Exercices, 6

Exercice 1. — 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 $x \mapsto \lambda \exp(-\lambda x) \mathbb{1}_{x \geq 0}$.

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'hypothèses 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 $n \to \infty$.

Exercice 2. — 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? A titre indicatif le quantile d'une loi $\chi^2(5)$ d'ordre 95% est 11.07.

Exercice 3. — 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

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

alors T suit une loi de Fisher sous H_0.

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

Exercice 4 (ANOVA). — On observe k échantillons gaussiens de tailles respectives $n_1, n_2, \ldots, n_k$, notés $X_{i\cdot} = (X_{i,j})_{1 \leq j \leq n_i}$, $\quad 1 \leq i \leq k$, 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 $\mu = \mathbb{E}[X]$ 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.

EXERCICE 5 (Un test asymptotique de gaussianité, le test de Jarque-Bera). — Soit $(X_1, \ldots, X_n)$ 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 $k = 1, \ldots, 4$,
$$T_n^{(k)} = \frac{\frac{1}{n}\sum_{i=1}^n X_i^k}{\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right)^{k/2}}.$$
Montrer que si F est une distribution gaussienne, on a
$$\frac{n}{15}\left(T_n^{(3)}\right)^2 + \frac{n}{96}\left(T_n^{(4)} - 3\right)^2 \xrightarrow{\mathscr{L}} \chi_2^2,$$
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 ?

EXERCICE 6 (Test du signe). — On observe n couples aléatoires $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ indépendants mais pas nécessairement de même loi. On suppose de plus que les variables X_i et Y_i sont indépendantes et qu'elles ont une loi diffuse pour tout $i \in \{1, \ldots, n\}$. On considère le test d'hypothèses
$$\begin{array}{ll} H_0 : & X_i = Y_i \text{ en loi pour tout } i, \\ H_1 : & \text{il existe } i \neq j \text{ tels que } X_i \neq Y_i \text{ en loi.} \end{array}$$

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

2. On pose $N = \sum_{i=1}^n \mathbb{1}_{X_i > Y_i}$. 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, sachant que le quantile d'ordre 0.975 d'une $\mathscr{B}(12, 0.5)$ est 9.

Exercice 7 (***-Modèle de Cauchy**). —

On suppose que l'on observe $X_1, \ldots, X_n$ i.i.d. de loi $m + \mathcal{C}(1)$, où m est un paramètre réel à estimer, et $\mathcal{C}(1)$ désigne une loi de Cauchy de paramètre 1.

1. On note $\hat{m}_{naive}$ l'estimateur $\frac{1}{n}\sum_{1=1}^{n} X_i$. Montrer que $\hat{m}_{naive}$ n'est pas consistant.
2. On note $\hat{m}_{EMV}$ un estimateur par maximum de vraisemblance. Exprimer $\hat{m}_{EMV}$ comme racine d'un polynôme.

Pour des raisons de commodité calculatoire, on suppose qu'on observe un nombre impair d'observations, c'est à dire $X_1, \ldots, X_{2n+1}$. On note $X_{(1)} < \ldots < X_{(2n+1)}$ l'échantillon trié (bien défini p.s.).

3. Que vaut la médiane empirique? On la notera $\hat{m}$.
4. Montrer que $\hat{m}$ est consistant.
5. Construire un intervalle de confiance de niveau $1-\alpha$ pour m basé sur $\hat{m}$.
6. Montrer que $\hat{m}$ est fortement consistant.

On admet le résultat suivant (se déduit du Théorème de Lindeberg Feller).
Soit $(X_{2n+1})_{n\geq 1}$ une suite de variables aléatoires de lois binômiales $\mathcal{B}(2n+1, p_{2n+1})$ satisfaisant $(2n+1)p_{2n+1}(1-p_{2n+1}) \to +\infty$ quand $n \to +\infty$. Alors

$$\frac{X_{2n+1} - (2n+1)p_{2n+1}}{\sqrt{(2n+1)p_{2n+1}(1-p_{2n+1})}} \rightsquigarrow \mathcal{N}(0,1).$$

7. Montrer que $\frac{2}{\pi}\sqrt{2n+1}(\hat{m}-m)$ converge en loi vers une loi $\mathcal{N}(0,1)$, et en déduire un intervalle de confiance asymptotique de niveau $1-\alpha$ sur m, où $\alpha \in]0,1[$.