CERESTA Editeur, 1 Avenue Herbillon 94160 Saint-Mandé (France), Octobre 1995.
- [17] L.HOUDA. *Module 7 lois de probabilité*. Département des Mathématiques et d'informatique, Université du Québec à Trois-Rivières, 2014.
- [18] S. MARKOVITCH. *Probabilités une approche expérimentale*. Hachette Livre, 2002.
- [19] A. MEOT. *Introduction aux statistiques inférentielles*. De Boek et Larcier, 2003.
- [20] R. RAKOTOMALALA. *Comparaison de populations, Tests non paramétriques*. Université Lumière Lyon 2, France, 2013.
- [21] P. RIBEREAU. *Cours de Statistiques Inférentielles*.
- [22] J. J. RUCH. *Statistique : Tests D'hypothèses (Préparation à l'agrégation)*. Bordeaux1, 2012-2013.
- [23] S. SEGNI. *Statistique et Probabilité*. Mémoire de master, Département des Mathématiques et d'informatique et des Sciences de la Matière, Université de Guelma, Algérie, 2015.
- [24] W. TORCHI. *Tests Statistiques et ses Applications*. Mémoire de master, Département de Mathématiques, Université de Biskra, Université de Biskra, Algérie, 2020.
- [25] B. TRIBOUT. *Statistique pour économistes et gestionnaires*. Université Robert Schuman, strasbourg, 2013.

## Résumé

L'objectif assigné à ce travail est l'étude des tests statistiques. Ces derniers sont utilisés dans presque tous les secteurs de l'activité humaine.

Les principaux tests paramétriques et non paramétriques ont été traités. Des exemples d'illustration des méthodes étudiées complètent l'étude théorique.

A l'aide du logiciel d'analyse statistique R, nous proposons quelques applications permettant d'illustrer les résultats théoriques

Mots clés : Test paramétrique, test non paramétrique, erreur de 1<sup>ère</sup> espèce, erreur de 2<sup>ème</sup> espèce, région critique, ...

## Abstract

The objective assigned to this work thesis is the study of statistical tests. These tests are used in almost all sectors of human activity.

The main parametric tests were discussed. Examples of illustration of the studied methods complete the theoretical study.

Using the statistical analysis software R, we propose some applications to illustrate the theoretical results.

Key words : Parametric test, non parametric test, error of 1<sup>ère</sup> kind, error of 2<sup>ème</sup> kind, critical region, ...