Université de Strasbourg — Ségolen Geffray
M1 - Magistère — geffray@math.unistra.fr
Statistique - études de cas — Année 2015/2016

Sujet 3: Tests de conformité sur le paramètre d'une loi exponentielle

---

• **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

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

• On souhaite ici comparer le paramètre d'une loi exponentielle à une norme fixée à partir d'un échantillon i.i.d. Soit $(X_1, \ldots, X_n)$ un échantillon i.i.d. de loi $\mathcal{E}(\lambda)$ où $\lambda > 0$. Formellement, on souhaite tester l'hypothèse nulle H_0: $\lambda = \lambda_0$ contre l'hypothèse alternative H_1: $\lambda \neq \lambda_0$ au niveau α.

• **Test asymptotique**:
Soit $\widehat{\lambda}_n = \dfrac{1}{\frac{1}{n}\sum_{i=1}^n X_i}$ l'EMV de λ. Soit la statistique de test:

$$T_n = \sqrt{n}\,\frac{\widehat{\lambda}_n - \lambda_0}{\lambda_0}\,.$$

Notons t_n sa réalisation. La région critique est

$$\mathcal{R} = \{(x_1, \dots, x_n) \in (\mathbb{R}^{*+})^n : |t_n| > F^{-1}_{\mathcal{N}(0,1)}(1 - \alpha/2)\}$$

en notant $F^{-1}_{\mathcal{N}(0,1)}(1 - \alpha/2)$ le fractile d'ordre $(1 - \alpha/2)$ de la loi gaussienne standard $\mathcal{N}(0, 1)$.

• **Test exact**:
On peut montrer que, sous H_0: $\lambda = \lambda_0$, on a $\sum_{i=1}^{n} X_i \sim \Gamma(n, \lambda_0)$. Par ailleurs, rappelons que si $W \sim \Gamma(a, b)$, alors, pour tout $c > 0$, on a $cW \sim \Gamma\left(a, \frac{b}{c}\right)$. On en déduit ici que, sous H_0: $\lambda = \lambda_0$, on a $2\lambda_0 \sum_{i=1}^{n} X_i \sim \Gamma(n, 1/2) = \chi^2(2n)$. Utilisons

$$T_n = 2\lambda_0 \sum_{i=1}^{n} X_i$$

comme statistique de test. Notons t_n sa réalisation. La région critique recherchée est

$$\mathcal{R} = \left\{(x_1, \dots, x_n) \in (\mathbb{R}^{*+})^n : \ t_n < F^{-1}_{\chi^2(2n)}(\alpha/2) \text{ ou } t_n > F^{-1}_{\chi^2(2n)}(1 - \alpha/2)\right\}$$

en notant $F^{-1}_{\chi^2(2n)}(\alpha/2)$ et $F^{-1}_{\chi^2(2n)}(1 - \alpha/2)$ les fractiles d'ordre $\alpha/2$ et $(1 - \alpha/2)$ de la loi du χ^2 à $(2n)$ degrés de liberté.

• **Test fondé sur une borne exponentielle**:
Le théorème des grandes déviations permet d'obtenir les deux inégalités suivantes:

$$\forall \delta > \frac{1}{\lambda_0}, \quad \mathbb{P}_{\lambda_0}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \geq \delta\right) \leq \exp\left(-n\left(\delta\lambda_0 - 1 - \log\left(\delta\lambda_0\right)\right)\right)$$

et

$$\forall 0 < \delta < \frac{1}{\lambda_0}, \quad \mathbb{P}_{\lambda_0}\left(\frac{1}{n}\sum_{i=1}^{n} X_i \leq \delta\right) \leq \exp\left(-n\left(\delta\lambda_0 - 1 - \log\left(\delta\lambda_0\right)\right)\right).$$

Notons $k_{\alpha/2,n}$ et $K_{1-\alpha/2,n}$ les deux solutions, respectivement inférieure et supérieure à $1/\lambda_0$, de l'équation en δ:

$$\delta\lambda_0 - 1 - \log\left(\delta\lambda_0\right) = \frac{1}{n}\log\left(\frac{2}{\alpha}\right).$$

La région critique est alors

$$\mathcal{R} = \left\{(x_1, \dots, x_n) \in (\mathbb{R}^{*+})^n : \frac{1}{n}\sum_{i=1}^{n} x_i < k_{\alpha/2,n} \text{ ou } \frac{1}{n}\sum_{i=1}^{n} x_i > K_{1-\alpha/2,n}\right\}.$$

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.

3. Illustrer de manière empirique à partir de données simulées la robustesse ou au contraire la sensibilité du test aux hypothèses sous-jacentes.
   NB: la loi de Weibull est une loi continue à valeurs dans $\mathbb{R}^{*+}$ qui inclut la loi exponentielle comme cas particulier.

Vous veillerez à faire varier tour à tour:

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