

STATISTIQUE ET PROBABILITES

STATISTIQUE INFERENTIELLE

ECHANTILLONNAGE – ESTIMATION – TESTS D’HYPOTHESES

Lamia soltani

Année universitaire : 2017 - 2018

Sommaire

Chapitre 1: Variables aléatoires

I. Introduction

II- Variable aléatoire discrète

II.1- Définition

II.2- Loi d'une variable aléatoire discrète

II.3 Fonction de répartition

II.4- Espérance mathématique, valeurs typiques

III- Lois usuelles discrètes

III.1- Loi Uniforme

III.2- Loi de Bernoulli

III.3- Loi Binomiale

III.4- Loi de Poisson

III.5- Loi géométrique

III.6- Loi hypergéométrique

III.7- Approximation de la loi binomiale par une loi de Poisson

IV- Variables aléatoires continues

IV.1- Définition

IV.2- Fonction de répartition

IV.3- Espérance et variance

V- Lois usuelles continues

V.1- Loi Uniforme

V.2- Loi exponentielle

V.3- Loi Normale

VI- Les lois dérivées de la loi normale

VI.1- La loi du Khi deux

VI.2 La loi de Student

VI.3 La loi de Fisher

VII- Théorème central limite

Chapitre 2: Echantillonnage

I-Echantillon aléatoire et ses caractéristiques

I.1- Moyenne empirique

I.2- La variance empirique

I.3- Fréquence

II-Estimateur et ses propriétés statistiques

II.1- Définition d'un estimateur

II.2- Propriétés statistiques d'un estimateur

Chapitre 3 : Méthodes d'estimation ponctuelle et Intervalle de Confiance

I-Estimation ponctuelle : Méthode du Maximum de vraisemblance

I.1- Estimateur de la moyenne

I.2- Estimation d'une proportion théorique

I.3- Estimation du paramètre d'une loi de Poisson

I.2- Estimation de la variance σ^2

II-Estimation par intervalle de confiance

II.1- Intervalle de confiance pour la moyenne dans le cas d'une loi normale

II.2- Intervalle de confiance pour la proportion d'une loi de Bernoulli

II.3- Intervalle de confiance pour la moyenne d'une loi inconnue

II.4- Intervalle de confiance de la variance d'une population normale

Chapitre 4 : Tests d'hypothèses

I-Principe des tests

I.1- Formulation des hypothèses à tester

I.2- Erreurs, seuil de signification et puissance d'un test

I.3- Variable de décision et région critique

II-Test de la moyenne d'une loi normale

II.1- Test de la moyenne avec la variance σ^2 est connue

II.2- Test de la moyenne avec la variance σ^2 est inconnue

III-Test de la moyenne d'une loi inconnue

III.1- Test de la moyenne lorsque la variance σ^2 est connue

III.2- Test de la moyenne lorsque la variance σ^2 est inconnue

IV-Test de la proportion d'une loi de Bernoulli

V-Test de la variance d'une loi normale

V.1- Test de la variance lorsque la moyenne μ connue

V.2- Test de la variance lorsque la moyenne μ est inconnue

VI-Test de la variance d'une loi inconnue

VI.1- Test de la variance lorsque la moyenne m est connue

VI.2- Test de la variance lorsque la moyenne m est inconnue

Chapitre 1: Variables aléatoires

I. Introduction

Dans de nombreux jeux, on fait intervenir le hasard en observant la somme des points marqués par deux dés. Considérons le jet d'un dé bleu et d'un dé rouge et notons S la somme des points obtenus. On modélise cette expérience en prenant l'équiprobabilité sur : $\Omega = \{1, 2, 3, 4, 5, 6\}^2$.

Un événement élémentaire ω est un couple (b, r) où b désigne le résultat du dé bleu et r celui du rouge. Pour tout événement élémentaire $\omega = (b, r)$, on a $S(\omega) = b + r$. Il est facile de représenter tous les cas possibles par un tableau à 36 cases:

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|----|----|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| 6 | 7 | 8 | 9 | 10 | 11 | 12 |

On a ainsi défini une application S de Ω dans l'ensemble des sommes de points possibles : $\{2, 3, \dots, 11, 12\}$. On dit que S est une variable aléatoire sur Ω . En fait, l'observation qui nous intéresse dans cette expérience, ce n'est pas l'événement ω , mais seulement $S(\omega)$. Ce que l'on aimerait connaître, c'est la probabilité que la somme des points prenne une valeur donnée, soit $P(S = k)$ pour k entier fixé entre 2 et 12. En utilisant l'équiprobabilité sur Ω et le tableau ci-dessus, nous obtenons immédiatement :

| x_i | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|--------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|----------------|
| $P(X = x_i)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ |

Cela revient à considérer un nouvel ensemble d'événements élémentaires : $\Omega' = S(\Omega) = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$ et à munir cet ensemble de la probabilité PS définie par le tableau des $P(X = x_i)$. Cette nouvelle probabilité s'appelle loi de la variable aléatoire X .

II- Variable aléatoire discrète

II.1- Définition : Soit (Ω, \mathcal{F}, P) un espace probabilisé. On appelle variable aléatoire discrète sur (Ω, \mathcal{F}, P) toute application X :

$$X : \Omega \rightarrow \mathbb{R}$$

$$\omega \mapsto X(\omega)$$

vérifiant les deux conditions :

(i) L'ensemble des images $X(\Omega) = \{X(\omega), \omega \in \Omega\}$ est une partie au plus dénombrable de \mathbb{R} . On peut donc numéroté ses éléments par des indices entiers:

$$X(\Omega) = \{x_0, x_1, \dots, x_k, \dots\}.$$

(ii) Pour tout $x_k \in X(\Omega)$, $A_k = \{\omega \in \Omega, X(\omega) = x_k\}$ fait partie de la famille \mathcal{F} d'événements auxquels on peut attribuer une probabilité par P .

Remarquons que la famille de tous les A_k forme une partition de Ω : on classe chaque élément de Ω selon son image par X . Il en résulte :

$$\sum_{x_k \in X(\Omega)} P(A_k) = \sum_{x_k \in X(\Omega)} P(X = x_k) = 1$$

Dans cette écriture, les sommes sont des séries convergentes si $X(\Omega)$ est infini et des sommes ordinaires lorsque l'ensemble $X(\Omega)$ est fini.

II.2- Loi d'une variable aléatoire discrète

Définition : Soit X une variable aléatoire discrète sur (Ω, \mathcal{F}, P) . On lui associe la fonction d'ensemble P_X définie sur la famille de toutes les parties de \mathbb{R} en posant :

$$p_k = P_X(\{x_k\}) = P(A_k) = P(X = x_k),$$

puis pour tout $B \subset \mathbb{R}$:

$$P_X(B) = \sum_{x_k \in B} P(X = x_k) = \sum_{x_k \in B} p_k.$$

La fonction d'ensembles P_X ainsi définie est une probabilité sur la famille $\mathcal{P}(\mathbb{R})$ de toutes les parties de \mathbb{R} . On l'appelle loi de la variable aléatoire X .

La représentation graphique de la loi de probabilité est un diagramme en bâton.

II.3 Fonction de répartition

Définition : On appelle fonction de répartition de la variable aléatoire X , la fonction F_X définie sur R par :

$$\forall x \in R, F_X(x) = P_X([-\infty, x]) = P(X \leq x).$$

On a aussi : $F_X(x) = \sum_{x_k \in X(\Omega), x_k \leq x} P(X = x_k)$.

Théorème 1 La fonction de répartition F_X d'une variable aléatoire discrète X est croissante sur R , continue à droite et limitée à gauche en tout point. Elle tend vers 0 en $-\infty$ et vers 1 en $+\infty$. Elle caractérise la loi de X , autrement dit : $F_X = F_Y$ si et seulement si les variables aléatoires X et Y ont même loi.

Voici à titre d'exemple la représentation graphique de F_X où X est la variable aléatoire somme des points de deux dés.

| x_i | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | ≥ 12 |
|--------------|----------------|----------------|----------------|----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------|
| $P(X = x_i)$ | $\frac{1}{36}$ | $\frac{2}{36}$ | $\frac{3}{36}$ | $\frac{4}{36}$ | $\frac{5}{36}$ | $\frac{6}{36}$ | $\frac{5}{36}$ | $\frac{4}{36}$ | $\frac{3}{36}$ | $\frac{2}{36}$ | $\frac{1}{36}$ | |
| $P(X < x_i)$ | 0 | $\frac{1}{36}$ | $\frac{3}{36}$ | $\frac{6}{36}$ | $\frac{10}{36}$ | $\frac{15}{36}$ | $\frac{21}{36}$ | $\frac{26}{36}$ | $\frac{30}{36}$ | $\frac{33}{36}$ | $\frac{35}{36}$ | 1 |

La représentation graphique de la fonction de répartition est un diagramme en escalier.

II.4- Espérance mathématique, valeurs typiques

Soit X une variable aléatoire dont l'ensemble des valeurs s'écrit $\{x_1, x_2, \dots, x_n\}$. L'espérance mathématique de X est le réel :

$$E(X) = \sum_{k=1}^n x_k P(X = x_k)$$

Propriété 1

Soit X une variable aléatoire. Si $m = \min(X(\Omega))$ et $M = \max(X(\Omega))$, alors $E(X) \in [m, M]$.

Corrolaire : En particulier, si X est une variable aléatoire positive, alors $E(X) \geq 0$.

Propriétés :

Linéarité de l'espérance

- Soit X une variable aléatoire et $(a, b) \in \mathbb{R}^2$. On a $E(aX + b) = aE(X) + b$
- Soient X et Y deux variables aléatoires indépendantes alors $E(X + Y) = E(X) + E(Y)$

Variance d'une variable aléatoire

Soit X une variable aléatoire. La variance de X est définie par :

$$V(X) = E\left((X - E(X))^2\right) = E(X^2) - (E(X))^2$$

On appelle **écart-type** de X le réel $\sigma(X) = \sqrt{V(X)}$

Propriétés :

Soit X une variable aléatoire et a et b deux réels.

- $V(X + b) = V(X)$ et $\sigma(X + b) = \sigma(X)$ (*invariance par translation*) ;
- $V(aX) = a^2V(X)$ et $\sigma(aX) = |a|\sigma(X)$;

Théorème : Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire d'espérance $E(X)$ et de variance $V(X)$. Alors, pour tout $\varepsilon > 0$

$$\text{on a : } P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}$$

III- Lois usuelles discrètes

III.1- Loi Uniforme

Elle modélise des situations d'équiprobabilités

III.1.1 Définition : Une variable aléatoire X suit une loi uniforme discrète lorsqu'elle prend ses valeurs dans l'ensemble $\{1, \dots, n\}$ avec équiprobabilités (des probabilités élémentaires identiques).

$$X \rightarrow U(1, 2, \dots, n) \text{ si : } \forall k = 1, \dots, n; \quad p[X = k] = \frac{1}{n}$$

Donc on dit que les événements $\{x = 1\}, \dots, \{x = n\}$ sont équiprobables.

III.1.2 Les caractéristiques :

Espérance mathématique :

$$E(X) = \frac{n+1}{2}$$

Variance :

$$V(X) = \frac{n^2 - 1}{12}$$

Remarque : La distribution de probabilité d'une loi Uniforme est représentée par un diagramme en bâtons.

Exemple :

On lance un dé (non pipé). Soit X une variable aléatoire qui représente les chiffres obtenus au lancer.

- a) Déterminer l'espace fondamental Ω et $X(\Omega)$.
- b) Donner la loi de probabilité.
- c) Calculer $E(X)$ et $V(X)$.
- d) Représenter graphiquement la distribution des probabilités.

III.2- Loi de Bernoulli

Cette loi modélise une expérience dont l'issue ne possède que deux alternatives de type « succès ou échec », « oui ou non », « vrai ou faux », « pile ou face »...etc. Un succès est représenté par l'événement $\{X=1\}$. Alors que $\{X=0\}$ correspond à un échec.

La variable aléatoire $X(\Omega) = \{0, 1\}$.

III.2.1- Définition

X suit la loi de Bernoulli de paramètre p ($p \in [0, 1]$), si $X(\Omega) = \{0, 1\}$ et sa fonction de probabilité est définie par : $p(X = k) = \begin{cases} p^k(1-p)^{1-k} & \text{si } X = 0, 1 \\ 0 & \text{Si non} \end{cases}$

On écrit $X \rightarrow B(1, p)$ avec $p = p(X = 1)$ et $p(X = 0) = 1 - p = q$.

Si A est un événement de probabilité p , son indicatrice définie par:

$$\mathbb{I}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \notin A \end{cases} = \begin{cases} 1 & \text{si } A \text{ est réalisé} \\ 0 & \text{si } A \text{ n'est pas réalisé} \end{cases}$$

est une variable aléatoire suivant la loi de Bernoulli de paramètre p . Réciproquement, si X est une v.a. de Bernoulli, on peut toujours écrire $X = \mathbb{I}_A$ en définissant $A = \{\omega \in \Omega, X(\omega) = 1\}$.

III.2.2- Les caractéristiques

Espérance mathématique :

$$X \rightarrow B(1, p) \text{ donc } E(X) = p$$

Variance :

$$X \rightarrow B(1, p) \text{ donc } V(X) = p(1 - p)$$

Remarque : La distribution de probabilité d'une variable de Bernoulli est représentée par un diagramme en bâtons.

Propriété :

Soient les deux variables aléatoires indépendantes X_1 et X_2 , $X_1 \rightarrow B(1, p)$ et $X_2 \rightarrow B(1, p)$

Soit la variable aléatoire $Z = X_1 + X_2$ n'est pas distribuée selon une loi de Bernoulli.

Exemple

On jette une pièce de monnaie équilibrée. Soit X une variable aléatoire qui est égale à 1 si pile, 0 si non.

- Déterminer l'espace fondamental Ω et $X(\Omega)$.
- Donner la loi de probabilité .
- Calculer $E(X)$ et $V(X)$.
- Représenter graphiquement la distribution des probabilités.

III.3- Loi Binomiale

III.3.1- Définition

La variable binomiale, X , représente le nombre de succès obtenus lors de la répétition de n épreuves identiques et indépendantes. Chaque épreuve ne peut donner que deux résultats possibles.

La variable aléatoire X suit la loi binomiale de paramètres n et p ($n \in \mathbb{N}^*$ et $p \in [0, 1]$) si l'ensemble des valeurs possibles est $X(\Omega) = \{0, 1, \dots, n\}$ et $\forall k = 0, 1, \dots, n$, $P(X = k) = C_n^k p^k (1 - p)^{n-k} = C_n^k p^k q^{n-k}$, avec $q = 1 - p$

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

Notation : $X \sim B(n, p)$. La formule ci-dessus définit bien une loi de probabilité puisque les $C_n^k p^k (1 - p)^{n-k}$ sont positifs et $\sum_{k=0}^n C_n^k p^k (1 - p)^{n-k} = (p + (1 - p))^n = 1$.

III.3.2- Les caractéristiques

Espérance mathématique :

Si $X \rightarrow B(n, p)$ donc $E(X) = np$

Variance :

Si $X \rightarrow B(n, p)$ donc $V(X) = npq$ avec $q = 1 - p$

Remarques :

La loi binomiale $B(n, p)$ est la loi du nombre de succès obtenus en une suite de n épreuves répétées indépendantes tel que chaque épreuve a une probabilité de succès égale à p . De même, soit A_1, \dots, A_n une famille de n événements mutuellement indépendants ayant tous même probabilité p et notons X_i la variable de Bernoulli indicatrice de A_i :

$$X_i(\omega) = \begin{cases} 1 & \text{si } \omega \in A_i \\ 0 & \text{si } \omega \in A_i^c \end{cases}$$

Alors la variable aléatoire $S_n = \sum_{i=1}^n X_i$ suit la loi binomiale $B(n, p)$

La distribution de probabilité d'une variable binomiale est représentée par un diagramme en bâtons.

Soient X_1 et X_2 deux variables aléatoires indépendantes : $X_1 \rightarrow B(n_1, p)$ et $X_2 \rightarrow B(n_2, p)$ alors $(X_1 + X_2) \rightarrow B(n_1 + n_2, p)$

Exemple :

Dans une population, on compte 35% sont de sexe masculin. On prend au hasard un échantillon de 40 individus. Soit la variable aléatoire X « nombre d'individus de sexe masculin dans l'échantillon tiré ».

- Donner la loi de probabilité de X .
- Donner la probabilité d'avoir 9 garçons dans l'échantillon tiré.
- Donner la probabilité d'avoir au moins un garçon dans l'échantillon tiré.
- Calculer $E(X)$ et $V(X)$.

III.4- Loi de Poisson

Cette loi est souvent utilisée dans la modélisation des files d'attente : nombre de clients en attente à une caisse de supermarché, nombre de conducteurs passés à un péage pendant une période de temps.

III.4.1- Définition

Soit X une variables aléatoires à valeurs dans N . On dit que X suit une loi de Poisson de paramètre λ avec $\lambda \in R^+$, notée $P(\lambda)$ si :

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

Notation $X \sim \mathcal{P}(\lambda)$.

On sait que la fonction exponentielle a un développement en série entière avec rayon de convergence infini. En particulier :

$$\forall \lambda > 0, e^\lambda = \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!}$$

On a donc:

$$\sum_{k=0}^{+\infty} P(X = k) = e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^\lambda = 1$$

III.4.2- Les caractéristiques

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

Remarque : soient X et Y deux variables aléatoires indépendantes :

$$\left\{ \begin{array}{l} X \rightarrow P(\lambda) \\ Y \rightarrow P(\mu) \end{array} \right\} \text{ donc } (X + Y) \rightarrow P(\lambda + \mu)$$

Exemple:

A un guichet d'une poste, on sait que le nombre moyen de clients par heure est 12. On suppose que le nombre de clients par heure est distribué selon une loi de Poisson.

- a) Quelle est la probabilité pour qu'en une heure, le guichetier s'occupe de plus de 15 clients.

III.5- Loi géométrique

On répète continuellement et de façon indépendante une épreuve de Bernoulli dont la probabilité de succès est P. Soit X le nombre d'épreuves nécessaires pour obtenir un premier succès. Alors X suit une loi géométrique de paramètre p, dénoté $X \rightarrow \text{Geom}(p)$. Avec $R_x = \{1, 2, \dots, n\}$.

Il est à signaler que le nombre d'épreuves n'est pas fixé à l'avance : on s'arrête lorsque le succès est obtenu pour la première fois. Ainsi le caractère aléatoire n'est pas le nombre de succès, mais le nombre d'épreuves : X représente le nombre d'épreuves nécessaires au premier succès.

Loi de probabilité :

$X \sim \text{Geom}(p)$ ou $X \sim G(p)$ ssi :

$$P(X = x) = (1 - p)^{(x-1)} \cdot p ; \text{ pour tout } x = 1, 2, \dots, n$$

La fonction de répartition d'une variable aléatoire qui suit la loi géométrique est :

$$F_X(x) = \begin{cases} 1 - (1 - p)^a & \text{si } x \in [a, a + 1[\text{ quelque soit } a \in \mathbb{N} \text{ et } a \geq 1 \\ 0 & \text{si non} \end{cases}$$

Les caractéristiques

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

$$V(X) = \frac{1-p}{p^2}$$

III.6- Loi hypergéométrique

On considère une population de taille N dont N_1 individus présentent exactement un certain caractère A . On prélève sans remise un échantillon de n individus. Soit X la variable aléatoire qui représente le nombre d'individus ayant le caractère A dans l'échantillon.

On dit que X suit la loi hypergéométrique de paramètre N , n et $\frac{N_1}{N}$:

$$X \sim H\left(N, n, \frac{N_1}{N}\right)$$

Loi de probabilité :

La loi de probabilité de X est donnée par :

$$P(X = k) = \frac{c_{N_1}^k c_{N-N_1}^{n-k}}{c_N^n}$$

Avec $\max(0; n - (N - N_1)) \leq k \leq \min(n, N_1)$

Les caractéristiques :

$$\text{Si } p = \frac{N_1}{N}$$

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

$$V(X) = n \cdot p \cdot (1 - p) \left(\frac{N-n}{N-1} \right)$$

III.7- Approximation de la loi binomiale par une loi de Poisson

Soit X une variable aléatoire discrète.

$$X \rightarrow B(n, p)$$

si $n \geq 50; p < 0.1$ et $np \leq 10$ alors

$$B(n, p) \approx P(np) \text{ avec } np = \lambda$$

IV- Variables aléatoires continues

IV.1- Définition

Une variable aléatoire continue est une variable qui prend ses valeurs dans un intervalle de \mathbb{R} .

IV.2- Fonction de répartition

Soit X une variable aléatoire, on appelle fonction de répartition de X la fonction définie sur \mathbb{R} par $F(x) = P(X \leq x)$.

Propriété 1 La définition nous permet d'écrire :

- $F(x) = P(X \in]-\infty ; x])$.
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$.
- $P(X > b) = 1 - P(X \leq b) = 1 - F(b)$.

Remarque On admet que pour une variable aléatoire continue, pour tout $a \in \mathbb{R}$:

- $P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b)$,
- $P(a < X) = P(a \leq X < b)$,
- $P(X > b) = P(X \geq b)$.

Propriété 2 La fonction de répartition F d'une variable aléatoire continue X a les propriétés suivantes :

- F est une fonction croissante, définie et continue sur \mathbb{R} .
- Pour tout $x \in \mathbb{R}$, $0 \leq F(x) \leq 1$.
- $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

Densité et loi de probabilité

Définition Dans le cas où F est dérivable, la fonction f dérivée de F est appelée densité de probabilité de X et pour tout x de \mathbb{R} , $F'(x) = f(x)$.

IV.3- Espérance et variance

Définition Dans le cas d'une variable aléatoire continue X , sous réserve que les intégrales généralisées existent on a :

$$\begin{aligned}
E(X) &= \int_{-\infty}^{+\infty} xf(x) dx . \text{ et } V(X) \\
&= E(X)^2 - [E(X)]^2 \text{ Erreur ! Signet non défini. et } \sigma(X) \\
&= \sqrt{V(X)} \text{ Erreur ! Signet non défini..}
\end{aligned}$$

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

Exemple : Soit f la fonction définie sur \mathbb{R} par :

$$f(x) = \frac{1}{2} e^{\frac{1}{2}x}$$

Montrer que f est une densité de probabilité.

Calculer l'espérance de la variable aléatoire continue précédente.

V- Loix usuelles continues

V.1- Loi Uniforme

V.1.1- Définition

La variable aléatoire X suit une loi uniforme sur $[a, b]$, $a < b$, notée $U_{[a, b]}$ si sa fonction de densité $f(x)$ est donnée par :

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } x \in [a, b] \\ 0 & \text{si non} \end{cases}$$

V.1.2- Les caractéristiques

$$E(X) = \int_{-\infty}^{+\infty} Xf(X) dX = \frac{b+a}{2}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{(b-a) \times (b+a)}{12} = \frac{(b-a)^2}{12}$$

Fonction de répartition :

$$\forall x \in [a, b], F(x) = \int_{-\infty}^x f(t) dt = \frac{x-a}{b-a}$$

Exercice

On considère que la mesure d'un caractère biologique chez des patients suit une loi Uniforme sur $[0, 20]$. Soit X la v.a qui associe à un patient sa mesure biologique.

$$X \rightarrow U_{[0, 20]}.$$

- 1) Quelle est la probabilité d'avoir une mesure entre 14 et 18.
- 2) Quelle est la probabilité d'avoir une mesure entre 4 et 8.

V.2- Loi exponentielle

V.2.1- Définition

La loi exponentielle modélise la survie au sens large (être humains, objets...). Elle a pour paramètre $\lambda > 0$.

X : temps de survie.

On dit que la variable aléatoire X est distribuée selon une loi exponentielle de paramètre $\lambda > 0$, si l'ensemble de ses valeurs possibles est $\Omega_x =]0, +\infty[$ et sa fonction densité est définie par :

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{si } x > 0 \\ 0 & \text{si non} \end{cases}$$

On écrit $X \rightarrow \exp(\lambda)$

V.2.2- Les caractéristiques

Si $X \rightarrow \exp(\lambda)$, alors $E(X) = \frac{1}{\lambda}$

Si $X \rightarrow \exp(\lambda)$, alors $V(X) = \frac{1}{\lambda^2}$

Propriété : Soient X et Y deux variables aléatoires indépendantes.

$$\left\{ \begin{array}{l} \text{Si } X \rightarrow \exp(\lambda) \\ \text{Si } Y \rightarrow \exp(\lambda) \end{array} \right\} (X + Y) \text{ ne suit pas } \exp(\lambda)$$

Exercice

Le temps de survie de patients atteints d'un certain cancer peut être modélisé par une loi exponentielle de moyenne 2 ans.

Soit X une v.a qui associe à un patient sa durée de survie en années.

Quelle est la probabilité de survivre au moins d'un an ? 2ans ? 5ans ?

V.3- Loi Normale

Définition : On dit que la variable aléatoire X est distribuée selon une loi normale de paramètre $\mu \in \mathbb{R}$ et $\sigma^2 > 0$, si l'ensemble de ses valeurs possibles est \mathbb{R} et sa fonction densité est définie par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \forall x \in \mathbb{R} \text{ où } \sigma > 0$$

On écrit $X \rightarrow N(\mu, \sigma^2)$

Avec : $E(X) = \mu$

$V(X) = \sigma^2$

Pour $\mu = 0$; $\sigma^2 = 1$ donc X suit la loi normale centrée réduite ($X \rightarrow N(0; 1)$). La densité de probabilité correspondante :

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

La densité de la loi normale centrée réduite est symétrique par rapport à zéro par conséquent :

- Soit $F(X)$: la fonction de répartition avec

$X \rightarrow N(0; 1)$ alors :

$$p(X < -x) = p(X > x) = 1 - p(X < x) \Rightarrow$$

$$F(-x) = 1 - F(x)$$

- $p(X < x) = \alpha$, X_α est le quantile d'ordre α .
- $p(X < x) = 1 - \alpha$, $X_{1-\alpha}$ est le quantile d'ordre $1 - \alpha$
- Le quantile d'ordre α est l'opposé du quantile d'ordre $1 - \alpha$ soit : $X_\alpha = -X_{1-\alpha}$

Tables statistiques de la loi normale centrée réduite

- **Table statistique de la fonction de répartition :**

Soit $X \rightarrow N(0; 1)$, On veut déterminer la probabilité $p(X < x)$ avec x connue.

$$P(X < 1.64) = F(1.64) = 0.95.$$

$$P(X > 1.64) = 1 - F(1.64) = 0.05$$

$$P(X < -1.96) = F(-1.96)$$

$$\text{On sait que } P(X < -1.96) = P(X > 1.96) = 1 - P(X < 1.96) = 0.025.$$

Cas d'une Normale : $N(\mu, \sigma^2)$ avec $\mu \neq 0$ et/ou $\sigma^2 \neq 1$

Soit $X \rightarrow N(\mu, \sigma^2)$. Pour déterminer $P(X < x)$ où x est donnée, il faut centrer et réduire la variable X en posant $Z = \frac{X - \mu}{\sigma}$. Soit $P(X < x) = P\left(\frac{X - \mu}{\sigma} < \frac{x - \mu}{\sigma}\right) = P(Z < z)$

Table statistique des quantiles de la loi normale centrée réduite :

Soit $X \rightarrow N(0; 1)$, On veut déterminer le quantile x tel que $p(X < x) = \alpha$ avec α connu.

Exemple 1 : Soit $X \rightarrow N(0; 1)$

- 1- Déterminer x tel que :
 - $P(X < x) = 0.95$.
 - $P(X < x) = 0.05$.

2- Déterminer x_1 et x_2 tel que $p(x_1 < X < x_2) = 0.95$

et $p(X < x_1) = 0.025$

Exemple 2 : Soit $X \rightarrow N(1; 4)$

- Déterminer $P(X < x) = 0.95$.

VI- Les lois dérivées de la loi normale

VI.1- La loi du Khi deux

Soient $X_1, X_2, \dots, X_i, \dots, X_n$ des variables aléatoires indépendantes avec $X_i \rightarrow N(0, 1)$. Soit la v.a $Y = \sum_{i=1}^n X_i^2, Y \rightarrow \chi^2(n)$. La densité de Y est donnée par :

$$f_Y(y) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} & \text{si } y > 0 \\ 0 & \text{si non} \end{cases}$$

Avec $\tau\left(\frac{n}{2}\right)$ est la fonction Gamma au point $n/2$, $\tau\left(\frac{n}{2}\right) = \int_0^{+\infty} u^{\frac{n}{2}-1} e^{-u} du$

Les caractéristiques :

Soit la v.a $Y \rightarrow \chi^2(n)$

$$E(Y) = n$$

$$V(Y) = 2n$$

Propriété :

Stabilité de la loi de Khi-deux à l'addition

Table statistique des quantiles de la loi du Khi deux :

Soit $U \rightarrow \chi^2(8)$. Déterminer u telle que $P(U < u) = 0.05$. La valeur u est le quantile d'ordre 5% de la loi du Khi deux à 8 degrés de liberté, soit $u = \chi_{0.05}^2(8) = 2.733$.

La valeur u est le quantile d'ordre 5% de la loi du Khi deux à 8 degrés de liberté, soit $u = \chi_{0.05}^2(8) = 2.733$.

Soit $U \rightarrow \chi^2(12)$. Déterminer u telle que $P(U > u) = 0.05$. $P(U > u) = 1 - P(U < u) = 0.05$ donc $P(U < u) = 0.95$ et la valeur u est le quantile d'ordre 95% de la loi du Khi deux à 12 degrés de liberté, soit $u = \chi_{0.95}^2(12) = 21.026$

VI.2 La loi de Student

Soient X et Y deux variables aléatoires indépendantes, $X \rightarrow N(0, 1)$ et $Y \rightarrow \chi^2(n)$ alors la variable aléatoire $T = \frac{X}{\sqrt{\frac{Y}{n}}} \rightarrow t(n)$

Exemple : Soient $X \rightarrow N(0, 1)$ et $Y \rightarrow \chi^2(n - 1)$ donc $T = \frac{X}{\sqrt{\frac{Y}{n-1}}} \rightarrow t(n - 1)$

La densité d'une variable aléatoire T est symétrique par rapport à zéro puisque

$$f(-t) = f(t)$$

La loi de Student est symétrique :

$$P(T < -t) = P(T > t) = 1 - P(T < t) \text{ donc } F(-t) = 1 - F(t)$$

Le quantile d'ordre α est l'opposé du quantile d'ordre $1-\alpha$.

Les caractéristiques :

Soit $T \rightarrow t(n)$

$$E(T) = 0$$

$$V(T) = \frac{n}{n-2}, n' \text{ existe que pour } n > 2$$

Table statistique des quantiles de la loi de Student :

Soit $T \rightarrow t(7)$. Déterminer le quantile t tel que $P(T < t) = 0.05$.

$$t = t_{0,05}(7) = -t_{0,95}(7) = -1,895$$

Soit $T \rightarrow t(15)$ Déterminer le quantile t tel que $P(T < t) = 0.95$

$$\text{donc } t = t_{0,95}(15) = 1,753$$

VI.3 La loi de Fisher

Soient X et Y deux variables aléatoires indépendantes avec $X \rightarrow \chi^2(n_1)$ et $Y \rightarrow \chi^2(n_2)$ alors

La variable $F = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \rightarrow F(n_1, n_2)$

F est distribué selon une loi de Fisher à (n_1, n_2) degrés de liberté.

$$\text{On a } P(F > F_p(n_1, n_2)) = p$$

$$\text{Donc } P(F < F_p(n_1, n_2)) = 1 - p$$

$F_p(n_1, n_2)$ est le quantile d'ordre $1 - p$

Les caractéristiques :

$$E(F) = \frac{n_2}{n_2-2} \text{ pour } n_2 > 2$$

$$V(F) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} \text{ pour } n_2 > 4$$

Remarque :

$$\text{Si } w \rightarrow F(n_1, n_2) \text{ donc } U = \frac{1}{W} = \frac{\frac{Y}{n_2}}{\frac{X}{n_1}} \rightarrow F(n_2, n_1)$$

$$\text{Donc } F_p(n_2, n_1) = \frac{1}{F_{1-p}(n_1, n_2)} \text{ avec } F_p(n_2, n_1) \text{ est un quantile.}$$

Une variable de Fisher $F(1, n_2)$ est équivalente au carré d'une variable de student $t(n_2)$.

$$\text{Si } F \rightarrow F(1, n_2) \text{ donc } F = \frac{\frac{X}{1}}{\frac{Y}{n_2}} = \frac{X}{\frac{Y}{n_2}} = \frac{T^2}{\frac{Y}{n_2}} \text{ où } T \rightarrow N(0, 1)$$

$$\text{Donc } F = \left(\frac{T}{\sqrt{\frac{Y}{n_2}}} \right)^2$$

Table statistique de la loi de Fisher

Soit $z \rightarrow F(n_1, n_2)$, déterminer $z/P(Z > z) = p$

Exemple

- 1) Soit $Z \rightarrow F(8, 20)$, Déterminer $z/P(Z > z) = 0,01$ donc $z = F_{0,01}(8, 20) = 3,564$
- 2) Soit $Z \rightarrow F(30, 25)$, Déterminer $z/P(Z > z) = 0,05$ donc $z = F_{0,05}(30, 25) = 1,919$
- 3) Soit $Z \rightarrow F(30, 25)$, Déterminer z_1 et z_2 tel que $P(z_1 < Z < z_2) = 0,98$
et $P(Z > z_2) = 0,01$ donc $z_2 = 2,538 = F_{0,01}(30, 25)$

$$\begin{aligned} P(z_1 < Z < z_2) &= P(Z < z_2) - P(Z < z_1) \\ &= [1 - P(Z > z_2) - P(Z < z_1)] = 0,98 \end{aligned}$$

$$\text{Donc } P(Z < z_1) = 0,99 - 0,98 = 0,01 \text{ Donc } P(Z > z_1) = 1 - 0,01 = 0,99$$

$$\text{Donc } z_1 = F_{0,99}(30, 25) = \frac{1}{F_{0,01}(25, 30)} = \frac{1}{2,453} = 0,407$$

VII- Théorème central limite

Soient X_1, X_2, \dots, X_n n variables aléatoires indépendantes et de même loi quelconque connue ou non telles que $E(X_i) = m$ et $V(X_i) = \sigma^2, i = 1, \dots, n$ et soit $S_n = \sum_{i=1}^n X_i$. Alors quelque soit le réel X , on dit que $Y_n = \frac{S_n - E(S_n)}{\sqrt{V(S_n)}} = \frac{S_n - n.m}{\sigma\sqrt{n}}$ converge en loi vers la loi $N(0, 1)$.

Le théorème central limite affirme que dans le cas général, la loi $N(0, 1)$ est une bonne approximation de la loi de Y_n , à condition que n soit assez grand.

Remarque :

Le théorème central limite est utilisé aussi pour l'approximation de certaines probabilités.

Exemple 1:

Soient X_1, X_2, \dots, X_n n variables aléatoires indépendantes et de même loi avec $E(X_i) = 1$ et $V(X_i) = 4$ et $E(X_i^4) = 34$ pour $i = 1, \dots, n$

1- Montrer que

$$T_n = \frac{\sum_{i=1}^n X_i - n}{2\sqrt{n}} \text{ converge en loi vers } N(0, 1) \text{ et } W_n =$$

$$\frac{\sum_{i=1}^n X_i^2 - 5n}{3\sqrt{n}} \text{ converge en loi vers } N(0, 1)$$

2- En déduire, pour $n = 100$, une approximation des probabilités $P(\sum_{i=1}^n X_i < 125)$ et $P(\sum_{i=1}^n X_i^2 < 455)$

Exemple 2 :

Soient X_1, X_2, \dots, X_n n variables aléatoires indépendantes et de même loi avec $E(X_i) = m$ et $V(X_i) = \sigma^2$.

Montrer que $\sqrt{n}(\bar{X}_n - m)$ converge en loi vers $N(0, \sigma^2)$ avec $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

Chapitre 2 Echantillonnage

Nous supposons dans la suite que l'on procède à un échantillonnage aléatoire, c'est-à-dire qu'on a tiré au sort un échantillon aléatoire de taille n parmi une population de taille N . Le tirage peut être effectué de deux façons donnant lieu à deux types d'échantillons.

Un échantillon est dit exhaustif s'il est constitué sans remise et non exhaustif s'il est constitué avec remise. Dans le cas d'un tirage non exhaustif, il y a indépendance entre les tirages. Dans le cas contraire, il n'y a pas d'indépendance entre les tirages.

I- Echantillon aléatoire et ses caractéristiques

On considère une machine qui produit des articles. On prélève un échantillon de n articles.

$$\text{Soit } \begin{cases} X_i = 1 \text{ si l'article produite est defectueuse} \\ 0 \text{ si non} \end{cases}$$

La séquence de variables aléatoires X_1, X_2, \dots, X_N s'interprète comme un échantillon aléatoire de la population étudiée selon la variable X .

Cet échantillon possède certaines caractéristiques :

I.1- Moyenne empirique

I.1.1- Définition

Si X_1, X_2, \dots, X_N est un échantillon indépendant et identiquement distribué (i.i.d), alors la moyenne empirique, notée \bar{X}_n , est défini par :

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

La moyenne empirique \bar{X}_n est une statistique, donc c'est une variable aléatoire dont les propriétés (espérance mathématique et variance) seront déduites à partir de celles des variables de l'échantillon.

I.1.2- La distribution de \bar{X}_n : cas d'un échantillon issu d'une population normale

Soit X_1, X_2, \dots, X_N est un échantillon aléatoire indépendant et identiquement distribué avec $X_i \rightarrow N(\mu, \sigma^2)$ ou μ et σ^2 sont inconnus.

Déterminer la loi de $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$

on a $X_i \rightarrow N(\mu, \sigma^2)$ donc $\sum_i X_i \rightarrow N(n \cdot \mu; n \cdot \sigma^2)$

donc $\frac{\sum_i X_i}{n} \rightarrow N\left(\mu; \frac{\sigma^2}{n}\right)$ donc $\bar{X}_n \rightarrow N\left(\mu; \frac{\sigma^2}{n}\right)$ ou encore

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\text{loi}} N(0, 1)$$

I.1.3- Distribution asymptotique de \bar{X}_n : cas d'un échantillon quelconque

Considérons l'échantillon aléatoire X_1, X_2, \dots, X_N indépendant et identiquement distribuée selon une loi quelconque, $E(X) = m$; $V(X) = \sigma^2$. Pour une taille de l'échantillon assez grande (généralement $n > 30$), on utilise le théorème central limite pour déduire la distribution limite de la moyenne empirique, soit :

$$\frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\text{loi}} N(0, 1)$$

Propriétés :

$$E(\bar{X}_n) = m$$

$$V(\bar{X}_n) = \frac{\sigma^2}{n}$$

Si $n \rightarrow +\infty$ alors $V(\bar{X}_n) \rightarrow 0$

Le tableau suivant donne la distribution limite de la moyenne empirique \bar{X}_n pour différentes lois.

Tableau 1 la distribution limite de la moyenne empirique \bar{X}_n pour différentes lois.

| X_1, X_2, \dots, X_N Indépendantes et de même loi | $E(X_i)$ | $V(X_i)$ | Distribution limite |
|--|--------------------|----------------------|--|
| Bernoulli : $X_i \rightarrow B(p)$ | p | $p(1 - p)$ | $T_n = \frac{\bar{X}_n - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{loi} Z \rightarrow N(0, 1)$ |
| Binomiale : $X_i \rightarrow B(K, p)$ | kp | $kp(1 - p)$ | $T_n = \frac{\bar{X}_n - kp}{\sqrt{\frac{kp(1-p)}{n}}} \xrightarrow{loi} Z \rightarrow N(0, 1)$ |
| Poisson : $X_i \rightarrow P(\lambda)$ | λ | λ | $T_n = \frac{\bar{X}_n - \lambda}{\sqrt{\frac{\lambda}{n}}} \xrightarrow{loi} Z \rightarrow N(0, 1)$ |
| Uniforme continue : $X_i \rightarrow U_{[a,b]}$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | $T_n = \frac{\bar{X}_n - \left(\frac{a+b}{2}\right)}{\sqrt{\frac{(b-a)^2}{12n}}} \xrightarrow{loi} Z \rightarrow N(0, 1)$ |
| Exponentielle : $X_i \rightarrow \exp(\theta)$ | $\frac{1}{\theta}$ | $\frac{1}{\theta^2}$ | $T_n = \frac{\bar{X}_n - \left(\frac{1}{\theta}\right)}{\sqrt{\frac{\left(\frac{1}{\theta}\right)^2}{n}}} \xrightarrow{loi} Z \rightarrow N(0, 1)$ |
| Exponentielle : $X_i \rightarrow \exp\left(\frac{1}{\theta}\right)$ | θ | θ^2 | $T_n = \frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta^2}{n}}} \xrightarrow{loi} Z \rightarrow N(0, 1)$ |

I.2- La variance empirique

I.2.1- Définition

Soit X_1, X_2, \dots, X_N est un échantillon indépendant et identiquement distribué d'une population de moyenne m et de variance σ^2 . La variance σ^2 est inconnue.

1^{er} Cas : m est connu

Dans ce cas la variance empirique, notée δ_n^2 est définie ainsi :

$$\delta_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$$

2^{eme} Cas : m est in connu

Dans ce cas la variance empirique, notée \bar{S}_n^2 est définie par :

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

La variance empirique corrigée (variance sans biais) :

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \quad \text{ou} \quad S_n^2 = \left(\frac{n}{n-1} \right) \bar{S}_n^2$$

Les caractéristiques de la variance :

$E(S_n^2) = \sigma^2$ Donc tous les échantillons ont la même espérance de la variance

$$V(S_n^2) = \frac{2\sigma^4}{n-1}$$

$$E(\bar{S}_n^2) = \left(\frac{n-1}{n} \right) \sigma^2 = \left(1 - \frac{1}{n} \right) \sigma^2 \text{ Donc tous les échantillons n'ont pas le même}$$

Esperance de la variance.

Si $n \rightarrow +\infty$ Donc $E(\bar{S}_n^2) = \sigma^2$ Donc le biais diminue

I.2.2- Distribution de S_n^2 : cas d'un échantillon issu d'une population normale

Soit X_1, X_2, \dots, X_N est un échantillon aléatoire indépendant et identiquement distribué avec $X_i \rightarrow N(\mu, \sigma^2)$ ou μ et σ^2 sont inconnus donc nous avons :

$$\frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{\sigma^2} \rightarrow \chi^2(n-1) \quad \text{donc} \quad \frac{(n-1)S_n^2}{\sigma^2} \rightarrow \chi^2(n-1)$$

I.3- Fréquence

I.3.1- Définition

Soit X_1, X_2, \dots, X_N est un échantillon aléatoire de taille n avec $X \rightarrow B(p)$

La fréquence $F = \frac{\sum_{i=1}^n X_i}{n}$ avec :

$$E(F) = p$$

$$V(F) = \frac{p \cdot q}{n}$$

I.3.2- Distribution de la fréquence

Si la taille de l'échantillon ($n \geq 30$) , on fait recours au théorème central limite alors

$$F \rightarrow N\left(p, \sqrt{\frac{p \cdot q}{n}}\right) \text{ donc } \frac{F-p}{\sqrt{\frac{p \cdot q}{n}}} \rightarrow N(0, 1)$$

II- Estimateur et ses propriétés statistiques

L'échantillonnage nous permet de déterminer des caractéristiques importantes d'une variable aléatoire à savoir l'espérance mathématique et la variance. Dans la pratique, on connaît souvent à l'avance la forme de la loi de la variable étudiée et il nous reste à déterminer certains paramètres par exemple si X suit la loi binomiale de paramètre p inconnu, on doit estimer ce paramètre.

Dans cette section nous allons définir un estimateur et nous allons étudier ses propriétés statistiques. Les méthodes utilisées pour déterminer un estimateur d'un paramètre seront présentées dans le chapitre suivant.

II.1- Définition d'un estimateur

Soit X_1, X_2, \dots, X_n un échantillon aléatoire indépendant et identiquement distribué de même loi que X ; $X \rightarrow L(\theta)$; ou la loi L est supposé connu et le paramètre θ est inconnu. Un estimateur de θ , noté $\hat{\theta}$, est une statistique $\hat{\theta} = T(X_1, \dots, X_n)$. Toute valeur $T_n(X_1, X_2, \dots, X_n)$ de cet estimateur $\hat{\theta}$ est appelée une estimation.

II.2- Propriétés statistiques d'un estimateur

Pour qu'un estimateur puisse fournir de bonnes estimations, il doit posséder certaines propriétés. Il doit être sans biais, convergent et efficace.

II.2.1- Estimation sans biais

On dit que $\hat{\theta}$ est un estimateur sans biais du paramètre θ si $E(\hat{\theta}) = \theta$

$E(\hat{\theta})$: esperance mathématique de $\hat{\theta}$.

Remarque :

- En général nous avons $E(\hat{\theta}) = \theta + B(\theta)$, $B(\theta)$ est appelé biais de l'estimateur $\hat{\theta}$.

$\hat{\theta}$ est un estimateur sans biais de θ donc $B(\theta) = 0$.

- Un estimateur $\hat{\theta}$ de θ est dit asymptotiquement sans biais si $\lim_{n \rightarrow +\infty} B(\theta) = 0$
ou encore $\lim_{n \rightarrow +\infty} E(\hat{\theta}) = \theta$

II.2.2- Estimateur convergent

Théorème :

On dit que $\hat{\theta}$ est un estimateur convergent de θ ssi :

$$\left\{ \begin{array}{l} \lim_{n \rightarrow +\infty} E(\hat{\theta}) = \theta \\ \lim_{n \rightarrow +\infty} V(\hat{\theta}) = 0 \end{array} \right\}$$

Exemple :

Soit X_1, X_2, \dots, X_N un échantillon aléatoire indépendant et identiquement distribué de même loi que X , $X \rightarrow \beta(p)$.

Soit $\hat{p} = \bar{X}_n$ est un estimateur de p .

- 1) Montrer que \hat{p} est un estimateur convergent ?

II.2.3- Estimateur efficace

On dit que $\hat{\theta}$ est un estimateur efficace de θ si :

$$\left\{ \begin{array}{l} E(\hat{\theta}) = \theta \\ V(\hat{\theta}) = \frac{1}{I_n(\theta)} \end{array} \right\}$$

Avec $I_n(\theta)$ est appelé quantité d'information rapportée par l'échantillon.

$$I_n(\theta) = E \left[\left(\frac{\partial \ln L(X_1, \dots, X_n, \theta)}{\partial \theta} \right)^2 \right] = -E \left[\frac{\partial^2 \ln L(X_1, \dots, X_n, \theta)}{\partial^2(\theta)} \right]$$

$L(X_1, \dots, X_n, \theta)$ est appelée distribution de probabilité conjointe de X_1, \dots, X_n : elle est appelée aussi la fonction de vraisemblance.

$$L(X_1, \dots, X_n, \theta) = \begin{cases} \prod_{i=1}^n P(X_i, \theta) & \text{dans le cas discret} \\ \prod_{i=1}^n f(X_i, \theta) & \text{dans le cas continu} \end{cases}$$

II.2.4- Comparaison de deux estimateurs

Soient $\hat{\theta}_1$ et $\hat{\theta}_2$ deux estimateurs sans biais de θ . On dit que $\hat{\theta}_1$ est plus efficace que $\hat{\theta}_2$ si $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.

Chapitre 3 Méthodes d'estimation ponctuelle et Intervalle de Confiance

I- Estimation ponctuelle : Méthode du Maximum de vraisemblance

Soit X une variable aléatoire, $X \rightarrow L(\theta)$ ou θ est inconnu et définie par sa fonction de probabilité $P(X, \theta)$ dans le cas où X est discrète et par $f(X, \theta)$ dans le cas continu. On tire un échantillon X_1, X_2, \dots, X_n indépendant et identiquement distribué de X . D'après la méthode du maximum de vraisemblance (MMV), l'estimateur $\hat{\theta}_{MV}$ du paramètre inconnu θ est celui qui maximise la fonction de vraisemblance $L(X_1, \dots, X_n, \theta)$.

L'estimateur $\hat{\theta}_{MV}$ est la solution du programme :

$$\text{Max}_{\theta} L(X_1, \dots, X_n, \theta)$$

Il est plus simple de maximiser le logarithme neperien de la fonction de vraisemblance par rapport à θ .

Pour trouver l'estimateur $\hat{\theta}_{MV}$, on peut procéder ainsi :

i) Ecrire la fonction de vraisemblance

$$L(X_1, \dots, X_n; \theta) = \prod_{i=1}^n f(x_i, \theta)$$

ii) Ecrire le logarithme de la fonction de vraisemblance

$$\ln L(X_1, \dots, X_n, \theta) = \ln \prod_{i=1}^n f(x_i, \theta)$$

iii) $\frac{\partial \ln L(X_1, \dots, X_n, \theta)}{\partial \theta} = 0$ donc trouver la valeur $\hat{\theta}$ pour laquelle cette dérivée est égale à zéro.

Cette dérivée partielle peut s'annuler si :

- $X(\Omega)$ ne dépend pas de θ
- $P(X, \theta)$ et $f(X, \theta)$ sont continues et dérivables par rapport à θ .

iiii) Calculer $\frac{\partial^2 \ln L(X_1, \dots, X_n, \theta)}{\partial^2(\theta)}$ et vérifier le signe de cette dérivée seconde au point $\hat{\theta}$, si ce signe est négatif, $\hat{\theta}$ est l'estimateur du maximum de vraisemblance de θ .

I.1- Estimateur de la moyenne

Soit une population de taille N (ou indéfinie) sur laquelle est défini un caractère quantitatif représenté par une variable aléatoire X d'espérance mathématique μ et d'écart type σ .

On prélève un échantillon de taille n , nous créons une suite de n variables aléatoires indépendantes, de même distribution que X ; notées X_1, \dots, X_n et prenant respectivement pour valeurs les valeurs prises par X sur chacun des n individus de l'échantillon.

Définition :

On définit la variable aléatoire \bar{X}_n comme la moyenne de l'échantillon

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

Avec :

$E(\bar{X}_n) = \mu$ donc La moyenne d'échantillon est un estimateur sans biais de la moyenne de la population.

La moyenne de la variable aléatoire \bar{X} est toujours égale à la moyenne de la population mère, celle d'où l'échantillon a été prélevé.

$$V(\bar{X}_n) = \frac{\sigma^2}{n}, \text{ on déduit } \sigma(\bar{X}_n) = \frac{\sigma}{\sqrt{n}}$$

On note que l'écart-type de la variable \bar{X} diminue quand la taille n de l'échantillon augmente.

Théorème :

La moyenne empirique $\bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$ est un estimateur sans biais et convergent de la moyenne de la population μ .

I.2- Estimation d'une proportion théorique

Dans une population, une proportion inconnue $p \in [0; 1]$ d'unités possèdent un caractère donné C. Nous nous proposons d'estimer cette proportion.

Soit X la variable aléatoire : $X = \begin{cases} 1 & \text{si l'unité choisie possède le caractère C} \\ 0 & \text{si non} \end{cases}$

On tire un échantillon de taille n . Soit \hat{p} la proportion d'individus ayant le caractère C dans cet échantillon.

$$\hat{p}(X) = \bar{X}_n = \frac{\sum_{i=1}^n X_i}{n}$$

$$E(\hat{p}) = p$$

$$V(\hat{p}) = \frac{p(1-p)}{n}$$

Théorème :

La proportion \hat{p} est un estimateur sans biais et convergent de p .

Remarque :

L'estimation d'une proportion p est un cas particulier de la moyenne empirique au sens où les variables aléatoires X_i considérées sont de Bernouilli de paramètre p .

I.3- Estimation du paramètre d'une loi de Poisson

Soit y_1, \dots, y_n un échantillon i.i.d d'une v.a $y \rightarrow P(\lambda)$

Déterminer l'estimateur du paramètre λ :

$$\hat{\lambda}_{MV} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}_n$$

Les propriétés statistiques de $\hat{\lambda}_{MV}$:

$$E(\hat{\lambda}_{MV}) = E(\bar{y}_n) = \lambda$$

$$V(\hat{\lambda}_{MV}) = V(\bar{y}_n) = \frac{\lambda}{n}, \text{ si } n \rightarrow +\infty \text{ donc } V(\hat{\lambda}_{MV}) \rightarrow 0$$

Donc $\hat{\lambda}_{MV}$ est un estimateur sans biais et convergent de λ .

$$I_n(\lambda) = \frac{1}{V(\hat{\lambda}_{MV})} \text{ Donc } \hat{\lambda}_{MV} \text{ est un estimateur efficace de } \lambda.$$

I.2- Estimation de la variance σ^2

I.2.1- Estimation de σ^2 en supposant μ connu

Théorème :

Lorsque la moyenne μ de la population est connu on a :

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \text{ est un estimateur sans biais et convergent de } \sigma^2 .$$

$$E(\hat{\delta}_n^2) = \sigma^2$$

$$Var(\hat{\delta}_n^2) = \frac{[\mu_4 - \sigma^4]}{n} \text{ avec } \mu_k = E[(X - \mu)^k]$$

Lorsque n est assez grand, le théorème central limite nous permet de déterminer la loi asymptotique de $\hat{\delta}_n^2$, on a : $\hat{\delta}_n^2 \rightarrow N\left(\sigma^2, \frac{(\mu_4 - \sigma^4)}{n}\right)$

I.2.2- Estimation de σ^2 en supposant μ inconnu

En général on ne connaît pas μ ; on le remplace par un estimateur et on introduit la variance empirique associée.

$$\bar{S}_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}, \quad \bar{S}_n^2 : \text{la variance calculée sur l'échantillon.}$$

Théorème :

La variance empirique \bar{S}_n^2 est un estimateur biaisé et convergent de σ^2 . Il est asymptotiquement sans biais.

Avec :

$$E(\bar{S}_n^2) = \left(\frac{n-1}{n}\right) \sigma^2$$

Théorème :

La variance empirique corrigée :

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ou $S_n^2 = \left(\frac{n}{n-1}\right) \bar{S}_n^2$ est un estimateur sans biais et convergent de la variance de la population (σ^2).

II- Estimation par intervalle de confiance

L'estimation ponctuelle $\hat{\theta}$ n'est qu'une valeur approchée du paramètre θ même dans le cas où l'estimateur obtenu possède de bonnes propriétés statistiques (sans biais, convergent et efficace). Puisqu'on ne peut s'attendre à ce qu'une estimation ponctuelle soit exactement égale à la valeur du paramètre de la population correspondant. Il serait plus pertinent de déterminer un intervalle par exemple de la forme $[L_1, L_2] = [\hat{\theta} - \Delta; \hat{\theta} + \Delta]$ qui contient la

valeur véritable de θ avec une probabilité donnée à l'avance et proche de l'unité. Une estimation par intervalle est souvent réalisée en ajoutant et en soustrayant une marge d'erreur à l'estimation ponctuelle. La forme générale d'une estimation par intervalle est : estimation ponctuelle \pm marge d'erreur.

Un intervalle de confiance de niveau $(1 - \alpha)$ pour θ est un intervalle $[L_1, L_2]$ tel que $p(L_1 \leq \theta \leq L_2) = 1 - \alpha$, θ est une constante inconnue.

La construction des intervalles de confiance ; de niveau $(1 - \alpha)$ donné à priori ; est basée essentiellement sur le choix adéquat d'une fonction de $\hat{\theta}$ et θ .

Soit $Q(\hat{\theta}, \theta)$ telle que :

- Sa loi est connue (normale, Khi-deux, Student, Fisher).
- C'est une fonction pivotale pour θ c.a.d sa loi ne dépend pas de θ .

Les étapes de construction d'un intervalle de confiance

- 1) Etant donné un niveau de confiance $(1 - \alpha)$ proche de 1, écrire $p(L_1 \leq \theta \leq L_2) = 1 - \alpha$
- 2) Chercher la statistique $Q(\hat{\theta}, \theta)$.
- 3) Ecriture de l'équation permettant de retrouver l'intervalle :
 $p(C_1 \leq Q(\hat{\theta}, \theta) \leq C_2) = 1 - \alpha$ où C_1 et C_2 sont les quantiles d'ordre respectifs α_1 et $1 - \alpha_2$ de la loi de $Q(\hat{\theta}, \theta)$, avec $\alpha_1 + \alpha_2 = \alpha$. Généralement on prend $\alpha_1 = \alpha_2 = \frac{\alpha}{2}$.
- 4) Détermination de C_1 et C_2 à partir d'une table statistique (loi normale, Khideux, Student)
- 5) Détermination de L_1 et L_2 par résolution de l'inéquation de la troisième étape par rapport au paramètre θ afin de ramener à l'écriture $p(L_1 \leq \theta \leq L_2) = 1 - \alpha$ où L_1 et L_2 seront en fonction de C_1 et C_2 .

II.1- Intervalle de confiance de la moyenne dans le cas d'une loi normale

Considérons X une variable aléatoire, $X \rightarrow N(\mu; \sigma^2)$ et soit $X_1; \dots; X_n$ un échantillon i.i.d de X . Avec $n \leq 30$.

Nous voulons déterminer un intervalle de confiance de niveau $(1 - \alpha)$ pour la moyenne inconnue μ .

II.1.1- Cas ou la variance σ^2 est connue

On sait que $\hat{\mu} = \bar{X}_n \xrightarrow{\text{loi}} N\left(\mu; \frac{\sigma^2}{n}\right)$ donc la statistique $Q(\bar{X}_n; \mu) = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0; 1)$

Théorème : $X \rightarrow N(\mu; \sigma^2)$, σ^2 est connue. L'intervalle de confiance de niveau $(1 - \alpha)$ pour la moyenne de la population est :

$$IC(\mu) = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}}; \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\frac{\alpha}{2}} \right]$$

Cas d'un intervalle unilatéral :

- L'intervalle de confiance de la forme $[L; +\infty[$ tel que $P(\mu \geq L) = 1 - \alpha$ donc

$$IC(\mu) = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} z_{1-\alpha}; +\infty \right]$$

- L'intervalle de confiance unilatéral de la forme $] -\infty; L]$ est tel que $P(\mu \leq L) = 1 - \alpha$

$$\text{Donc } IC(\mu) = \left] -\infty; \bar{X}_n + \frac{\sigma}{\sqrt{n}} z_{1-\alpha} \right]$$

II.1.2- Cas ou la variance σ^2 est inconnue

D'après les propriétés de la moyenne empirique dans le cas de normalité on a

$$\bar{X}_n \rightarrow N\left(\mu; \frac{\sigma^2}{n}\right) \text{ donc } \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0; 1)$$

$$\text{On sait que } \frac{(n-1)S_n^2}{\sigma^2} \rightarrow \chi^2(n-1)$$

$$\text{Donc } \frac{\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2} \cdot \frac{1}{n-1}}} = \frac{\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}}{\frac{S_n}{\sigma}} = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \rightarrow t(n-1)$$

Théorème : Soit $X \rightarrow N(\mu; \sigma^2)$, σ^2 est inconnue. L'intervalle de confiance de niveau $(1 - \alpha)$ pour la moyenne μ est :

$$IC(\mu) = \left[\bar{X}_n - t_{1-\frac{\alpha}{2}}(n-1) \frac{S_n}{\sqrt{n}}; \bar{X}_n + t_{1-\frac{\alpha}{2}}(n-1) \frac{S_n}{\sqrt{n}} \right]$$

II.2- Intervalle de confiance pour la proportion d'une loi de Bernoulli

Soit X_1, X_2, \dots, X_n est un échantillon indépendant et identiquement distribué d'une variable aléatoire X qui suit la loi de Bernoulli de paramètre p . La taille de l'échantillon est égale à $n > 30$.

Déterminer un intervalle de confiance $[L_1; L_2]$ de niveau $(1 - \alpha)$ pour la proportion p tel que $p(L_1 \leq p \leq L_2) = 1 - \alpha$.

\hat{p} est un estimateur sans biais de p

\hat{p} converge en loi vers la loi $N\left(p; \frac{p(1-p)}{n}\right)$ si $n.p \geq 5$ et $n(1-p) \geq 5$

D'après T.C.L $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\text{loi}} N(0; 1)$. Cette statistique ne peut pas être utilisée pour la construction de l'intervalle de confiance puisque les bornes dépendront du paramètre inconnu p . Alors on remplace p par \hat{p} . Nous aurons :

$$Q(\hat{p}, p) = \frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \xrightarrow{\text{loi}} N(0; 1)$$

L'intervalle de confiance de niveau $(1 - \alpha)$ pour la proportion p d'une loi de Bernoulli est :

$$IC(p) = \left[\hat{p} - \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{1-\frac{\alpha}{2}}; \hat{p} + \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} z_{1-\frac{\alpha}{2}} \right]$$

Exemple

Le commanditaire d'une émission télévisée veut déterminer un intervalle de confiance de niveau 95% pour la proportion des téléspectateurs qui regardent régulièrement son émission.

Parmi un échantillon de 900 téléspectateurs pris au hasard, 200 affirment qu'ils regardent régulièrement l'émission.

Déterminer l'intervalle de confiance de niveau de 95% pour la proportion p .

II.3- Intervalle de confiance pour la moyenne d'une loi inconnue

Soit X_1, X_2, \dots, X_n est un échantillon indépendant et identiquement distribué d'une variable aléatoire X dont la loi est inconnue. $E(X) = m$ et $V(X) = \sigma^2$. La taille de l'échantillon est égale à $n > 30$.

Déterminer un intervalle de confiance pour la moyenne m de la population.

Selon le théorème central limite \bar{X}_n est un estimateur de m .

Nous avons deux cas :

- La variance σ^2 est connue.
- La variance σ^2 est inconnue.

II.3.1- Cas où la variance σ^2 est connue

D'après le théorème central limite : $Q(\hat{\theta}, \theta) = Q(\bar{X}_n, m) = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0, 1)$

$$\text{Donc } p\left(-z_{1-\frac{\alpha}{2}} \leq \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \leq z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$\text{Donc } p\left(\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} \leq m \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$$

Théorème :

Lorsque σ^2 est connu, l'intervalle de confiance de niveau $(1 - \alpha)$ pour la moyenne m de la population est :

$$IC(m) = \left[\bar{X}_n - \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}}, \quad \bar{X}_n + \frac{\sigma}{\sqrt{n}} \cdot z_{1-\frac{\alpha}{2}} \right]$$

II.3.2- Cas où la variance σ^2 est inconnue

Dans ce cas on doit remplacer σ^2 par son estimateur sans biais :

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ ou $S_n^2 = \left(\frac{n}{n-1}\right) \bar{S}_n^2$, $\bar{S}_n^2 = \frac{\sum_{i=1}^n (X_i - \bar{X}_n)^2}{n}$, \bar{S}_n^2 : la variance calculée sur l'échantillon.

D'après le théorème central limite : $Q(\hat{\theta}, \theta) = Q(\bar{X}_n, m) = \frac{\bar{X}_n - m}{\frac{S_n}{\sqrt{n}}} \rightarrow N(0, 1)$

Théorème :

Lorsque σ^2 est inconnu, l'intervalle de confiance de niveau $(1 - \alpha)$ pour la moyenne m de la population est :

$$\left[\bar{X}_n - z_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}}; \bar{X}_n + z_{1-\frac{\alpha}{2}} \frac{S_n}{\sqrt{n}} \right]$$

Exemple

A la buvette d'un établissement la dépense globale par jour de 100 étudiants, tirés au hasard de la population estudiantine d'un établissement, est de 125 dinars. Ainsi la somme des carrés des dépenses individuelles est égale à 172.

- 1) Déterminer un intervalle de confiance de niveau de confiance de 90% pour la dépense moyenne dans la population.

II.4- Intervalle de confiance de la variance d'une population normale**II.4.1- Estimation de la variance quand la moyenne est connue**

On considère la variable aléatoire X suit la loi normale de moyenne μ et de variance σ^2

Lorsque la moyenne μ est connue, l'estimateur sans biais et convergent de σ^2 est :

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

La statistique $Q(\hat{\sigma}^2; \sigma^2) = \frac{n\hat{\sigma}^2}{\sigma^2} \rightarrow \chi^2(n)$

un intervalle de confiance de niveau $(1 - \alpha)$ de σ^2 est : $\left[\frac{n\hat{\sigma}^2}{U_2}; \frac{n\hat{\sigma}^2}{U_1} \right]$

Où U_1 est le quantile d'ordre $\frac{\alpha}{2}$ de la loi de $\chi^2(n)$.

U_2 est le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi de $\chi^2(n)$.

Avec $p(\chi^2 \leq U_1) = \frac{\alpha}{2}$ et $p(\chi^2 \geq U_2) = \frac{\alpha}{2}$

II.4.2- Estimation de la variance quand la moyenne est inconnue

Lorsque μ est inconnu, $S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x}_n)^2$ est un estimateur sans biais de σ^2 telle

$$\text{que : } Q(S_n^2, \sigma^2) = \frac{(n-1)S_n^2}{\sigma^2} \rightarrow \chi^2(n-1)$$

L' intervalle de confiance de niveau $(1 - \alpha)$ de σ^2 est :

$$\left[\frac{(n-1)S_n^2}{U_2}, \frac{(n-1)S_n^2}{U_1} \right]$$

Où U_1 est le quantile d'ordre $\frac{\alpha}{2}$ de la loi de $\chi^2(n-1)$.

U_2 est le quantile d'ordre $\left(1 - \frac{\alpha}{2}\right)$ de la loi de $\chi^2(n-1)$.

Chapitre 4 Tests d'hypothèses

I- Principe des tests

Le principe des tests d'hypothèse est de poser une hypothèse de travail et de prédire les conséquences de cette hypothèse pour la population ou l'échantillon. On compare ces prédictions avec les observations et l'on conclut en acceptant ou en rejetant l'hypothèse de travail à partir de règles de décision objectives. Il faut donc être très attentif à la formulation des hypothèses, afin d'être sûr qu'elles sont appropriées et que les conclusions du test d'hypothèses fournissent bien les informations souhaitées par le chercheur.

Pour tester une hypothèse, différentes étapes doivent être suivies :

- 1- Formuler les hypothèses du test H_0 et H_1 .
- 2- Fixer le seuil de signification α .
- 3- Choisir un test statistique ou une statistique pour contrôler H_0 .
- 4- Définir la distribution de la statistique sous l'hypothèse « H_0 est réalisée ».
- 5- Définir le niveau de signification du test ou région critique notée α .
- 6- Calculer, à partir des données fournies par l'échantillon, la valeur de la statistique.
- 7- Prendre une décision concernant l'hypothèse posée et faire une interprétation.

I.1- Formulation des hypothèses à tester

L'hypothèse nulle : elle consiste à dire qu'il n'existe pas de différence entre les paramètres comparés. Elle est notée H_0 . C'est une hypothèse de travail.

L'hypothèse alternative : est la négation de H_0 . Cette hypothèse correspond à l'hypothèse de recherche.

Considérons une population dont on veut étudier selon un caractère (variable aléatoire) X suivant une loi (connue ou non) de paramètre inconnu θ et soit X_1, X_2, \dots, X_n un échantillon i.i.d de X .

Dans ce cas le test d'hypothèse peut prendre deux formes :

$\begin{cases} H_0: \theta = \theta_0 \\ H_1: \theta \neq \theta_0 \end{cases}$: il s'agit d'un test bilatéral.

$\begin{cases} H_0 = \theta \leq (=) \theta_0 \\ H_1 = \theta > \theta_0 \end{cases}$ Il s'agit d'un test unilatéral à droite.

ou encore

$\left\{ \begin{array}{l} H_0 = \theta \geq (=) \theta_0 \\ H_1 = \theta < \theta_0 \end{array} \right\}$ Il s'agit d'un test unilatéral à gauche.

Exemple 1

Considérons un modèle de voiture particulier qui consomme, en moyenne, un litre de carburant tous les 24 kilomètres. Un groupe de recherche a mis au point un nouveau moteur spécialement conçu pour augmenter le nombre de kilomètres effectués avec un litre de carburant. Pour évaluer les performances du nouveau moteur, plusieurs prototypes ont été construits, installés sur des voitures et soumis à des tests de conduite par le groupe de recherche. Il est à noter que le groupe de recherche cherche à prouver que le nouveau moteur augmente en moyenne le nombre de kilomètres effectués avec un litre de carburant.

Les hypothèses formulées sont : $H_0: \mu \leq 24$

$$H_1: \mu > 24$$

Exemple 2

Un producteur de boissons non alcoolisées qui prétend que les bouteilles de deux litres contiennent, en moyenne, au moins 2,028 litres de son produit. Un échantillon de bouteilles de deux litres est sélectionné et leur contenance est mesurée pour tester l'affirmation du fabricant. Dans ce type de test d'hypothèses, on suppose généralement que l'affirmation du fabricant est vraie, à moins que l'échantillon ne prouve le contraire.

Les hypothèses formulées sont : $H_0: \mu \geq 2,028$

$$H_1: \mu < 2,028$$

I.2- Erreurs, seuil de signification et puissance d'un test

Définition :

L'erreur commise en rejetant l'hypothèse nulle H_0 , alors qu'en réalité elle est vraie, s'appelle erreur de première espèce de probabilité α définie par :

$$\alpha = P(\text{rejeter } H_0 \text{ sachant } H_0 \text{ est vraie})$$

Avec α : s'appelle seuil de signification.

L'erreur commise en acceptant l'hypothèse nulle H_0 , alors qu'en réalité elle est fausse, s'appelle erreur de deuxième espèce de probabilité β définie par :

$$\beta = P(\text{accepter } H_0 \text{ sachant } H_0 \text{ est fausse})$$

On appelle puissance du test notée π , la probabilité de rejeter l'hypothèse H_0 lorsque cette hypothèse est fausse, soit :

$$\pi = P(\text{rejeter } H_0 \text{ sachant } H_0 \text{ est fausse}) = 1 - \beta$$

Dans un problème de test, on peut rencontrer les situations suivantes :

| Décision Réalité | H_0 est vraie | H_0 est fausse |
|---------------------|--|---|
| Accepter H_0 | $p(\text{accepter } H_0 \text{ sachant } H_0 \text{ vraie})$ $= 1 - \alpha$ | $p(\text{accepter } H_0 \text{ sachant } H_0 \text{ fausse})$ $= \beta$ |
| Rejeter H_0 | $p(\text{rejeter } H_0 \text{ sachant } H_0 \text{ vraie})$ $= \alpha$ | $p(\text{rejeter } H_0 \text{ sachant } H_0 \text{ fausse})$ $= 1 - \beta$ |

I.3- Variable de décision et région critique

Soient $T(X_1, \dots, X_n)$ une fonction des variables aléatoires de l'échantillon ; $T(X_1, \dots, X_n)$ est appelée variable de décision.

$T_0^c(X_1, \dots, X_n)$: variable de décision calculée.

I.3.1- Définition

On appelle région de rejet de l'hypothèse H_0 ou région critique du test, l'ensemble des échantillons pour lesquels on rejette H_0 en faveur de H_1 , soit :

$$w = \{(X_1, \dots, X_n) : H_0 \text{ est rejetée}\} \text{ telle que } P(w \text{ sachant } H_0 \text{ est vraie}) = \alpha$$

On appelle région d'acceptation de l'hypothèse H_0 , l'ensemble des échantillons pour lesquels on accepte H_0 ,

Soit $\bar{W} = \{(X_1, \dots, X_n) : H_0 \text{ est acceptée}\}$ telle que

$$P(\bar{W} \text{ sachant } H_0 \text{ est vraie}) = 1 - \alpha.$$

I.3.2- Cas de test d'hypothèse bilatéral

$$\text{On a : } W = \left\{ T_0(X_1, \dots, X_n) < q_{\frac{\alpha}{2}} \text{ ou } T_0(X_1, \dots, X_n) > q_{1-\frac{\alpha}{2}} \right\}$$

Dans ce cas la décision du test est :

- Si $T_0^c(X_1, \dots, X_n) < q_{\frac{\alpha}{2}}$ ou $T_0^c(X_1, \dots, X_n) > q_{1-\frac{\alpha}{2}}$ alors on rejette l'hypothèse nulle H_0 en faveur de l'hypothèse alternative H_1 .
- Si $q_{\frac{\alpha}{2}} \leq T_0^c(X_1, \dots, X_n) \leq q_{1-\frac{\alpha}{2}}$ alors on accepte l'hypothèse nulle H_0 .

I.3.3- Cas de test d'hypothèse unilatéral à droite

Le test d'hypothèse est sous cette forme :

$$\begin{cases} H_0 = \theta \leq (=) \theta_0 \\ H_1 = \theta > \theta_0 \end{cases}$$

La région critique $w = \{T_0(X_1, \dots, X_n) > q_{1-\alpha}\}$

Dans ce cas la décision du test est :

- Si $T_0^c(X_1, \dots, X_n) > q_{1-\alpha}$ alors on rejette l'hypothèse nulle H_0 en faveur de l'hypothèse alternative H_1 .
- Si $T_0^c(X_1, \dots, X_n) \leq q_{1-\alpha}$ alors on accepte l'hypothèse nulle H_0 .

I.3.4- Cas de test d'hypothèse unilatéral à gauche

Le test d'hypothèse est sous cette forme :

$$\begin{cases} H_0 = \theta \geq (=) \theta_0 \\ H_1 = \theta < \theta_0 \end{cases}$$

La région critique $W = \{T_0(X_1, \dots, X_n) < q_\alpha\}$.

Dans ce cas la décision du test :

- Si $T_0^c(X_1, \dots, X_n) < q_\alpha$ alors on rejette l'hypothèse nulle H_0 en faveur de l'hypothèse alternative H_1 .
- Si $T_0^c(X_1, \dots, X_n) \geq q_\alpha$ alors on accepte l'hypothèse nulle H_0 .

Remarque : La variable de décision utilisée $T(X_1, \dots, X_n)$ est en général une fonction de θ et de l'estimateur sans biais $\hat{\theta}$. Donc $T_0(X_1, \dots, X_n)$ est une fonction de la valeur θ_0 donnée en H_0 et de $\hat{\theta}$.

II- Test de la moyenne d'une loi normale

On considère un échantillon X_1, \dots, X_n indépendant et identiquement distribuée d'une variable aléatoire X qui suit la loi normale $N(\mu, \sigma^2)$ où la moyenne est μ et la variance est σ^2 . L'objectif est de tester la moyenne (μ) de la population.

II.1- Test de la moyenne avec la variance σ^2 est connue

X suit la loi normale $N(\mu, \sigma^2)$; la variance σ^2 est connue.

On sait que la moyenne empirique \bar{X}_n est un estimateur sans biais de μ , tel que :

$$\bar{X}_n \rightarrow N\left(\mu, \frac{\sigma^2}{n}\right)$$

Dans ce cas **la variable de décision** est la statistique :

$$T(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0, 1)$$

Sous l'hypothèse $H_0: \mu = \mu_0$ ($\mu \leq \mu_0$ ou $\mu \geq \mu_0$) on a :

$$T_0(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu_0}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0, 1)$$

Différents types de test sur la moyenne μ au seuil α peuvent être envisagés et ceci en fonction de l'objectif de l'étude.

Exemple :

Soit l'hypothèse du test suivant :

$$\left\{ \begin{array}{l} H_0 : \mu = 0 \\ H_1 : \mu = -0.2 \end{array} \right\}$$

Soit la variable aléatoire $X \rightarrow N(\mu, 0,1)$; On a $n = 25$; $\bar{X}_n = -0,09$; $\alpha = 5\%$.

- 1- Définir la région critique du test.
- 2- Donner, au seuil de signification $\alpha = 5\%$, la décision du test.

II.2- Test de la moyenne avec la variance σ^2 est inconnue

X suit la loi normale $N(\mu, \sigma^2)$; la variance σ^2 est inconnue.

On sait que la moyenne empirique \bar{X}_n et la variance empirique (corrigée)

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont des estimateurs sans biais de μ et σ^2 respectivement.

Pour le test de la moyenne, la variable de décision est la statistique :

$$T(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} \rightarrow t(n-1)$$

Sous l'hypothèse $H_0: \mu = \mu_0$ ($\mu \leq \mu_0$ ou $\mu \geq \mu_0$) on a :

$$T_0(X_1, \dots, X_n) = \frac{\bar{X}_n - \mu_0}{\frac{S_n}{\sqrt{n}}} \rightarrow t(n-1)$$

La région critique et la décision se différencient d'un cas à un autre.

Exemple :

Considérons le test d'hypothèses suivant :

$$\left\{ \begin{array}{l} H_0 : \mu = 40 \\ H_1 : \mu \neq 40 \end{array} \right\}$$

Un échantillon de taille égale à 25 a fourni une moyenne égale à $\bar{X} = 37,4$ et d'écart type empirique égale à 11,55

- 1- Calculer la valeur de la statistique de test.
- 2- Au seuil de signification de 5%, quelle est votre conclusion ?

III- Test de la moyenne d'une loi inconnue

On considère un échantillon X_1, \dots, X_n indépendant et identiquement distribuée d'une variable aléatoire X dont **la loi est inconnue** avec $E(X) = m, V(X) = \sigma^2$. La taille de l'échantillon $n > 30$. Pour tester la moyenne m de la population, on utilise la loi limite de la moyenne empirique en faisant appel au théorème central limite.

III.1- Test de la moyenne lorsque la variance σ^2 est connue

Le test de la moyenne m est fondé sur **la variable de décision** :

$$T(X_1, \dots, X_n) = \frac{\bar{X}_n - m}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{\text{loi}} N(0, 1)$$

La région critique et la décision se diffèrent d'un test à un autre.

III.2- Test de la moyenne lorsque la variance σ^2 est inconnue

Le test de la moyenne est fondée sur **la variable de décision** :

$$T(X_1, \dots, X_n) = \frac{\bar{X}_n - m}{\frac{S_n}{\sqrt{n}}} \xrightarrow{\text{loi}} N(0, 1)$$

La région critique et la décision dépend du type de test envisagé.

IV- Test de la proportion d'une loi de Bernoulli

Soit X_1, \dots, X_n un échantillon indépendant et identiquement distribuée d'une variable aléatoire X qui suit la loi de Bernoulli de paramètre p . La taille de l'échantillon $n > 30$. Nous voulons tester au seuil α la proportion p .

$\hat{p} = \bar{X}_n$ est un estimateur sans biais de p .

Pour le test de la proportion, la variable de décision est définit ainsi :

$$T(X_1, \dots, X_n) = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow{\text{loi}} N(0; 1)$$

Sous l'hypothèse $H_0: p = p_0$ on a :

$$T_0(X_1, \dots, X_n) = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \xrightarrow{loi} N(0; 1)$$

La région critique dépend du type de test envisagé.

Exemple :

Business Week a publié des informations sur les habitudes de sommeil des Américains. L'article notait que la privation de sommeil cause de nombreux problèmes. Par exemple, le manque de sommeil peut être à l'origine d'accidents de la route mortels. 51% des conducteurs adultes admettent avoir conduit alors qu'ils étaient somnolents. Un chercheur a émis l'hypothèse selon laquelle ce problème serait encore plus sérieux pour les personnes travaillant de nuit.

- Formuler les hypothèses qui permettront de déterminer si plus de 51% des personnes travaillant de nuit admettent avoir conduit alors qu'ils étaient somnolents.
- Sur un échantillon de 500 travailleurs de nuit, 232 ont admis avoir conduit dans un état de somnolence. Quelle est la proportion d'échantillon ? Quelle est la valeur p ?
- Au seuil de signification $\alpha = 0.01$, quelle est votre conclusion ?

V- Test de la variance d'une loi normale

V.1- Test de la variance lorsque la moyenne μ connue

On considère un échantillon X_1, \dots, X_n indépendant et identiquement distribuée d'une variable aléatoire X qui suit la loi normale $N(\mu, \sigma^2)$ où la moyenne est μ est connue et la variance est σ^2 est inconnue. On sait que $\delta_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ est un estimateur sans biais de σ^2 tel que $\frac{n\delta_n^2}{\sigma^2} \rightarrow \chi^2(n)$.

Pour le test de la variance σ^2 , nous pouvons choisir comme variable de décision la statistique suivante :

$$T(X_1; \dots, X_n) = \frac{n\delta_n^2}{\sigma^2} \rightarrow \chi^2(n)$$

Sous l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$, on a :

$$T_0(X_1; \dots, X_n) = \frac{n\delta_n^2}{\sigma_0^2} \rightarrow \chi^2(n)$$

La région critique et la décision se diffèrent d'un test à un autre. Elles font appel aux quantiles de la loi de Khi deux à (n) degrés de liberté.

Exemple :

Dans une société donnée, le salaire est distribuée selon la loi normale de moyenne $\mu = 700$ dinars. Le responsable de cette société affirme que la variance des salaires $\sigma^2 = 120$. Pour vérifier cette assertion, nous avons tiré un échantillon de 30 salaires dans le secteur. Les résultats révèlent que $\sum_{i=1}^{30} (X_i - \mu)^2 = 2013.560$.

Cet échantillon est-il représentatif de la société entière au risque de 5% ?

V.2- Test de la variance lorsque la moyenne μ est inconnue

On considère un échantillon X_1, \dots, X_n indépendant et identiquement distribuée d'une variable aléatoire X qui suit la loi normale $N(\mu, \sigma^2)$ où la moyenne est μ et la variance est σ^2 sont inconnues.

On sait que la moyenne empirique \bar{X}_n et la variance empirique (corrigée)

$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ sont des estimateurs sans biais de μ et σ^2 tel que

$$\frac{(n-1)S_n^2}{\sigma^2} \rightarrow \chi^2(n-1)$$

Pour le test de la variance σ^2 nous pouvons choisir comme variable de décision la statistique :

$$T(X_1; \dots, X_n) = \frac{(n-1)S_n^2}{\sigma^2} \rightarrow \chi^2(n-1)$$

Sous l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$, on a :

$$T_0(X_1; \dots, X_n) = \frac{(n-1)S_n^2}{\sigma_0^2} \rightarrow \chi^2(n-1)$$

Différentes types de tests sur la variance σ^2 , au seuil de signification α , peuvent être envisagés.

La région critique, la décision du test, la puissance du test font appel aux quantiles de la loi de Khi deux à $(n-1)$ degrés de liberté.

VI- Test de la variance d'une loi inconnue

Soit X_1, \dots, X_n un échantillon indépendant et identiquement distribuée d'une variable aléatoire X dont la loi est **inconnue**, avec $E(X) = m$ et $V(X) = \sigma^2$, toutes les deux sont inconnues. La taille n de l'échantillon est assez grande ($n > 30$).

VI.1- Test de la variance lorsque la moyenne m est connue

On sait que $\delta_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - m)^2$ est un estimateur sans biais de σ^2

Pour le test de la variance σ^2 nous pouvons choisir comme variable de décision la statistique :

$$T(X_1, \dots, X_n) = \sqrt{n} \frac{(\delta_n^2 - \sigma^2)}{\sigma^2 \sqrt{2}} \xrightarrow{\text{loi}} N(0; 1)$$

Sous l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$, on a :

$$T_0(X_1; \dots, X_n) = \sqrt{n} \frac{(\delta_n^2 - \sigma_0^2)}{\sigma_0^2 \sqrt{2}} \xrightarrow{\text{loi}} N(0; 1)$$

Différentes types de tests sur la variance σ^2 , au seuil de signification α , peuvent être envisagés.

VI.2- Test de la variance lorsque la moyenne m est inconnue

On sait que la moyenne empirique \bar{X}_n et la variance empirique (corrigée)

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2 \text{ sont des estimateurs sans biais de } m \text{ et } \sigma^2$$

Pour le test de la variance σ^2 nous pouvons choisir comme variable de décision la statistique :

$$T(X_1, \dots, X_n) = \sqrt{n} \frac{(s_n^2 - \sigma^2)}{\sigma^2 \sqrt{2}} \xrightarrow{\text{loi}} N(0; 1)$$

Sous l'hypothèse $H_0 : \sigma^2 = \sigma_0^2$, on a :

$$T_0(X_1; \dots, X_n) = \sqrt{n} \frac{(s_n^2 - \sigma_0^2)}{\sigma_0^2 \sqrt{2}} \xrightarrow{\text{loi}} N(0; 1)$$

Différentes types de tests sur la variance σ^2 , au seuil de signification α , peuvent être envisagés.

REFERENCES

Jamel Fakhfakh, 2008, *Statistique inférentielle : Théorèmes limites – Echantillonnage - Estimation – Tests d’hypothèses – Introduction à l’économétrie*, Tome 3

David R. Anderson, Dennis J. Sweeney, Thomas A. Williams, 2007, *Statistiques pour l’économie et la gestion*, deuxième édition, Bruxelles : De Boeck.

Besma Belhadj, 2005, *Statistique descriptive – calcul des probabilités : Cours – exercices et examens corrigés*, centre de publication universitaire Tunis.