République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ MOHAMED KHIDER, BISKRA

Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie

DÉPARTEMENT DE MATHÉMATIQUES

Mémoire présenté en vue de l'obtention du Diplôme :

MASTER en Mathématiques

Option : statistique

Par

Lobna Rehab

Titre :

Sur le test statistique d'ajustement

Membres du Comité d'Examen :

Pr.	**Benatia Fatah**	UMKB	Président
Pr.	**Brahimi Brahim**	UMKB	Encadreur
Dr.	**Abdelli Jihane**	UMKB	Examinateur (rice)

Juin 2024

Dédicace

Je dédie ce humble travail

À mes chers parents,

Merci pour tout l'amour, le soutien et les grands sacrifices que vous m'avez donnés.

À mes chères sœurs et mon frère,

Merci pour votre amour et votre soutien constants.

À mes chers professeurs,

Merci pour tout ce que vous m'avez appris en matière de connaissances, de sagesse et de valeurs morales.

REMERCIEMENTS

Louange à Allah, Seigneur des univers ! Je Lui rends grâce pour la science et la foi qu'Il m'a accordées. Je Lui demande de me faire profiter de ce que j'ai appris et de rendre ma connaissance bénéfique pour moi et pour les autres.

Obtenant mon diplôme, je tiens à exprimer ma profonde gratitude à tous ceux qui ont contribué à mon parcours académique.

À mes parents, source d'affection et d'amour, merci pour votre foi en mes capacités et votre soutien constant.

À mes chers professeurs, soleil de la connaissance qui a illuminé mon chemin, merci pour votre patience et votre dévouement, et en particulier au professeur **Brahimi Brahim** qui a été pour moi un phare qui a guidé mes pas dans mon voyage de la connaissance, et le **Pr :Benatia Fatah** Il a présidé le jury de discussion avec habileté et sagesse et **Dr** :**Abdelli Jihane** elle a contribué à la discussion de ma thèse et l'a enrichie de ses questions précieuses.

À mes chères sœurs, **Amina, Nour alhouda** et mon frère **Rami,** merci pour tout l'amour, le soutien et l'encouragement que vous me donnez. Vous êtes les étoiles de ma vie, la lumière de mon chemin et mon soutien à chaque étape.

À mes chers camarades, **Wafa Merzougui, Hasna, Masouda, Iman ;** compagnons de route et partenaires de réussite, merci d'avoir été présents dans ma vie.

À tous ceux qui m'ont soutenu et encouragé, merci pour vos paroles d'encouragement et de soutien.

Enfin, je tiens à me remercier pour ma foi en mes capacités, ma patience et ma persévérance dans la réalisation de mon objectif.

Avec toute ma gratitude, ma reconnaissance et mon appréciation.

Table des matières

Table des figures

Liste des tableaux

Introduction

Dans cette thèse, nous aborderons une problématique d'une grande importance pratique : celle de déterminer si un ensemble de mesures indépendantes d'un phénomène aléatoire, dont la loi de probabilité est inconnue, peut être associé à une loi spécifique, telle que la loi normale ou loi exponentielle par exemple. Pour ce faire, nous utiliserons des méthodes d'ajustement visant à comparer ces mesures à une loi théorique choisie.

Ce mémoire est composé de trois chapitres.

Chapitre 1 : Dans ce chapitre, quelques notions générales sur les lois de probabilité sont présentées, telles que les variables aléatoires, la fonction de répartition, la fonction de répartition empirique, les lois de probabilité discrètes et continues.

Chapitre 2 : Ce chapitre est dédié aux test statistique d'ajustement. Tout d'abord, nous aborderons les tests statistiques de manière générale, en discutant de leurs définitions telles que les hypothèses nulle et alternative, la p-valeur, les erreurs et les risques, la puissance, ainsi que la région de rejet et région critique. Ensuite, nous introduirons le test statistique d'ajustement et généralité sur le test d'ajustement , et quelque test comme le test du Khi-deux, de Kolmogrov-Smirnov, et lilliefors,....

Chapitre 3 : Ce chapitre est application sur R.

Chapitre 1

Les lois de probabilité

1.1 Lois de probabilité

Les lois de probabilités sont des règles mathématiques qui décrivent le comportement des variables aléatoires. Elles permettent de quantifier l'incertitude associée à des évènements aléatoires. Les principales lois de probabilités comprennent la loi de probabilité uniforme, la loi binomiale, la loi de poisson et la loi normale(ou gaussienne)...

Chacune de ces lois a ses propres caractéristiques et est utilisée dans différents contextes pour modéliser des phénomènes aléatoires.

1.1.1 Variables aléatoires réelles

Généralités sur les Variables aléatoires

Les variables aléatoires sont des concepts fondamentaux en probabilité et en statistiques. Elles représentent des quantités qui prennent des valeurs au hasard en fonction des résultats d'une expérience aléatoire. On distingue deux types de variables aléatoires, discrètes : qui prennent des valeurs dénombrables, et continues :

qui prennent des valeurs dans un intervalle de nombres réels.

Les variables aléatoires sont souvent associées à des fonctions de probabilité ou à des densités de probabilité, qui décrivent la probabilité d'obtenir chaque valeur possible de la variable aléatoire. Elles sont largement utilisées dans de nombreux domaines, tels que la finance, la physique, la biologie et bien d'autres, pour modéliser et analyser des phénomènes aléatoires.

Fonction de répartition

La fonction de répartition est utilisée pour définir de façon unifiée la loi de probabilité d'une Variables aléatoire (v.a) qu'elle soit discrète ou continue. Si cette fonction est connue, il est possible de calculer la probabilité de toute intervalle et donc, en pratique, de tout événement.

Soit X une v.a définie sur un espace de probabilités (Ω, Γ, P).

Définition 1.1.1 *On appelle fonction de répartition de X, que l'on note F, la fonction définie de $\mathbb{R}$ dans $[0;1]$ par :*

$$F(x) := P(X \leq x), x \in \mathbb{R}.$$

La valeur prise par la fonction de répartition au poit x est donc la probabilité de l'événement $]-\infty, x]$.

Propriété 1.1.1 *Les propriétés principales de la fonction de répartition sont les suivantes :*

- F est non décroissante.
- F continue à droite en tout point x de $\mathbb{R}$.
- $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$.

Remarque 1.1.1 *Les résultats suivants sont des conséquences directes de la définition.*

1. Si X est discrète de valeurs $x_1, x_2, ..., x_n$; alors :

$$F(x) = \sum_{x_k \leq x} P(X = x_k).$$

2. Si X est continue de densité f ; alors :

$$F(x) = \int_{-\infty}^{x} f(t)\, dt$$

dans ce cas, la probabilité de tout intervalle réel de bornes a et b; avec $a \leq b$ est égale à $F(b) - F(a)$

Nous utiliserons principalement la fonction de survie associée à une variable aléatoire et définie par :

$$\overline{F}(x) = P(X > x) = 1 - F(x)$$

Fonction de répartition empirique

Définition 1.1.2 *La fonction de répartition empirique F_n associée à un échantillon $(X_1, X_2, ..., X_n)$ est définie par :*

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \coprod_{(x_i \leq x)}, x \in \mathbb{R}.$$

C'est la proportion des éléments de l'échantillon qui sont inférieurs ou égaux à x. En d'autres termes, la fonction de répartition empirique est la moyenne empirique des fonctions d'indicatrices des événements $(X_i \leq x)$.

Sa représentation par les statistiques d'ordre est donnée par :

$$F_n(x) = \begin{cases} 0 & \text{si} \quad x_{(1)} \\ \frac{i}{n} & \text{si} \quad x_{(i)} \leq x < x_{(i+1)}, i = 1, 2, ..., n-1, \\ 1 & \text{si} \quad x_{(n).} \end{cases}$$

F_n est une fonction en escaliers qui fait des sauts de hauteur $1/n$ en chaque point de l'échantillon.

1.1.2 lois de probabilité discètes

Loi de Bernoulli $\mathcal{B}(p)$

Soit une expérience dont le résultat est aléatoire et soit A un évènement défini sur cette expérience. Soit X la variable aléatoire prenant la valeur 1 quand A est réalisé et 0 quand $\bar{A}$ est réalisé. On dit que X est une variable aléatoire de Bernoulli s'il existe p et q dans $\mathbb{R}$ vérifiant :

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

de moyenne

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} p_i x_i = P(X = 1) * 1 + P(X = 0) * 0 = p * 1 + q * 0 \\ &= p \end{aligned}$$

et de variance

$$\begin{aligned} Var(X) &= E(X^2) - (E(X))^2 = \sum_{i=1}^{n} p_i {x_i}^2 - E(X)^2 = p - p^2 \\ &= p\,(1 - p) \end{aligned}$$

Exemple 1.1.1 *On lance un pièce de monnaie on a :*

$$A : pile$$

$$\bar{A} : face$$

$$p = q = \frac{1}{2}$$

On fait l'experiance plusieurs fois on trouve :

$$A = 1, 6$$

$$\bar{A} = 2, 3, 4, 5$$

$$p = \frac{1}{3} \ et \ q = \frac{2}{3}$$

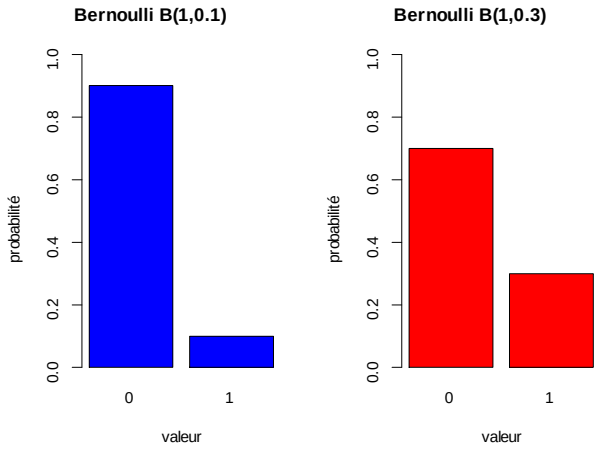

FIG. 1.1 – loi bernoulli

Loi de Binomiale $\mathcal{B}(n, p)$

La loi binomiale est une distribution de probabilité discrète qui décrit le nombre de succès k dans un nombre fixe n d'essais indépendants, où chaque essai a une probabilité de succès p et une probabilité d'échec $1 - p$.

La formule de la loi binomiale est donnée par :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}.$$

- n est le nombre total d'essais,
- k est le nombre de succès que nous voulons obtenir,
- p est la probabilité de succès pour chaque essai,
- $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ est le coefficient binomial qui calcule le nombre de façons de choisir k succès parmi n essais.

de moyenne

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} p_i x_i \\ &= np. \end{aligned}$$

et de variance

$$\begin{aligned} Var(X) = E(X^2) - (E(X))^2 &= \sum_{i=1}^{n} p_i {x_i}^2 - E(X)^2 \\ &= np(1-p). \end{aligned}$$

Exemple 1.1.2 *Supposons qu'un fabricant de composants électroniques affirme que* 10% *des composants qu'il produit sont défectueux. Un inspecteur décide de prélever un échantillon aléatoire de* 20 *composants de la production actuelle pour les tester et vérifier si l'affirmation du fabricant est correcte. Nous pouvons modéliser ce scénario à l'aide de la loi binomiale. Dans ce cas :*

- $n = 20$ (nombre total de composants dans l'échantillon),
- $p = 0.10$ (probabilité qu'un composant soit défectueux).

Maintenant, supposons que nous voulions calculer la probabilité que parmi les 20 composants testés, exactement 3 soient défectueux. En utilisant la formule de loi de binomiale :

$$P(X = 3) = \binom{20}{3} \times (0.10)^3 \times (1 - 0.10)^{20-3} \approx 0.001$$

$$E(X) = np = 20 * 0.10 = 2 \text{ et } Var(X) = np(1 - p) = 1.8$$

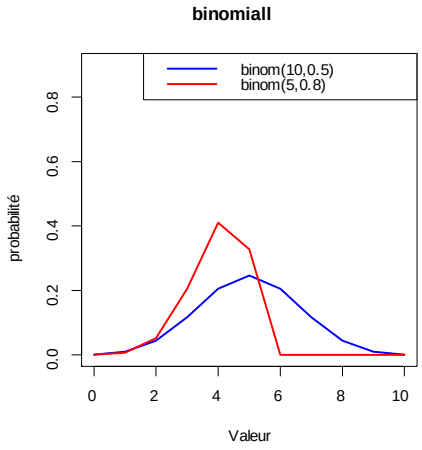

FIG. 1.2 – Loi de Binomiale

Loi de poisson $\mathcal{P}(\lambda)$

Soit X une variable aléatoire pouvant prendre toutes les valeurs entières $0, 1, ..., n$ vérifiant :

$$P(X = k) = e^{-\lambda}\lambda^k/k!$$

On dit que X suit une loi de poisson $P(\lambda)$ de paramètre λ de moyenne :

$$E(X) = \sum_{i=1}^{n} p_i x_i = \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} x_i$$
$$= \lambda$$

et de variance

$$Var(X) = E(X^2) - (E(X))^2 = \sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} x_i^2 - [\sum_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} x_i]^2$$
$$= \lambda$$

Exemple 1.1.3 *Supposons qu'un magasin reçoive en moyenne* 2 *clients par heure à sa caisse. Nous voulons modéliser le nombre de clients qui arrivent à la caisse en une heure en utilisant la loi de Poisson.*

Dans ce cas, la variable aléatoire X représente le nombre de clients arrivant à la caisse en une heure. La moyenne du nombre de clients par heure est $\lambda = 2$. Maintenant, supposons que nous voulions calculer la probabilité de recevoir exactement 1 *client à la caisse en une heure.*

En utilisant la formule de la loi de Poisson :

$$P(X = 1) = \frac{e^{-2}2^1}{1!} = \frac{2e^{-2}}{1} \approx 0.27067$$

Donc, la probabilité de recevoir exactement 1 *client à la caisse en une heure est d'environ* 0.27067 *ou* 27.07%.

$$E(X) = Var(X) = \lambda = 2$$

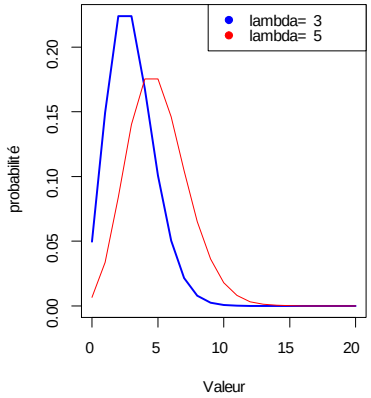

FIG. 1.3 – Loi de poisson

Loi géométrique $\mathcal{G}(p)$

Définition 1.1.3 *On dit d'une variable aléatoire X suit une loi géométrique de paramètre p, si ona :*

$$P(X = k) = pq^{k-1}$$

de moyenne

$$E(X) = \frac{1}{p}$$

et de variance

$$Var(X) = \frac{q}{p^2}$$

Exemple 1.1.4 *Dans une population de vanneaux, la probabilité de décès d'un oiseau au cours d'une année est constante et égale à $\frac{1}{3}$, quel est l'age moyen d'une oiseau ?*

La durée de vie X d'une oiseau suit une distribution géométrique. Et l'on peut écrire :

$$P(X = n) = pq^{n-1}$$

avec $q = \frac{2}{3}$ et $p = \frac{1}{3}$, ainsi la probabilité qu'un vanneau meure à deux ans est :

$$P(X = n) = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}$$

La durée moyenne de vie d'un vanneau est alors

$$E(X) = \frac{1}{p} = 3 \text{ ans}$$

1.1.3 lois de probabilité continues

Loi normale ou Gouss $\mathcal{N}(m, \sigma^2)$

Définition 1.1.4 *On appelée variable aléatoire normale ou gaussienne toute variable aléatoire absolument continue*

dont la densité de probabilité f est définie par :

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

m étant une constant réelle,σ une constante réelle strictement positive.

On utilise la notation suivante :

$$X \rightsquigarrow \mathcal{N}(m, \sigma^2)$$

La fonction f définit une densité. En effet :

$$\int_{-\infty}^{+\infty} f(x)dx = 1$$

de Fonction de répartition :

$$F(X) = P(X < a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

Moments :

L'espérance et la variance d'une variable normale sont respectivement données par :

$$E(X) = m \quad \text{et} \quad Var(X) = \sigma^2$$

Variable normale centrée réduite : Si $m = 0$ et $\sigma = 1$, et on note alore

$$X \rightsquigarrow \mathcal{N}(0, 1)$$

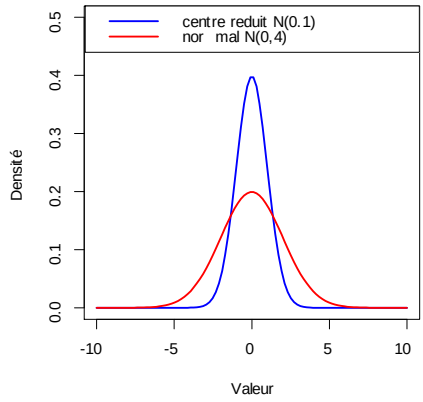

FIG. 1.4 – Loi normale

Loi exponentielle $\mathcal{E}(\lambda)$

Définition 1.1.5 *Une variable aléatoire X suit une loi expenentielle de paramètre λ ($\lambda \in \mathbb{R}^{+*}$).*

Si X est une variable aléatoire absolument continue dont la densité de probabilité est définie par :

$$f(x)\begin{cases} 0 & \text{si} \quad x<0 \\ \lambda\exp(-\lambda x) & \text{si} \quad x\geq 0 \end{cases}$$

On note alore $X \rightsquigarrow \mathcal{E}(\lambda)$

Sa fonction de répartition :

$$F(x)=\begin{cases} 0 & \text{si} \quad x<0 \\ 1-\lambda\exp(-\lambda x) & \text{si} \quad x\geq 0 \end{cases}$$

Moments :

Si $X \rightsquigarrow \mathcal{E}(\lambda)$:

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_{0}^{+\infty} x\lambda\exp(-\lambda x)dx=\frac{1}{\lambda}$$

à l'aide d'une intégration par parties :

Espérance et variance :

$$E\left(x\right)=1/\lambda \text{ et } Var(x)=1/\lambda^{2}$$

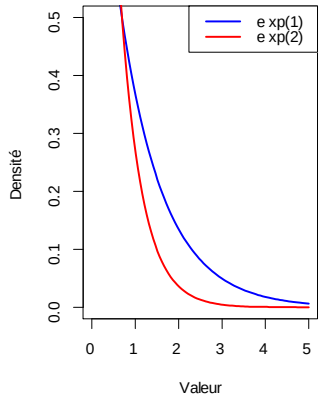

FIG. 1.5 – Loi exponentielle

Loi uniforme $\mathcal{U}([a,b])$

On dit que la loi de probabilité d'une variable aléatoire réele est uniforme sur un segment $[a;b]$, avec $0 \leq a < b$.

si sa probabilité f est définie par :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si pour} \quad x \in [a;b] \\ 0 & \text{si pour} \quad x < a \text{ ou } x > b \end{cases}$$

On note alore $X \rightsquigarrow \mathcal{U}([a;b])$.

Propriétés 1.1.1 *On a bien une densité de probabilité puisque :*

$f(x) \geq 0 \qquad \forall x \in \mathbb{R}$

f est continue sur $]-\infty;a[\cup]a;b[\cup]b;+\infty[$.

$$\int_{-\infty}^{\infty} f(t)dt = \int_{-\infty}^{a} f(t)dt + \int_{a}^{b} f(t)dt + \int_{b}^{\infty} f(t)dt = 0 + 1 + 0 = 1.$$

Fonction de répartition :

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt$$

On sait que :

donc si $X \rightsquigarrow \mathcal{U}([a;b])$:

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

Par intégration, on obtient :

$$E(x) = \frac{(b+a)}{2} \text{ et } Var(x) = \frac{(b-a)^2}{12}$$

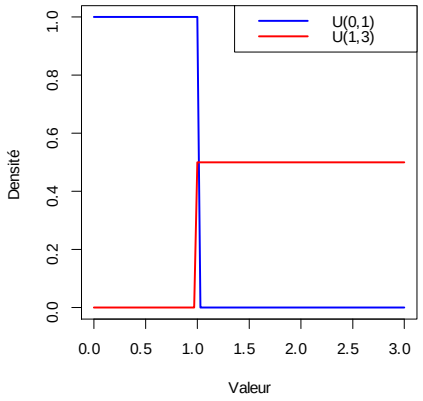

FIG. 1.6 – Loi uniforme

Loi de gamma $\Gamma(t, \lambda)$

Définition 1.1.6 *On dit que X suit la loi gamma de paramètres $t > 0$ et $\lambda > 0$, notée $\gamma(t, \lambda)$ ou $\Gamma(t, \lambda)$, si elle admet pour densité de probabilité est donné par :*

$$f(x) = \begin{cases} \dfrac{\lambda e^{-\lambda x}(\lambda x)^{t-1}}{\Gamma(t)} & \text{si} \qquad x \geq 0 \\ 0 & \text{sinon} \end{cases}$$

Γ est la fonction eulérienne définie par l'intégrale pour $t > 0$

$$\Gamma(t) = \int_0^{\infty} e^{-\gamma}\gamma^{t-1}d\gamma$$

Une telle variable aléatoire X admet alors une espérance et une variance donnés par :

$$E(X) = \frac{t}{\lambda} \text{ et } Var(X) = \frac{t}{\lambda^2}$$

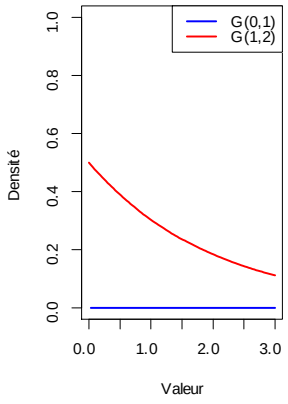

FIG. 1.7 – Loi de gamma

Loi de bêta $\mathcal{B}(\alpha, \beta)$

Cette loi propose un cadre pour représenter des mesures allant de 0 à 1, notamment pour des taux ou des proportions. Elle est caractérisée par une densité de probabilité :

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)}x^{\alpha}(1-x)^{\beta} & \text{si} \quad x \in]0;1[\\ 0 & \text{si} \quad x \notin]0;1[\end{cases}$$

avec $\alpha > -1$ et $\beta > -1$.

Pour $\alpha = \beta = 0$ on a la loi uniforme $\mathcal{U}[0,1]$. Pour α et β strictement positifs elle admet un mode en $x = \frac{\alpha}{(\alpha+\beta)}$.

Sachant que , pour tout $\alpha > -1$ et tout $\beta > -1$, ona :

$$\int_0^1 x^\alpha (1-x)^\beta dx = \frac{\Gamma(\alpha+1)\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}$$

on calcule aisément, pour $X \rightsquigarrow \mathcal{B}(\alpha, \beta)$

$$E(X) = \frac{\alpha+1}{\alpha+\beta+2}$$
$$Var(X) = \frac{(\alpha+1)(\beta+1)}{(\alpha+\beta+2)^2(\alpha+\beta+3)}$$

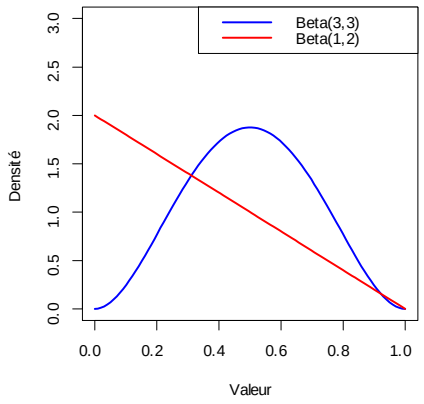

FIG. 1.8 – Loi de bêta

Chapitre 2

Test statistique d'ajustement

2.1 Qu'est-ce qu'un test statistique ?

Dans le domaine de la statistique, les tests statistiques sont utilisés pour analyser les données et tester les hypothèses statistiques. Ces tests se concentrent généralement sur la comparaison d'une ou plusieurs variables entre deux groupes ou plus. Cependant, ces tests supposent souvent que les données suivent une certaine distribution, telle que la distribution normale. Avant de se fier à ces hypothèses, il est essentiel de vérifier dans quelle mesure les données correspondent à ces distribution théoriques.

Ces tests se basent sur deux hypothèses principales : l'hypothèse nulle notée H_0 et l'hypothèse alternative notée H_1.

2.1.1 L'hypothèse nulle et l'hypothèse alternative :

Définition 2.1.1 *L'hypothèse nulle* (H_0) *Représente généralement l'absence de différence ou d'effet entre les groupes ou les conditions étudiées. Elle est souvent formulée de manière à être rejeter, car le but du test est déterminer si les données fournissent suffisamment de preuves pour rejeter cette hypothèse et accepter l'hypo-*

thèse alternative.

Remarque 2.1.1 *Cette hypothèse est formulée dans le but d'être rejetée.*

Définition 2.1.2 *L'hypothèse alternative* (H_1) *Elle représente l'état alternative testé, supposant l'existence d'un effet ou d'une différence différente de l'hypothèse de base.*

Remarque 2.1.2 *La décision de rejeter* H_0 *signifie que* H_1 *est réalisée ou* H_1 *est vraie.*

En général, un test statistique est conçu pour évaluer dans quelle mesure les données contredisent l'hypothèse nulle, et ainsi tester le degré de soutien des données à l'hypothèse alternative. Cela nous permet de comprendre s'il existe une différence ou un effet significatif dans les données analysées qui mérite d'être examiné.

2.1.2 P-Valeur

En statistique la P-valeur est la probabilité pour un modèle statistique donne sous l'hypothèse nulle d'obtenir la même valeur ou une valeur encore plus extrême que celle observée sur l'échantillon. On va comparer un seuil de signification α et P-valeur pour accepter ou rejeter H_0 comme suit

– Si $P\text{-}valeur \leq \alpha$ on va rejeter l'hypothèse H_0.

– Si $P\text{-}valeur > \alpha$ on va accepter l'hypothèse H_0.

On peut alors interpréter la P-valeur comme le plus petit seuil de significativité pour lequel H_0 est acceptée.

2.1.3 Erreurs et risques

Lors de la prise de décision (choisir H_0 ou H_1) quatre situations sont possibles :

– Accepter H_0 et elle est vraie (bonne décision)

– Rejeter H_0 et elle est fausse (bonne décision)

– Rejeter H_0 et elle est vraie (appellé erreur de première espèce)

– Accepter H_0 et elle est fausse (appellé erreur de deuxième espèce)

Le risque est la probabilité de l'erreur, on a donc deux types de risque : risque de premier espèce et risque de deuxième espèce,le premier est généralement noté par "α" et le second "β".

$$\begin{cases} \alpha = \alpha(\theta) = \mathcal{P}(\text{rejeter } H_0 \backslash H_0 \text{ est vraie}) = \mathcal{P}(H_1/H_0). \\ \beta = \beta(\theta) = \mathcal{P}(\text{accepter } H_0 \backslash H_0 \text{ est fausse}) = \mathcal{P}(H_0/H_1). \end{cases}$$

Les tests sont fondés sur le schéma suivant :

décision|vraie	H_0	H_1
H_0	$1-\alpha$	β
H_1	α	$1-\beta$

2.1.4 Puissance d'un test

Définition 2.1.3 *La puissance d'un test est la probabilité de rejeter l'hypothèse nulle H_0 quand l'alternative H_1 est vraie. On la note par*

$$\pi := P\left[\textit{rejeter } H_0 \backslash H_1 \textit{ est vraie}\right] = 1 - \beta.$$

Lorsque H_1 est composite, la puissance est variable sur Θ_1. De même lorsque H_0 est composite, le risque de première espèce est variable sur Θ_0. On définit alors une fonction sur l'ensemble Θ qu'on appel le fonction puissance

$$\pi(\theta) := P_0[\textit{rejeter } H_0], \theta \in \Theta.$$

1. Si $\theta \in \Theta_0, \pi(\theta) = \alpha(\theta)$ c'est le risque de première espèce.

2. Si $\theta \in \Theta_1, \pi(\theta) = 1 - \beta(\theta)$ c'est la puissance du test.

2.1.5 Région de rejet et région critique

Définition 2.1.4 *La région de rejet d'un test est l'ensemble des points* $(X_1, ..., X_n)$ *de* $\mathbb{R}^n$ *pour lequel l'hypohèse nulle* H_0 *est écartée au profit de l'hypothèse alternative* H_1. *On appelle aussi région critique du test et on la note généralement par* W. *Elle est définie par la relation*

$$\mathcal{P}(W \backslash H_0) = \alpha$$

Le complémentaire de la région critique est appelée région d'acceptation du test. Elle est notée par $\bar{W}$ et est définie par

$$\mathcal{P}(\bar{W} \backslash H_0) = 1 - \alpha$$

Maintenant, plongeons dans le monde des tests statistiques ajustement. Alors que les tests statistiques traditionnels nous aident à comparer des moyennes, à évaluer des corrélations ou à tester des hypothèses, les tests statistiques d'ajustement sont un peu différents.Ils sont spécialement conçus pour évaluer à quel point nos données correspondent à une distribution de probabilité théorique spécifique.

En d'autres termes, les tests d'ajustement nous aident à répondre à des questions comme"Est-ce que mes données suivent une distribution normale ?", ou "Est-ce qu'elles correspondent à une distribution spécifique, comme la distribution binomiale ou exponentielle ?". En explorant les tests statistiques d'ajustement, nous allons plonger dans les fondements de la statistique et apprendre à évaluer la structure sous-jacente de nos données de manière plus approfondie. Cela nous aidera à affiner nos modèles statistiques et à améliorer la précision de nos analyses.

2.2 Test d'ajustement

Il est fréquent qu'en statistique on se pose le problème de savoir si une série de données peut être considérée comme pouvant être ajustée à une loi de probabilité déterminée. Cela permet de modéliser le phénomène étudié.

Majoritairement, le test du khi-deux est utilisé pour tester la modélisation par une loi discrète telle que la loi uniforme, la loi binomiale ou la loi de Poisson...

Pour tester la modélisation par une loi continue, comme la loi normale ou la loi exponentielle..., on utilise le test de Kolmogorov-Smirnov ou le test de Shapiro.

Il existe une grande variété de tests d'ajustement, parmi lesquels on peut citer :

-Le test du Khi-deux ,basés sur les effectifs.

-Les tests qui reposent sur la fonction de répartition empirique :tests de Kolmogorov-Smirnov,d'Anderson-Darling et de Cramer-von Mises.

-Les tests d'ajustement appliqués à la loi normale, appelés tests de normalité, comme les tests de Lilliefors, de Jarque-Bera et de Shapiro-Wilk.

2.2.1 Principe des tests ajustement

Le principe des tests d'ajustement consiste à comparer une distribution observée avec une distribution théorique pour évaluer leur compatiblité. Voici un résumé :

Hypothèse Nulle (H_0) :

Pour tous les tests d'ajustement, l'hypothèse nulle est généralement que la distribution observée correspond à la distribution théorique, sans spécifier la nature de cette correspondance.

Statistique de Test :

Chaque test d'ajustement utilise une statistique de test spécifique pour mesurer l'écart entre la distribution observée et la distribution théorique. Cela peut varier en fonction du test utilisé.

Distribution de Référence :

La statistique de test suit généralement une distribution de référence sous l'hypothèse nulle, telle que la distribution khi-deux pour le test du khi-deux ou la distribution normale pour le test de Kolmogorov-Smirnov...etc.

Décision :

La décision de rejeter ou d'accepter l'hypothèse nulle est basée sur la comparaison de la statistique de test avec une valeur critique ou sur le calcul d'une p-valeur associée à la statistique de test.

2.2.2 Hypothéses de test

Les hypothèses d'ajustement statistique sont fondamentales pour l'analyse des données, car elles permettent de déterminer si les observations correspondent à un modèle théorique donné. Ces hypothèses sont exprimées sous forme de deux propositions, l'hypothèse nulle (H_0) et l'hypothèse alternative (H_1), qui sont formulées mathématiquement comme suit :

- **L'hypothèse nulle** (H_0) : est une affirmation de départ qui stipule qu'il n'y pas de différence significative entre les données observées et les données attendues selon un modèle théorique.
- **l'hypothèse alternative** (H_1) : en revanche suggère qu'il existe une différence significative entre les données observées et les données attendues, indiquant ainsi

un ajustement inadéquat du modèle théorique aux données.

Nous écrivons ces hypothèses en abrégé comme suit :

$$\begin{cases} H_0 : F = F_0 \\ H_1 : F \neq F_0 \end{cases} \text{ ou bien } \begin{cases} H_0 : X \sim \text{une certaine loi spécifie,} \\ H_1 : X \nsim \text{cette loi.} \end{cases}$$

2.3 Quelques tests d'ajustement

Chacune de tests utilise des méthodes spécifiques pour comparer la distribution observée avec la distribution théorique et décider si elles sont compatibles. On va voir quelques exemples :

2.3.1 Test de Khi-deux

À partire de l'échantillon $X = (X_1, ..., X_n)$ on peut vérifié la qualité d'ajustement à une distribution théorique spécifiée par l'hypothèse nulle H_0. Posons P_0 une loi donnée et considérons le problème du test :

$$\begin{cases} H_0 : & (P = P_0) \\ H_1 : & (P \neq P_0) \end{cases} , où\ P_0(\theta_k) = P_{0k}.$$

Intuitivement, si les X_i suivent la loi P_0, la distance de Khi-deux $D(P_n, P_0)$ entre P_n et P_0 sera petite(Q décroit vers 0), par ailleurs on sait que si les X_i suivent la loi P_0, alors suit asymptotiquement une loi du χ^2 à $(m-1)$ dégrés de liberté.

– La statistique de Khi-deux définie par :

$$Q = D(P_n, P_0) = \frac{\sum (N_k - np_{0k})^2}{np_{0k}}. \tag{2.1}$$

– La région critique : on rejette H_0 si

$$Q > \chi^2{}_{(m-1)}(1-\alpha) = q_\alpha$$

Remarque 2.3.1 *Pour tester $H_0(P = P_{0,\theta})$, où $P_{0,\theta}$ est une famille de loi ($\theta \in \mathbb{R}^+$), alors H_0 est rejetté si :*

$$Q > \chi^2{}_{(m-r-1)}(1-\alpha) = q_{\alpha r}$$

Exemple 2.3.1 *Une bureau d'études a proposé trois modèles de design différents pour le produit d'un entreprise ; le responsable marketing affirme que le premier design sera deux fois plus populaire que le deuxième modèle sera trois fois plus populaire que le troisième. Lors d'un test de marché auprès de* 213 *personnes,* 111 *ont préféré le premier modèle,* 62*préféraient le deuxième modèle, et les autres préféraient le troisième modèle. Est-ce que ces résultats sont cohérents avec les affirmations du responsable marketing ?*

	M_1	M_2	M_3
n_i	111	62	40
P_k	$6\alpha = 0.6$	$3\alpha = 0.3$	$\alpha = 0.1$

ona :

$$\sum P_k = 1 \Rightarrow 10\alpha = 1 \Rightarrow \alpha = \frac{1}{10} = 0.1$$

Nous appliquons maintenant le test de khi-deux d'ajustement :

$$Q = \sum_{k=1}^{m=3} \frac{(N_k - nP_k)^2}{nP_k} = \frac{(111 - 0.6(213))^2}{0.6(213)} + \frac{(62 - 0.3(213))^2}{0.3(213)} + \frac{(40 - 0.1(213))^2}{0.1(213)}$$

$$= 18.68 > q_{(1-\alpha)} = \chi^2{}_{(m-1)}(0.95) = 0.1$$

$$\Longrightarrow \text{on accepte : } H_1(F \neq P_k)$$

Code R

```
X=c(111,62,40)

chisq.test(X)

Chi-squared test for given probabilities

data : X

X-squared = 37.2113, p-value=8.311e-09
```

2.3.2 Test de Kolmogorov-Smirnov

C'est le plus populaire parmi les tests d'adéquation qui sont basés sur la fonction de répartition empirique. Il a été proposé par Andreï N. Kolmogorov en 1933 et étendu par Vladimir I. Smirnov en 1939.

1.Statistique du test

La distance utilisée pour définir la statistique D_{n}de ce test est celle de la norme uniforme.

La statistique de Kolmogorov-Smirnov est alors définie par :

$$D_{KS} := \sup_{x\in\mathbb{R}} |F_n(x) - F_0(x)| . \tag{2.2}$$

Proposition 2.3.1 *La statistique de Kolmogorov-Smirnov s'écrit comme suit :*

$$D_{KS} = \max_{1\leq i\leq n} \max \left\{ \left| F_0(x_{(i)}) - \frac{i}{n} \right| , \left| F_0(x_{(i)}) - \frac{i-1}{n} \right| \right\} \tag{2.3}$$

2.Principe du test

On calcule la distance entreF_n et F_0 en utilisant la relalion (2.3) puis on décide du rejet ou non du modèle proposé. L'exécution de test de Kolmogorov-Smirnov est

donnée par les étapes suivantes :

1. Classer les valeurs observées par ordre croissant ;
2. Calculer, pour $i = 1$, les valeurs absolues des écarts

$$\left| F_0(x_{(i)}) - \frac{i}{n} \right| \text{ et } \left| F_0(x_{(i)}) - \frac{i-1}{n} \right| ;$$

3. Prendre le plus grand des deux écarts absolus ;
4. Répéter les étapes 2 et 3 pour $i = 2, ..., n$;
5. La valeur de la distance de Kolmogorov-Smirnov est égale au maximum des plus grands écarts.

On rejette H_0 si $\sqrt{n}D_{KS} > d_{n,\alpha}$, où ($d_{n,\alpha}$ est le quantile théorique lu à partir la table de Kolmogorov-Smirnov).

α	Valeur critique $d_{n,\alpha}$
0.10	$\frac{1.223}{\sqrt{n}}$
0.05	$\frac{1.358}{\sqrt{n}}$
0.01	$\frac{1.629}{\sqrt{n}}$

TAB. 2.1 – Les valeur critique du test de Kolmogorov-Smirnov en fonction de n.

Exemple 2.3.2 *On souhaite étudier le temps X (en mois) mais par* 10 *étudiants (diplomés) pour obtenir un emplois. On prend :*

$$3.5, 16, 18, 14, 26, 17.5, 12, 22.5, 36, 10.$$

On cherche à tester $H_0(X \rightsquigarrow Exp(\lambda = 1/5))$ avec un risque $\alpha = 0.05$.

Solution 2.3.1 *Sous R ;*

```
X<-c(3.5,16,18,14,26,17.5,12,22.5,36,10)

ks.test(X,"exp",lambda=1/5)

One sample Kolmogorov Smirnov test

data :X

D =0.88248 ;p-value=3.442exp(07)

alternative hypothes is :two-sided

p-value=0.003
```

2.3.3 Test de Cramer-von Mises

Le test était développé par Harald Cramer et Richard E. von Mises (1928-1930), Voir (**??**). Ce test est une variante au test de Kolmogrov-Smirnov. Il permet égalment de tester toute forme de différenciation entre les distributions. Sa particularité est qu'il exploite différemment les fonctions de répartition empirique au lieu de se focaliser sur l'écart maximal, il compile tous les écarts sous la forme d'une somme des carrés des différences.

1. **La statistique du test :**

$$\begin{aligned} \mathbf{W}_n^2 &= n \int_{-\infty}^{+\infty} [F_n(x) - F_0(x)]^2 dF_0(x) \\ &= \frac{1}{12n} + \sum_{i=1}^{n} (\frac{2i-1}{2n} - F_0(x_i))^2 \end{aligned} \tag{2.4}$$

2. **Principe du test**

On calcule la distance entre F_n et F_0 en utilisant la relation(2.4), puis on décide du rejet ou non du modèle proposé. L'exécution du test de Cramer-von Mises est donnée par les étapes suivantes :

1. Classer les valeurs observées par ordre croissant ;

2. Utiliser la fonction de répartition de la loi pour obtenir les valeurs de $F_0(x_i)$, pour $i = 1, ..., n$;
3. Calculer $\sum_{i=1}^{n}(\frac{2i-1}{2n} - F_0(x_i))^2$ puis la valeur de la statistique $\mathbf{W}_n^2$.

On rejette l'hypothèse H_0 si cette dernière est supérieure à une certaine valeur critique ($\mathbf{W}_{crit}^2$) :

$$\mathbf{W}_n^2 \geq \mathbf{W}_{crit}^2$$

pour un niveau α donnée et pour $n = 30$, les valeurs critiques sont résumées dans le tableau suivant :

α	$\mathbf{W}_{crit}^2$
0.10	0.172
0.05	0.218
0.01	0.33

TAB. 2.2 – Valeurs critiques du test Cramer-von Mises

Exemple 2.3.3 *Pour population Ω, on veut étudier la conformité de la distribution d'une v.a continue (X) à une distribution normale, on dispose pour cela un échantillon de taille $n = 30$ observations suivants :*

$$X = (14, 14, 18, 17, 16, 17, 2, 6, 9, 13, 12, 0, 6, 2, 20,$$
$$21, 28, 30, 21, 32, 10, 20, 23, 22, 10, 13, 11, 13, 13, 12)$$

Calculs du test

1. Trier des données brutes en ordre croissant $(x_{(i)})_{i=1}^{n}, n = 30$;

2. Centrage et réduction des valeur de X ;

$$z_{(i)} = \frac{x_{(i)} - \bar{X}}{S}, \text{ où } \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{445}{30} = 14.83 \text{ et } S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2} = 7.8743$$

3. Trouver les valeurs de Z correspondantes avec le tableau de loi normale ;
4. Utilisons la fonction de répartition de la loi normale centrée et réduite pour extraire les fréquences théoriques $F_0(x_{(i)})$;
5. calculer la statistique

$$\mathbf{W}_n^2 = \frac{1}{12n} + \sum_{i=1}^{n}(\frac{2i-1}{2n} - F(x_i))^2 = 0.041701 < \mathbf{W}_{crit}^2 = 0.218$$

Alors au risque $\alpha = 0.05$; on accepte l'hypothèse H_0.

Code R :

X=c(14,14,18,17,16,17,2,6,9,13,12,0,6,2,20,21,28,30,21,32,10,20,23,22,10,13,11,13,13,12)

cvm.test(X)

Cramer-von Mises normality test

data :X

W =0.041701,p-value=0.6374

alternative hypothesis :two-sided

Commentaire : On remarque que $p - value > \alpha$ $(0.6374 > 0.05)$, donc on accepte l'hypothèse de la normalité (H_0) au seuil de risque $\alpha = 5\%$.

2.3.4 Test de normalité de Lilliefors

Ce test est une variante du test de Kolmogorov-Smirnov, sous l'hypothèse de normalité (à chercher à tester $H_0 : P \rightsquigarrow Gaussienne$), où les paramètres μ, σ de la loi

sont estimès à partir des données. Les hypothèses suivantes à tester sont donc :

$$\begin{cases} H_0 : \text{ les données suivent une loi normale,} \\ H_1 : \text{ les données ne suivent pas une loi normale.} \end{cases}$$

– La statistique du test est :

$$L_n = \sqrt{n} \max_{1 \leq i \leq n} \max \left\{ \left| F_0(\frac{x_{(i)} - \bar{x}}{S_x}) - \frac{i}{n} \right|, \left| F_0(\frac{x_{(i)} - \bar{x}}{S_x}) - \frac{i-1}{n} \right| \right\} \tag{2.5}$$

où $\bar{x}$ est la moyenne empirique et S_x est l'écart type empirique.

On rejette H_0 si $L_n > D_{crit}$ (D_{crit} la valeur critique de test Lilliefors).

Valeurs critiques :

La table des valeurs critique D_{crit} pour les petites valeur de n et différentes valeur de α doivent être utilisées lorsque les effectifs sont élevés, typiquement $n > 30$, il est possible d'approcher la valeur critique à l'aide de formules simples :

α	Valeur critique D_{crit}
0.10	$\frac{0.805}{\sqrt{n}}$
0.05	$\frac{0.886}{\sqrt{n}}$
0.01	$\frac{1.031}{\sqrt{n}}$

TAB. 2.3 – Les valeurs critiques du test de Lilliefors en fonction de n.

Exemple 2.3.4 *Les observations ci-dessous correspondent à la hauteur de* 8 *arbres dans une forêt. Peut-on considérer que la distribution de ces hauteurs est gaussienne ?*

$$23.4; 24.6; 25.1; 26.3; 26.8; 27.2; 27.6; 28.3.$$

On applique le test de lilliefors :

$$L_n = \sqrt{n} \max_{1\leq i\leq n} \max \left\{ \left| F_0(\frac{x_{(i)} - \bar{x}}{S_x}) - \frac{i}{n} \right|, \left| F_0(\frac{x_{(i)} - \bar{x}}{S_x}) - \frac{i-1}{n} \right| \right\}$$

et ona $F_0 \rightsquigarrow \mathcal{N}(0,1)$, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = 26.16, S_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 1.66$

i	1	2	3	4	5	6	7	8
x_i	23.4	24.6	25.1	26.3	26.8	27.2	27.6	28.3
$\frac{x_i-\bar{x}}{S_x}$	-1.66	-0.34	-0.64	0.08	0.39	0.63	0.87	1.29
$F_0(\frac{x_i-\bar{x}}{S_x})$	0.05	0.17	0.26	0.53	0.65	0.73	0.81	0.90
$\frac{i-1}{n}$	0	0.12	0.25	0.37	0.5	0.62	0.75	0.87
$\frac{i}{n}$	0.12	0.25	0.37	0.5	0.62	0.75	0.87	1
$F_0(\frac{x_{(i)}-\bar{x}}{S_x}) - \frac{i}{n}$	-0.07	-0.08	-0.11	0.03	0.03	-0.02	-0.06	-0.1
$F_0(\frac{x_{(i)}-\bar{x}}{S_x}) - \frac{i-1}{n}$	0.05	0.05	0.01	0.16	0.15	0.11	0.06	0.03

$$\Longrightarrow \max_{1\leq i\leq n} \max \left\{ \left| F_0(\frac{x_{(i)} - \bar{x}}{S_x}) - \frac{i}{n} \right|, \left| F_0(\frac{x_{(i)} - \bar{x}}{S_x}) - \frac{i-1}{n} \right| \right\} = 0.16$$

$$\Longrightarrow L_n = \sqrt{8} * 0.16 = 0.35 > q_l(1-\alpha) = 0.287$$

$$\Longrightarrow H_0(X \nsim Gaussienne)$$

2.3.5 Test de Shapiro-Wilk

Très populaire, le test de Shapiro-Wilk est basé sur la statistique W. En comparaison des autres tests, il est particulièrement puissant pour les petits ectifs ($n \leq 50$). La statistique du test s'écrit :

$$W = \frac{\left[\sum_{i=1}^{\left[\frac{n}{2}\right]} \alpha_i (x_{(n-i+1)} - x_{(i)})\right]^2}{\sum_i^n (x_i - \bar{x})^2} \tag{2.6}$$

où

-$(x_{(i)})$ correspond à la série des données triées ;

-$\left[\frac{n}{2}\right]$ est la partie entière du rapport $\frac{n}{2}$;

-α_i sont des constantes générées à partir de la moyenne et de la matrice de variance co-variance des quantiles d'un échantillon de taille n suivant la loi normale. Ces constantes sont fournies dans des tables spécifiques.

La statistique W peut donc être interprétée comme le coefficient de détermination (le carré du coefficient de corrélation) entre la série des quantiles générées à partir de la loi normale et les quantiles empiriques obtenues à partir des données. Plus W est élevé, plus la compatibilité avec la loi normale est crédible. La région critique, rejet de la normalité, s'écrit :

$$R.C. : W < W_{crit}$$

Les valeurs seuils W_{crit} pour différents risques α et e ectifs n sont lues dans la table de Shapiro-Wilk.

Exemple 2.3.5 *On prend le même exemple (2.3.3)*

Calculs du test

1. Trier les données x_i, nous obtenons la série $x_{(i)}$;
2. Calculer les quantités

$$(x_{(n-i+1)} - x_{(i)}),\ i = 1, ..., \left[\frac{n}{2}\right]$$

3. Lire dans la table pour $n = 30$ et $i = 1, ..., 15$ les valeur de coefficient α_i.

4. Calculer la statistique

$$W = \frac{\left[\sum_{i=1}^{\left[\frac{n}{2}\right]} \alpha_i (x_{(n-i+1)} - x_{(i)})\right]^2}{\sum_i^n (x_i - \bar{x})^2} = 0.97656;$$

5. Pour une risque $\alpha = 0.05$, le seuil critique lue dans la table de Shapiro-Wilk

$$W_{crit} = 0.927;$$

6. Comparer entre W et W_{crit}

$$W > W_{crit}$$

Alors au risque $\alpha = 0.05$, on accepte l'hypothèse H_0.

Code R :

X=c(14,14,18,17,16,17,2,6,9,13,12,0,6,2,20,21,28,30,21,32,10,20,23,22,10,13,11,13,13,12)

shapiro.test(X)

Shapiro-Wilk normality test

W=0.97656, p-value=0.7285

Commentaire : On remarque que $p - value > \alpha$

donc on accepte l'hypothèse H_0 au risque $\alpha = 5\%$ (la v.a X suit une loi normale).

Chapitre 3

Application sur R

introduction

Le test du Khi-deux d'ajustement reste pertinent pour les grands échantillons, permettant de vérifier si une distribution observée suit une distribution théorique attendue, même avec un nombre élevé de données.

Exemple pratique

Supposons que vous travailliez dans une entreprise de vente en ligne et que vous analysez les données d'achat des clients sur une période d'un an. Vous souhaitez savoir si la répartition des montants d'achat suit une distribution normale, comme on le suppose souvent pour ce type de données.

Étapes

1. **Données simulées :** Simulez un échantillon de 100 montants d'achat (en euros) suivant une distribution normale avec une moyenne de 50 euros et un écart-type de 10 euros.On considére les deux hypothèses suivants :

$$\begin{cases} H_0 : & X \sim \mathcal{N}(50, 10^2) \\ H_1 : & X \nsim \mathcal{N}(50, 10^2) \end{cases}$$

2. **Distribution théorique attendue :** La Distribution théorique attendue est

une distribution normale avec les paramètres définis précédemment (moyenne=50,écart-type=10).

3. **Test du Khi-deux d'ajustement :** Utilisez la fonction (`chisq.test()`) pour comparer la distribution observée des montants d'achat (`montants-d'achat)` à la distribution théorique normale.

4. **Interprétation des resultats :**

– *Valeur du Khi-deux* : Représente la force de la différence entre la distribution observée et distribution théorique attendue. On a ($X - squared = 163.7842$).

– *Valeur p (p-value) :* Si elle est inférieure au seuil de signification (généralement 0.05),on rejette l'hypothèse nulle (H_0) selon laquelle la pièce suit une distribution équitable.

5. **Conclusion :**

D'aprés la valeur p ($p - value = 0.031 < \alpha = 0.05$), on rejette l'hypothèse nulle H_0. Cela suggère que la distribution des montants d'achat ne semble pas compatible avec une distribution normale, comme attendu pour ce type de donné. D'autre part, l'examen du **Q-Q plot** et **plot** donné par la figure (3.1)et (3.2), permet de conclure que les données ne suit pas une loi normale.

Code R pour Q-Q plot :

```
montants_achat <- rnorm(100, mean = 50, sd = 10)
chi_resultat<-chisq.test(table(montants_achat),p=dnorm(unique(montants_achat),
mean=50,sd=10))
qqnorm(montants_achat)
qqline(montants_achat,col="red")
```

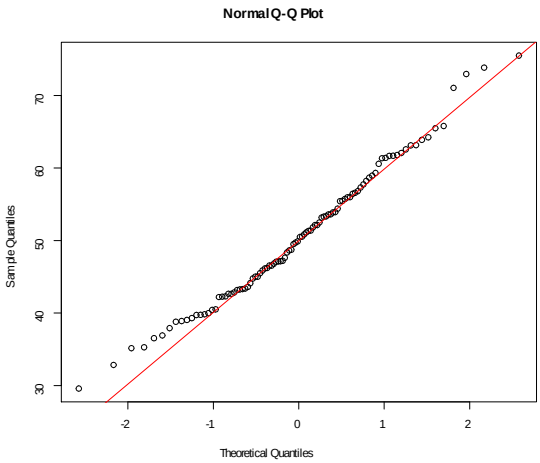

FIG. 3.1 – **Q-Q plot** Normal

Cod R pour plot

montants_achat <- rnorm(100, mean = 50, sd = 10)

chi_resultat<-chisq.test(table(montants_achat),p=dnorm(unique(montants_achat),

mean = 50,sd=10))

plot(density(montants_achat), main = "Density Plot of montants_achat", xlab = "Value", ylab = "Density", col = "red", lwd = 2)

curve(dnorm(x, mean(montants_achat), sd(montants_achat)), col = "blue", lwd = 2, add = TRUE)

legend("topright", legend = c("montants_achat Density", "Normal Density"), col = c("red", "blue"), lty=1,cex=0.8, pt.cex=0.8, inset=c(0.0001,0.0001))

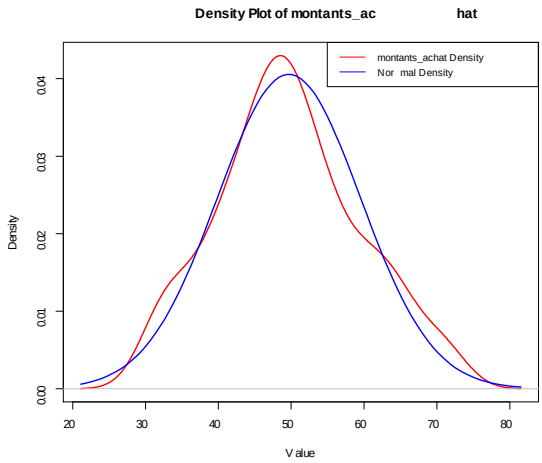

FIG. 3.2 – densité plot des montants d'achat

Conclusion

L'objectif de cette étude est l'application de tests d'ajustement à un échantillon statistique. Pour cela, une de tests statistiques et des principes fondamentaux liées à ces tests a été initialement établie.

Par la suite, une analyse des tests statistiques a été entreprise pour vérifier la conformité d'une variable aléatoire mesurée dans une population avec une loi de probabilité théorique prédéterminée(tests d'ajustement). En particulier, lnous avons étudié le test du Khi-deux, ainsi que le test de Kolmogrov-Smirnov... Toutes ces modalités de tests d'ajustement ont été explorées afin d'évaluer la normalité et ont été appliquées à l'aide d'exemples et du logiciel R. Tous les tests mentionnés ont été exécutés.

En conclusion, cette étude pourrait bénéficier d'une extension pour inclure l'analyse d'ajustment en cas multivarié.

Bibliographie

[1] Pablo M. C., (2014). Cramer-Von Mises Statistic for Repeated Measures. Revista Colombiana de Estadística

[2] Gusmia, N. H., (2018). Tests d'ajustement à une distribution basés sur la fonction de répartition empirique (mémoire). Université Mohamed Khider de biskra.

[3] Dhiab, M. Z, (2022). Tests d'ajustement et applications (mémoire). Université Mohamed Khider de biskra.

[4] Rahmouni, Y, (2019). Tests de normalité (mémoire). Université Mohamed Khider de biskra.

[5] ATMANI, F, (2020). Sur quelques tests non paramétriques et applications (mémoire). Université Mohamed Khider de biskra.

[6] ZAHNIT, S, (2020). Tests de normalité des données (mémoire). Université Mohamed Khider de biskra.

[7] Achour, C, (2021). Tests de normalité et applications (mémoire). Université Mohamed Khider de biskra.

[8] Saporta, G. (2006). Probabilité, analyse de données et statistique. Technip.

[9] Yahia.Dj. (2023). Cours de deuxième MASTER de SNP. Université Mohamed Khider de biskra.

[10] Lejeune, M. (2010). Statistique : la théorie et ses applications. Springer.

[11] Veysseyre, R. (20,2006). Aide-mémoire : statistique et probabilités pour l'ingénieur. Dunod.

[12] TORCHI, W, (2020). Tests statistiques et ses applications (mémoire). Université Mohamed Khider de biskra.

[13] Bouzid Bouchra, (2022). Sur la symétrie des lois de probabilité : une approche géométrique (mémoire). Université de Saïda Dr. Moulay Tahar.

[14] Mansouri, F, (2020). Tests de Normalité Multivariée (mémoire). Université Mohamed Khider de biskra.

[15] Barkati, R, (2019). Sur la loi normale et applications (mémoire). Université Mohamed Khider de biskra.

[16] Rakotomalala, R. (2008). Tests de normalité–Techniques empiriques et tests statistiques. Université Lumière Lyon 2, Version, 2.

Annexe A : Abréviations et Notations

Les différentes abréviations et notations utilisées tout au long de ce mémoire sont expliquées ci-dessous :

$E(X)$	:	Espérence mathématique.
$V(X)$	:	Variance mathématique.
X	:	Variable aléatoire
Ω	:	Ensemble de population.
$V.a$	:	variable aléatoire
$\mathcal{N}(0,1)$	:	Loi normale centrée réduite.
$\mathcal{E}(\lambda)$	:	Loi exponentielle
$\mathcal{U}([a,b])$	:	Loi uniforme
$F(x)$	:	fonction de répartition
$\bar{F}(x)$	:	fonction de survie
$F_n(x)$	:	fonction de répartition empirique
$f(x)$	:	fonction de densité
H_0	:	hypothèse nulle
H_1	:	hypothèse alternative
α	:	risque de premièr espèce
β	:	risque de deuxième espèce
$D_{K.S}$	:	la statistique de test Kolmogrov-Smirnov
L_n	:	la statistique de test Lillifors
W_n^2	:	la statistique de test Cramer-Von Mises
W	:	la statistique de test Shapiro-Wilk
$R.C$	:	région critique

Résumé :

Les tests statistique d'ajustement sont des outils mathématiques utilisés pour évaluer la qualité d'adéquation d'un modèle statistique à des données collectées. En d'autres termes, ces tests nous aident à déterminer si notre modèle représente avec précision le phénomène que nous étudions.

الملخص:

الاختبارات الإحصائية لقياس مدى ملاءمة النموذج هي أدوات رياضية تستخدم لتقييم جودة ملاءمة نموذج إحصائي للبيانات التي تم جمعها. بعبارة أخرى، تساعدنا هذه الاختبارات في تحديد ما إذا كان نموذجنا يمثل بدقة الظاهرة التي ندرسها.

Abstract:

Statistical goodness-of-fit tests are mathematical tools used to assess the quality of fit of a statistical model to collected data. In other words, these tests help us determine whether our model accurately represents the phenomenon we are studying.