
Sujet 4: Tests de conformité sur une moyenne

• **Quelques rappels sur la définition mathématique des tests d'hypothèses:**

Soit X une variable aléatoire à valeurs dans un ensemble \mathcal{X} , de loi P_θ où $\theta \in \Theta$. Supposons que l'on veuille effectuer un test statistique d'hypothèses portant sur θ . On formule deux hypothèses contradictoires notées H_0 et H_1 dont on suppose que l'une et seulement l'une est vraie. Mathématiquement, formuler H_0 et H_1 revient à choisir deux sous-ensembles disjoints de Θ notés Θ_0 et de Θ_1 de sorte que l'hypothèse H_0 s'écrit alors $\{\theta \in \Theta_0\}$ tandis que l'hypothèse H_1 s'écrit $\{\theta \in \Theta_1\}$. Soit (X_1, \dots, X_n) un échantillon i.i.d. issu d'une variable parente X et soit $(x_1, \dots, x_n) \in \mathcal{X}^n$ une réalisation de (X_1, \dots, X_n) . Construire un test de H_0 contre H_1 revient à construire une région critique \mathcal{R} de telle sorte que l'on rejette H_0 lorsque $(x_1, \dots, x_n) \in \mathcal{R}$ et que l'on s'assure que le risque de 1ère espèce est inférieur ou égal à α .

Dans ce cadre, on parle de **fonction de risque de 1ère espèce** définie sur Θ_0 pour

$$\theta \rightarrow \alpha(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \in \mathcal{R}).$$

On appelle **taille du test** la probabilité maximale de rejeter H_0 alors que H_0 est vraie:

$$\sup_{\theta \in \Theta_0} \{\mathbb{P}_\theta((X_1, \dots, X_n) \in \mathcal{R})\}.$$

On dit qu'un test est de **niveau** α si sa taille est égale à α ie si

$$\sup_{\theta \in \Theta_0} \{\mathbb{P}_\theta((X_1, \dots, X_n) \in \mathcal{R})\} = \alpha.$$

On parle de **fonction de risque de 2nde espèce** définie sur Θ_1 pour

$$\theta \rightarrow \beta(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \notin \mathcal{R}).$$

On parle de **fonction puissance** définie sur Θ_1 pour

$$\theta \rightarrow 1 - \beta(\theta) = \mathbb{P}_\theta((X_1, \dots, X_n) \in \mathcal{R}).$$

• On souhaite ici comparer une moyenne à une norme fixée à partir d'un échantillon i.i.d. (X_1, \dots, X_n) de variable parente X . Formellement, on souhaite tester l'hypothèse nulle H_0 :

$\mathbb{E}[X] = m_0$ contre une hypothèse alternative H_1 au niveau α . Soit $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ un estima-

teur de $\mathbb{E}[X]$. Notons $S_n'^2 = \frac{1}{n-1} \sum_{i=1}^n (\bar{X}_n - X_i)^2$ un estimateur de $\text{Var}(X)$.

• **1er cas:** $n \geq 30$ (pas de condition sur la distribution autre que l'existence de moments d'ordre 1 et 2).

Soit la statistique de test:

$$T_n = \sqrt{n} \frac{\bar{X}_n - m_0}{\sqrt{S_n'^2}}.$$

- Lorsque $H_1: \mathbb{E}[X] \neq m_0$, la région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : |t_n| > F_{\mathcal{N}(0,1)}^{-1}(1 - \alpha/2)\}.$$

- Lorsque $H_1: \mathbb{E}[X] > m_0$, la région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : t_n > F_{\mathcal{N}(0,1)}^{-1}(1 - \alpha)\}.$$

- Lorsque $H_1: \mathbb{E}[X] < m_0$, la région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : t_n < F_{\mathcal{N}(0,1)}^{-1}(\alpha)\}.$$

• **2ème cas:** les X_i suivent la loi gaussienne (pas de condition sur la taille de l'échantillon).
Soit la statistique de test:

$$T_n = \sqrt{n} \frac{\bar{X}_n - m_0}{\sqrt{S_n'^2}}.$$

- Lorsque $H_1: \mathbb{E}[X] \neq m_0$, la région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : |t_n| > F_{T(n-1)}^{-1}(1 - \alpha/2)\}.$$

- Lorsque $H_1: \mathbb{E}[X] > m_0$, la région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : t_n > F_{T(n-1)}^{-1}(1 - \alpha)\}.$$

- Lorsque $H_1: \mathbb{E}[X] < m_0$, la région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in \{0, 1\}^n : t_n < F_{T(n-1)}^{-1}(\alpha)\}.$$

NB: ce test s'appelle le test de Student.

Exercice 1.

1. Illustrer de manière empirique à partir de données simulées le comportement de la taille du test.
2. Illustrer de manière empirique à partir de données simulées le comportement de la fonction puissance.

Vous veillerez à faire varier tour à tour:

- le risque de 1ère espèce α ,
- la taille de l'échantillon n ,
- la variabilité du phénomène.

Le test de Student est réputé robuste aux écarts à la normalité. Qu'en pensez-vous?
Dans le cas d'un test unilatéral, que se passe-t-il si l'on se trompe dans le choix de H_0 et de H_1 , au sens où aucune des deux hypothèses n'est vraie en réalité?