Feuille de travaux dirigés 5 : Tests, Neyman-Pearson

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 des carrés de n variables normales centrées réduites indépendantes :

$$X \sim \chi_n^2 \iff X \overset{\text{loi}}{=} \sum_1^n U_i^2 \text{ où } U_i \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$$

Exercice 2 (Dosages)**:**

On souhaite tester si la concentration d'un produit est la même dans deux bassins différents A et B dans lesquels vivent des poissons d'élevage. A cet effet, des dosages sont effectuées dans les deux bassins et ont donné les résultats suivants :
Bassin A : 12, 14, 13, 13 (en mg/l)
Bassin B : 11, 13, 12 (en mg/l)

On admet que le résultat d'un dosage est une réalisation d'une variable aléatoire gaussienne dont l'espérance est la concentration du produit dans le bassin choisi (μ_A et μ_B respectivement pour les bassins A et B). On admet également que tous les dosages sont effectués de manière indépendante. On supposera que l'écart-type, pour la méthode de mesure utilisée, est égal à $\sigma = 1$ mg/l.

On cherche un test de l'hypothèse H_0 : "$\mu_A = \mu_B$" contre H_1 : "$\mu_A > \mu_B$".

1. On note n_A et n_B les tailles respectives des échantillons A et B et $\bar{X}_A$ et $\bar{X}_B$ les moyennes empiriques respectives des échantillons A et B. On considère la statistique de test

$$T(X) = \bar{X}_A - \bar{X}_B.$$

   Quelle est la loi de $T(X)$ sous l'hypothèse nulle ?

   Dans la suite on suppose que la seule observation disponible pour le statisticien est $T(X)$.
2. On considère provisoirement le test de l'hypothèse nulle H_0 contre

$$\tilde{H}_1 : \mu_A - \mu_B = \Delta$$

où $\Delta > 0$ est fixé. Montrer que le test de Neyman-Pearson de niveau $\alpha = 5\%$ revient à comparer $T(X)$ un seuil C que l'on précisera en fonction n_A, n_B, σ, et du quantile de niveau u de la loi gaussienne standard $\mathcal{N}(0,1)$, noté $q_{\mathcal{N}}(u)$, où u est à déterminer.

3. Quelle est la réponse de votre test pour les valeurs numériques données dans l'énoncé ? On donne $q_{\mathcal{N}}(0.95) \simeq 1.645, q_{\mathcal{N}}(0.975) \simeq 1.96$, $\sqrt{2} \simeq 1.414$.
4. Montrer que le test construit à la question 2 est aussi un test de niveau $\alpha = 5\%$ de H_0 contre H_1.
5. Montrer que le test construit à la question 2 est uniformément plus puissant de niveau α pour tester l'hypothèse H_0 contre H_1.

Exercice 3 (Test pour une certification bio)**:**
Pour avoir la certification “bio”, un fabriquant 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 en base 10 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 fabriquant, 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 fabriquant 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%.