

Feuille de travaux dirigés 5 : Test de Neyman-Pearson, maximum de vraisemblance et méthode des moments

Exercice 1 (Test de Neyman-Pearson : gaussiennes à moyenne connue):

1. Soit Y un vecteur gaussien centré de taille n . On veut tester l'hypothèse $H_0 : Y \sim \mathcal{N}(0, \Sigma_0)$ versus $H_1 : Y \sim \mathcal{N}(0, \Sigma_1)$ où Σ_0, Σ_1 sont inversibles. Montrer que le test de Neyman-Pearson revient à comparer $y^T(\Sigma_1^{-1} - \Sigma_0^{-1})y$ à un seuil.
2. Soient X et V deux variables gaussiennes réelles, centrées, de variances respectives σ_X^2 et σ_V^2 . La variable X est un signal utile et V est un bruit de mesure. L'observation est donnée par $Y = X + V$. On recoit n observations indépendantes. Proposer un test au niveau α permettant de détecter la présence du signal X .
3. Pour le test précédent, donner la valeur du seuil en fonction des quantiles de la loi du chi-deux. On précise que la loi du chi-deux à n degrés de libertés est la loi suivie par la somme de n variables normales centrées réduites indépendantes :

$$X \sim \chi_n^2 \iff X \stackrel{\text{loi}}{=} \sum_{i=1}^n U_i^2 \text{ où } U_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$$

Exercice 2 (Test pour une certification bio):

Pour avoir la certification “bio”, un fabricant de produits “bios” doit garantir pour chaque lot un pourcentage d’OGM inférieur à 1%. Il prélève donc $n = 25$ produits par lot et teste si le pourcentage d’OGM est inférieur à 1%. On note X_i le logarithme du pourcentage d’OGM du paquet numéro i .

Modèle : On suppose que les X_i sont indépendants et suivent une loi gaussienne $\mathcal{N}(\theta, 1)$.

1. Pour $\theta_1 > \theta_0$, montrer que le test de Neyman-Pearson de niveau α de $H_0 : \theta = \theta_0$ contre $H_1 : \theta = \theta_1$ est de la forme $\bar{X}_n > t_{n,\alpha}$.
2. Pour le fabricant, le pourcentage d’OGM est inférieur à 1% sauf preuve du contraire. Il veut tester l'hypothèse $H_0 : \theta \leq 0$ contre $H_1 : \theta > 0$ et il souhaite que pour $\theta \leq 0$ le test se trompe avec une probabilité inférieure à 5%. Calculer un seuil $t_{25,5}$ tel que

$$\sup_{\theta \leq 0} \mathbb{P}_\theta(\bar{X}_{25} > t_{25,5}) = 5\%.$$

On pourra utiliser que $\mathbb{P}(Z > 1.645) \approx 5\%$, pour $Z \sim \mathcal{N}(0, 1)$.

3. Une association “anti-OGM” veut s’assurer qu’il n’y a effectivement pas plus de 1% d’OGM dans les produits labélisés “bio”. En particulier, elle s’inquiète de savoir si le test parvient à éliminer les produits pour lesquels le pourcentage d’OGM dépasse de 50% le maximum autorisé. Quelle est la probabilité que le test ne rejette pas H_0 lorsque le pourcentage d’OGM est de 1.5% ?
4. Scandalisée par le résultat précédent, l’association milite pour que le test du fabricant prouve effectivement que le pourcentage d’OGM est inférieur à 1%. Pour elle, le pourcentage d’OGM est supérieur à 1% sauf preuve du contraire, donc H_0 est $\theta > 0$ et H_1 est $\theta \leq 0$. Proposer un test de H_0 contre H_1 tel que la probabilité que le test rejette à tort H_0 soit inférieure à 5%.

Exercice 3 (Maximum de vraisemblance et test de Fisher pour le modèle linéaire simple):
Pour tout $i = 1 \cdots n$, on considère

$$Y_i = x_i \theta_1 + V_i \quad (1)$$

où $V_i \sim \mathcal{N}(0, \sigma^2)$ et où les V_i sont indépendants. Les coefficients x_i sont des variables déterministes connues. Les paramètres θ_1 et σ^2 sont inconnus. On observe $Y = (Y_1, \dots, Y_n)$. On paramètre le modèle pour Y par $\theta = (\theta_1, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$.

1. Montrer que $Y = \theta_1 A + V$, où A est un vecteur constant que l'on déterminera et V est un vecteur Gaussien dont on donnera la matrice de variance-Covariance en fonction de σ^2 . En déduire la densité $p_\theta(y)$ de Y par rapport à la mesure de Lebesgue n -dimensionnelle dy_1, \dots, dy_n .
2. Exprimer l'estimateur $\hat{\theta} = (\hat{\theta}_1, \hat{\sigma}^2)$ du maximum de vraisemblance pour le paramètre de $\theta = (\theta_1, \sigma^2)$.
3. On s'intéresse aux hypothèses :

$$\begin{aligned} H_0 : \quad & \theta_1 = 0 \\ H_1 : \quad & \theta_1 \neq 0 . \end{aligned}$$

Fournir un test de niveau α donné en utilisant un critère basé sur la statistique du rapport de vraisemblance « généralisé », c'est-à-dire sur la statistique

$$Z(Y) = \frac{\sup_{\theta \in \Theta} p(\theta, Y)}{\sup_{\theta \in \Theta_0} p(\theta, Y)},$$

où $\theta \mapsto p(\theta, Y) = p_\theta(Y)$ est la vraisemblance du modèle. Pour cela, on exprimera Z sous la forme d'un rapport d'estimateurs de variance. En conclure que le test de Fisher revient à rejeter H_0 lorsque la variance estimée sous l'hypothèse nulle est « trop » élevée.