Université de Strasbourg
M1 - Magistère
Statistique - études de cas

Ségolen Geffray
geffray@math.unistra.fr
Année 2015/2016

Sujet 5: Tests de conformité sur une variance

• **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 $\{\theta_0 \in \Theta_0\}$ tandis que l'hypothèse H_1 s'écrit $\{\theta_1 \in \Theta_1\}$. Soit $(X_1, ..., X_n)$ un échantillon i.i.d. issu d'une variable parente X et soit $(x_1, ..., x_n) \in \mathcal{X}^n$ une réalisation de $(X_1, ..., X_n)$. 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 $(x_1, ..., x_n) \in \mathbb{R}$ 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

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

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

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

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

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

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

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

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

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

• Soit une variable aléatoire X dont on souhaite comparer la variance à une norme fixée. Formellement, on souhaite tester l'hypothèse nulle H_0: $\mathrm{Var}(X) = \sigma_0^2$ contre une hypothèse alternative H_1 au niveau α. Soit un échantillon i.i.d. $(X_1, \ldots, X_n)$ distribué comme X. Soit

$$S_n'^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\overline{X}_n - X_i\right)^2$$

un estimateur de $\mathrm{Var}(X)$ en notant $\overline{X}_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

• **1er cas**: Lorsque $n \geq 30$, on considère la statistique de test:

$$T_n = \sqrt{n}\frac{S_n'^2 - \sigma_0^2}{\sqrt{2\sigma_0^4}}.$$

Sous H_0, la statistique T_n suit asymptotiquement la loi $\mathcal{N}(0,1)$. Sous H_1, la statistique T_n diverge presque-sûrement vers ∞.

- Lorsque H_1: $\mathrm{Var}(X) \neq \sigma_0^2$, la région critique est
$$\mathcal{R} = \{|T_n| > F^{-1}_{\mathcal{N}(0,1)}(1-\alpha/2)\}.$$
- Lorsque H_1: $\mathrm{Var}(X) < \sigma_0^2$, la région critique est
$$\mathcal{R} = \{T_n < F^{-1}_{\mathcal{N}(0,1)}(\alpha)\}.$$
- Lorsque H_1: $\mathrm{Var}(X) > \sigma_0^2$, la région critique est
$$\mathcal{R} = \{T_n > F^{-1}_{\mathcal{N}(0,1)}(1-\alpha)\}.$$

• **2ème cas**: Lorsque l'échantillon est gaussien, on considère la statistique de test:

$$T_n = \frac{(n-1)S_n'^2}{\sigma_0^2}.$$

Sous H_0, la statistique T_n suit la loi $\chi^2(n-1)$.

- Lorsque H_1: $\mathrm{Var}(X) \neq \sigma_0^2$, la région critique est
$$\mathcal{R} = \{t_n < F^{-1}_{\chi^2(n-1)}(\alpha/2) \text{ ou } t_n > F^{-1}_{\chi^2(n-1)}(1-\alpha/2)\}.$$
- Lorsque H_1: $\mathrm{Var}(X) < \sigma_0^2$, la région critique est
$$\mathcal{R} = \{T_n < F^{-1}_{\chi^2(n-1)}(\alpha)\}.$$
- Lorsque H_1: $\mathrm{Var}(X) > \sigma_0^2$, la région critique est
$$\mathcal{R} = \{T_n > F^{-1}_{\chi^2(n-1)}(1-\alpha)\}.$$

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 étudié en 2ème cas 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 vrai en réalité?