VI. Tests

VI.1. Neyman-Pearson : loi exponentielle. Soit $n \geq 1$ un entier. On observe

$$X_1, \ldots, X_n$$

où les variables aléatoires X_i sont indépendantes, de même loi exponentielle de paramètre $\lambda > 0$, c'est-à-dire de densité

$$x \rightsquigarrow \lambda \exp\big(-\lambda x\big) 1_{\{x \geq 0\}}.$$

(1) Ecrire le modèle statistique engendré par l'observation de $(X_1, \ldots, X_n)$.
(2) Calculer l'estimateur du maximum de vraisemblance $\widehat{\lambda}_n^{\mathtt{mv}}$ de λ.
(3) Montrer que $\widehat{\lambda}_n^{\mathtt{mv}}$ est asymptotiquement normal et calculer sa variance limite.
(4) Soient $0 < \lambda_0 < \lambda_1$. Construire un test d'hypothèse de

$$H_0 \,:\, \lambda = \lambda_0 \quad \text{contre} \quad H_1 \,:\, \lambda = \lambda_1$$

de niveau α et uniformément plus puissant. Expliciter le choix du seuil définissant la région critique. Montrer que l'erreur de seconde espèce de ce test tend vers 0 lorsque $n \to \infty$.

VI.2. Test de Fisher. Soit E un sous-espace vectoriel de $\mathbb{R}^n$ de dimension $p < n$. On observe un vecteur gaussien X de dimension n et de loi $\mathcal{N}(\mu, \sigma^2 I_n)$ où $\mu \in E$ et $\sigma^2 > 0$ sont inconnus. On s'intéresse aux hypothèses :

$$H_0 : \mu \in F \quad \text{contre} \quad H_1 : \mu \notin F$$

où F est un sous espace vectoriel de E de dimension $q < p$. On note X_V la projection orthogonale de X sur un sous-espace vectoriel V.

(1) Montrer que $\|X - X_E\|^2$ et $\|X_E - X_F\|^2$ sont indépendants et décrire leur loi.
(2) En déduire que si

$$T = \frac{\|X_E - X_F\|^2/(p-q)}{\hat{\sigma}^2} \quad \text{où } \hat{\sigma}^2 = \frac{1}{n-p}\|X - X_E\|^2$$

alors T suit une loi de Fisher sous H_0.
(3) Si $c > 0$ est une constante montrer que la puissance du test $1_{\{T>c\}}$ au point (μ, σ^2) est une fonction croissante de $\|\mu - \mu_F\|^2/\sigma^2$.

VI.3. ANOVA. On observe k échantillons gaussiens de tailles respectives $n_1, n_2, \ldots, n_k$:

$$X_{i\cdot} = (X_{i,j})_{1 \leq j \leq n_i}, \quad 1 \leq i \leq k$$

tels que

$$X_{i,j} \sim \mathcal{N}(m_i, \sigma^2)$$

où $m = (m_1, \ldots, m_k) \in \mathbb{R}^k$ et $\sigma^2 > 0$ sont inconnus. on s'intéresse aux hypothèses :

$$H_0 : m_1 = m_2 = \cdots = m_k \quad \text{contre} \quad H_1 : \ \exists i \neq i' \text{ tels que } m_i \neq m_{i'}$$

(1) On considère le vecteur aléatoire $X = (X_{ij})_{1 \leq i \leq k, 1 \leq j \leq n_i}$ de dimension $n = n_1 + \cdots + n_k$. Montrer que $\mu = E(X)$ appartient à un sous-espace vectoriel E de dimension k. Calculer X_E.

(2) Montrer que l'hypothèse nulle s'écrit $H_0 : \mu \in F$ où F est un sous-espace vectoriel de E de dimension 1. Calculer X_F.

(3) En déduire la forme du test de Fisher dans ce contexte.

VI.4. Echantillonnage : Un test asymptotique de gaussianité. Soient $X_1, \dots, X_n$ un n-échantillon de loi inconnue F ayant au moins un moment d'ordre 4 et de moyenne nulle et de variance non-nulle.

(1) On pose, pour $k = 1, \dots, 4$

$$T_n^{(k)} = \frac{\frac{1}{n}\sum_{i=1}^n X_i^k}{\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{k/2}}.$$

Montrer que si F est une distribution gaussienne on a

$$\frac{n}{15}\left(T_n^{(3)}\right)^2 + \frac{n}{24}\left(T_n^{(4)} - 3\right)^2 \xrightarrow{d} \chi_2^2,$$

où χ_2^2 désigne la loi du χ^2 à 2 degrés de liberté.

(2) En déduire un test de l'hypothèse nulle H_0 : "F est gaussienne" contre l'alternative H_1 : "F n'est pas gaussienne".

(3) Le test est-il consistant ?

VI.5. Neyman-Pearson. On considère le modèle statistique $(\mathbb{R}^n, (P_\theta^n)_{\theta\in\mathbb{R}})$ associé à l'observation d'un n-échantillon gaussien de la loi $\mathcal{N}(\theta, 1)$ où $\theta \in \mathbb{R}$ est inconnu. On s'intéresse au test de l'hypothèse nulle $H_0 : \theta = 0$ contre $H_1 : \theta \neq 0$.

(1) Soit $\alpha \in]0, 1[$. Construire un test φ de niveau α dont la fonction puissance est paire.

(2) Montrer qu'il n'existe pas de test φ^* de niveau α qui soit uniformément plus puissant c'est-à-dire tel que pour tout autre test ψ on ait

$$P_{\theta_0}^n(\psi = 1) \leq \alpha \Longrightarrow \Big(\forall \theta \in H_1 \quad P_\theta^n(\psi = 1) \leq P_\theta^n(\varphi^* = 1)\Big)$$

(3) Pour $\theta \in \mathbb{R}$ on définit la probabilité Q_θ^n sur $\mathbb{R}^n$ par

$$Q_\theta^n = \frac{1}{2}\left(P_\theta^n + P_{-\theta}^n\right).$$

Calculer le processus de vraisemblance $(L_n(\theta))_{\theta\in\mathbb{R}}$ du modèle $(Q_\theta^n)_{\theta\in\mathbb{R}}$.

(4) Montrer que

$$\frac{L_n(\theta)}{L_n(0)} = \exp(-n\theta^2/2)\cosh(n\theta\bar{X}_n) \quad \text{avec} \quad \bar{X}_n = n^{-1}\sum_{i=1}^n X_i.$$

En déduire que dans le modèle $(\mathbb{R}^n, (Q_\theta^n)_{\theta\in\mathbb{R}})$ il existe un test φ^* de niveau de α uniformément plus puissant.

(5) Montrer que dans le modèle $(\mathbb{R}^n, (P_\theta^n)_{\theta\in\mathbb{R}})$, φ^* admet une fonction puissance paire et que φ^* est uniformément plus puissant parmi les tests dont la fonction puissance est paire.

VI.6. Exercice. On lance 60 fois un dé et on obtient les résultats suivants :

1	2	3	4	5	6
10	13	8	12	9	8

Au seuil de $0,025$ peut-on conclure que le dé est bien équilibré?

VI.7. Exercice. [Test du signe] On observe n couples aléatoires

$$(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$$

indépendants mais pas nécessairement de même loi. On suppose de plus que X_i et Y_i ont une loi diffuse pour tout $i \in \{1, \dots, n\}$. On considère le test d'hypothèse

$$H_0 : X_i = Y_i \text{ en loi pour tout } i$$

contre

$$H_1 : \text{il existe } i \neq j \text{ tels que } X_i \neq Y_i \text{ en loi}$$

(1) Montrer que $P\big[X_i = Y_i\big] = 0$ et en déduire que sous H_0 on a $P\big[X_i > Y_i\big] = \frac{1}{2}$.

(2) On pose

$$N = \sum_{i=1}^{n} 1_{\left\{X_i > Y_i\right\}}.$$

Quelle est la loi de N sous H_0 ?

(3) En déduire que le test défini par la région critique

$$\left\{\left|N - \frac{n}{2}\right| \geq c\right\}$$

permet de construire un test de niveau inférieur à $\alpha \in (0, 1)$ de H_0 contre H_1 pour un choix $c = c(\alpha) > 0$ que l'on précisera. Parmi tous les choix possibles de $c(\alpha)$, lequel préférer ?

(4) Les moyennes générales de la première et de la deuxième année de cinquième de 12 redoublants ont été relevées:

Elève	1	2	3	4	5	6	7	8	9	10	11	12
Année 1	8,2	6,9	7,0	6,2	6,4	6,3	7,2	7,6	7,8	6,4	7,3	8
Année 2	10,1	6,7	7,9	10,5	5,3	8,3	9,6	11,4	7,9	9,9	10,0	9,8

Le redoublement a-t-il une influence sur la moyenne générale ?