

---

**Sujet 1: Tests de conformité sur le paramètre d'une loi de Poisson**

---

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

Soit  $X$  une variable aléatoire à valeurs dans un ensemble  $\mathcal{X}$ , de loi  $P_\theta$  où  $\theta \in \Theta$ . Supposons que l'on veuille effectuer un test statistique d'hypothèses portant sur  $\theta$ . 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  $\Theta$  notés  $\Theta_0$  et de  $\Theta_1$  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  $(X_1, \dots, X_n)$  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  $\alpha$ .

Dans ce cadre, on parle de **fonction de risque de 1ère espèce** définie sur  $\Theta_0$  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**  $\alpha$  si sa taille est égale à  $\alpha$  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  $\Theta_1$  pour

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

On parle de **fonction puissance** définie sur  $\Theta_1$  pour

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

• On souhaite ici comparer le paramètre  $\lambda$  d'une loi de Poisson à une norme fixée à partir d'un échantillon i.i.d. Soit  $(X_1, \dots, X_n)$  un échantillon i.i.d. de loi  $\mathcal{P}(\lambda)$  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  $\alpha$  où la valeur  $\lambda_0$  est fixée par l'utilisateur.

• **Test asymptotique:**

Soit  $\hat{\lambda}_n = \frac{1}{n} \sum_{i=1}^n X_i$  l'EMV de  $\lambda$ . Soit la statistique de test:

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

Notons  $t_n$  sa réalisation. La région critique est

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

en notant  $F_{\mathcal{N}(0,1)}^{-1}(1 - \alpha/2)$  le fractile d'ordre  $(1 - \alpha/2)$  de la loi gaussienne standard  $\mathcal{N}(0, 1)$ . En pratique, il est recommandé que  $n$  soit assez grand pour que  $n\lambda_0 > 10$ .

• **Test exact:**

On sait que, sous  $H_0: \lambda = \lambda_0$ , on a  $\sum_{i=1}^n X_i \sim \mathcal{P}(n\lambda_0)$ . On cherche le plus grand  $k_{\alpha/2} \in \mathbb{N}$  tel que

$$\mathbb{P}_{\lambda_0} \left( \sum_{i=1}^n X_i < k_{\alpha/2} \right) \leq \alpha/2$$

et le plus petit  $K_{1-\alpha/2} \in \mathbb{N}$  tel que

$$\mathbb{P}_{\lambda_0} \left( \sum_{i=1}^n X_i > K_{1-\alpha/2} \right) \geq \alpha/2.$$

La région critique est alors

$$\mathcal{R} = \left\{ (x_1, \dots, x_n) \in \mathbb{N}^n : \sum_{i=1}^n x_i \in \left( \bigcup_{k=0}^{k_{\alpha/2}-1} \{k\} \right) \cup \left( \bigcup_{k=K_{1-\alpha/2}+1}^{\infty} \{k\} \right) \right\}$$

avec la convention  $\bigcup_{k=0}^{k_{\alpha/2}-1} \{k\} = \emptyset$  si  $k_{\alpha/2} = 0$ .

La fonction `poisson.test` du logiciel R implémente ce test.

• **Test fondé sur une borne exponentielle:**

Le théorème des grandes déviations permet d'obtenir les deux inégalités suivantes:

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

et

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

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

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

La région critique est alors

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

### 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 binomiale négative est une loi discrète à valeurs dans  $\mathbb{N}$  qui inclut la loi de Poisson comme cas particulier.

Vous veillerez à faire varier tour à tour:

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