

Exercices, 2

EXERCICE 1. — Soient X_1, \dots, X_n i.i.d de loi $\mathcal{N}(\theta, 1)$, où θ est un paramètre réel.

1. Donner un intervalle de confiance pour θ au niveau de risque 5% de la forme $[\hat{\theta}_n^-, +\infty[$.
2. En déduire un test de niveau 5% pour les hypothèses $H_0 : \theta = 0$ et $H_1 : \theta > 0$.
3. Donner le modèle de l'expérience statistique, avec deux possibilités de mesures dominantes. Dans ces deux cas, donner l'expression du test de rapport de vraisemblance $T_{0,\mu}$ pour les hypothèses $H_0 : \theta = 0$ et $H_1 : \theta = \mu$, où $\mu > 0$. Quel test retrouve-t-on?
4. Soit φ un autre test de niveau α pour $H_0 : \theta = 0$ et $H_1 : \theta = \mu$. Montrer que $T_{0,\mu}$ est plus puissant que φ . (Indication: On pourra regarder $\int_{\mathbb{R}} (\varphi - T_{0,\mu})(f_\mu - t f_0)(x) dx$, pour un t bien choisi).
5. Construire le test de rapport de vraisemblance au niveau 5% pour les hypothèses $H_0 : \theta = 0$ et $H_1 : \theta > 0$.

EXERCICE 2 (Test de la médiane). — On considère un échantillon X_1, \dots, X_n de variables iid réelles de loi diffuse, et $m \in \mathbb{R}$. On souhaite tester l'hypothèse H_0 : "la médiane de la loi des X_i est égale à m ", contre l'hypothèse alternative H_1 : "la médiane de la loi des X_i est strictement inférieure à m ". On considère la statistique de test

$$S = \sum_{i=1}^n \mathbb{I}_{\{X_i > m\}}.$$

1. Montrer que sous H_0 , la loi de la statistique S est libre de celle de X .
2. Le 11 janvier 2014, la vitesse maximale sur le périphérique parisien est passée de 80 km/h à 70 km/h. Un maître de conférences s'interroge de l'impact de cette mesure sur son temps de trajet total en voiture entre une ville de la banlieue Sud et l'Université Paris-Dauphine. Une série de données a permis de montrer qu'avec un périphérique limité à 80 km/h, le temps de trajet médian est de 38 min. Suite au passage à une vitesse limitée à 70 km/h, il recueille un échantillon de $n = 23$ temps de trajet en minutes¹ :

26 29 31 26 32 29 37 37 29 37 30 47 41 37 45 35 35 36 32 30 43 30 35.

Ayant le sentiment que son temps de trajet a diminué, il souhaite tester l'hypothèse nulle que ces observations sont des réalisations d'une loi dont la médiane est toujours de 38 minutes, contre l'hypothèse alternative que la médiane est moindre. Calculer la p -valeur du test à l'aide de la table ci-dessous.

k	3	4	5	6	7	8	9	10	11	12
$P[S \leq k]$	0.0002	0.0013	0.0053	0.0173	0.0466	0.105	0.2024	0.3388	0.5	0.6612
k	13	14	15	16	17	18	19	20	21	22
$P[S \leq k]$	0.7976	0.895	0.9534	0.9827	0.9947	0.9987	0.9998	1	1	1

EXERCICE 3. — On se donne X_1, \dots, X_n i.i.d de loi $\mathcal{U}(]0, \theta[)$, et on note $M_n = \max_{j=1, \dots, n} X_j$.

1. Si $\theta = 1$ et $t < 1$, que vaut $P_1(M_n \leq 1 - t)$? En déduire un test $T_{1,1-}$ de niveau α pour les hypothèses $H_0 : \theta = 1$ contre $H_1 : \theta < 1$.

¹Source des données: Waze.

- Donner le test du rapport de vraisemblance pour les hypothèses $H_0 : \theta = 1$ contre $H_1 : \theta < 1$, au niveau α .
- Soit $\theta < 1$. Que vaut $P_\theta(\{T_{1,1-} = 1\})$? En déduire que la puissance uniforme de $\{T_{1,1-}$ sur $\{\theta < 1\}$ vaut α .
- Donner le test du rapport de vraisemblance pour les hypothèses $H_0 : \theta = 1$ contre $H_1 : \theta > 1$, noté $T_{1,1+}$, au niveau α . Quelle est sa puissance minimale sur $\{\theta > 1\}$?
- Donner un test de niveau α pour $H_0 : \theta = 1$ contre $H_1 : \theta > 1$, plus puissant que $T_{1,1+}$ (pour n'importe quel $\theta > 1$ et en puissance minimale)?

EXERCICE 4. — On note p la probabilité qu'un enfant né vivant soit un garçon. On suppose que les enfants sont de sexe indépendants, et que cette probabilité est la même pour toutes les grossesses.

- Il y a eu en France métropolitaine en 2015 $n = 760\,421$ naissances², dont 389 181 garçons. Tester l'hypothèse $p = \frac{1}{2}$ contre l'alternative pertinente à l'aide d'un test de rapport de vraisemblance.
- En 1920, il y a eu 838 137 naissances dont 432 044 garçons. Tester l'hypothèse $p_{2015} = p_{1920}$.

EXERCICE 5 (Test composite). — On considère deux hypothèses composites sur un paramètre θ : $H_0 : \theta \in \Theta_0$ contre $H_1 : \theta \in \Theta_1$, où l'espace des valeurs possibles pour θ est l'union disjointe de Θ_0 et Θ_1 . On veut tester H_1 contre H_0 au vu de données x ; on se donne une fonction de test φ et des régions d'acceptation et de rejet R et A :

$$\varphi(x) = \begin{cases} 1 & \text{si } x \in R \\ 0 & \text{si } x \in A. \end{cases}$$

On dit que la fonction de test φ est *uniformément la plus puissante* (UPP) de niveau α si pour tout autre fonction de test φ' telle que

$$\sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\varphi'(X)] = \alpha' \leq \alpha = \sup_{\theta \in \Theta_0} \mathbb{E}_\theta[\varphi(X)]$$

on a

$$\forall \theta \in \Theta_1, \mathbb{E}_\theta[\varphi'(X)] = 1 - \beta' \leq 1 - \beta = \mathbb{E}_\theta[\varphi(X)].$$

Au vu d'une fonction de vraisemblance L , on appelle rapport de vraisemblance la quantité

$$\Lambda(x) = \frac{\sup_{\theta \in \Theta_1} L(\theta; x)}{\sup_{\theta \in \Theta_0} L(\theta; x)}$$

et on appelle test de rapport de vraisemblance un test dont la fonction est de la forme

$$\varphi(x) = \mathbb{I}_{\{\Lambda(x) \geq c\}}$$

pour une constante c .

- On considère un échantillon $X = (X_1, \dots, X_n)$ d'observations iid de la loi $\text{Ber}(\theta)$, $\Theta_0 = \{\theta_0\}$, $\Theta_1 = \{\theta_1\}$ avec $\theta_1 > \theta_0$. Montrer que le test de rapport de vraisemblance est de la forme $\varphi(X) = \mathbb{I}_{\{\sum_i X_i \geq K\}}$ où K dépend de θ_0 et α mais pas de θ_1 .
- On considère $\Theta_0 = \{\theta_0\}$ et $\Theta_1 =]\theta_0, 1]$ (on suppose que les valeurs $\theta < \theta_0$ sont physiquement impossibles). Montrer que le test de rapport de vraisemblance est UPP.
- On considère $\Theta_0 = [0, \theta_0]$ et $\Theta_1 =]\theta_0, 1]$. Montrer que le test de rapport de vraisemblance est UPP.
- On considère $\Theta_0 = [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ avec $0 \leq \varepsilon < \max(\theta_0, 1 - \theta_0)$ et $\Theta_1 = [0, 1] \setminus \Theta_0$. Montrer qu'il n'existe pas de test UPP.

²Source : Insee, statistiques de l'état civil.