TP 7 - Test de Neyman Pearson. Rapport de vraisemblance monotone.

Exercice 1

On considère un échantillon $(X_1, \dots, X_n)$ de loi $\mathcal{N}(\theta, \sigma^2)$. On souhaite tester l'hypothèse H_0 : « $\theta = \theta_0$ »contre l'alternative H_1 : « $\theta = \theta_1$ avec $\theta_0 < \theta_1$

1. Montrer que le test de rapport de vraisemblance est fondé sur la statistique $S = X_1 + \cdots + X_n$.
2. Sachant que si Y suit une loi $\mathcal{N}(0,1)$, alors $\mathbb{P}(Y > 1.64) = 0.05$, expliciter la région de rejet du test de niveau 5% dans notre cas.
3. Proposer une estimation par minoration de la puissance de ce test.
4. Que suggérez-vous pour augmenter la puissance du test ?
5. On désire cette fois tester l'hypothèse $\mathcal{H}_0$ « $\theta \leq 0$ »contre l'hypothèse $\mathcal{H}_1$ « $\theta \geq 0$ ». Montrer que le modèle est une famille à rapport de vraisemblances monotones.
6. Donner le test de puissance maximale de niveau α donné.
7. Un radar actif de surveillance aérienne a des caractéristiques telles qu'une éventuelle cible réfléchit 20 impulsions lors d'un balayage. À l'aide d'un traitement adapté, ces N impulsions réfléchies en cas de présence de la cible fournissent un vecteur d'observations (z_i) avec
$$\mathcal{H}_1 : z_i = A + b_i \qquad \text{en présence de cible}$$
$$\mathcal{H}_0 : z_i = b_i \qquad \text{en l'absence de cible}$$
où les b_i sont des variables aléatoires gaussiennes $\mathcal{N}(0, \sigma^2)$ indépendantes modélisant les divers bruit.
   (a) Donner le test de Neyman Pearson de niveau α de $\mathcal{H}_0$ contre $\mathcal{H}_1$. Application numérique $A = 1, \sigma = 0.6$ et $\alpha = 10^{-6}$.
   (b) Écrire un code matlab simulant n échantillons, chacun de taille N sous l'hypothèse nulle. Appliquer pour chacun des échantillons le test de N-P et calculer la fréquence empirique des fausses détections.
   (c) Faire la même chose sous l'hypothèse $\mathcal{H}_1$ et calculer la fréquence empirique des cibles non détectées.
8. Une antenne de communication cherche à détecter l'émission d'un signal. À cause des perturbations atmosphériques, même en l'absence de signal, l'antenne perçoit du « bruit ». On procède à n mesures et on modélise le bruit par des v.a.i.i.d. de loi gaussienne $\mathcal{N}(0, \sigma^2)$ et le signal par des v.a.i.i.d. $\mathcal{N}(0, \xi^2)$, signal et bruit étant considérés indépendants. On veut tester l'hypothèse $\mathcal{H}_0$ « pas de signal »contre l'alternative « signal ».
   (a) Formaliser le problème de test dans notre situation.
   (b) Déterminer la statistique de test du rapport de vraisemblance. En déduire que le test est fondé sur une statistique de loi du χ^2.
   (c) Comment déterminer la région de rejet pour obtenir un test de niveau α donné et calculer sa puissance.

Exercice 2

Soient X et Y deux variables aléatoires. On veut tester l'hypothèse $\mathcal{H}_0$: « X et Y ont même loi »contre $\mathcal{H}_1$: « X et Y n'ont pas même loi ». On dispose pour cela d'un échantillon $(X_1, \dots X_m)$ de loi celle de X et d'un échantillon $(Y_1, \dots Y_n)$ de loi celle de Y.

Toutes ces $m+n$ variables aléatoires étant indépendantes. On range par ordre croissant les $m+n$ valeurs effectivement observées et après effacement des indices on obtient un « mot »de la forme $XXYXYYYXYXX$. On compte alors le nombre R de blocs de lettres identiques (les longueurs ou « runs »ou composantes connexes) dans ce mot, (sur cet exemple, $R = 7$ soit 4 longueurs de X et 3 de Y).

L'idée intuitive de ce test est que c'est sous $\mathcal{H}_0$ que le mélange entre les valeurs de X et celles de Y sera le plus intense et donc le nombre de longueur le plus élevé.

1. Sous $\mathcal{H}_0$, calculer la loi de R par les formules suivantes où l'on suppose $m \leq n$:

$$P(R = 2k) \qquad 1 \leq k \leq m$$

$$P(R = 2k+1) \qquad 1 \leq k < m$$

$$P(R = 2m+1)$$

2. Lorsque m et n sont grands, quelle est la distribution (sous $\mathcal{H}_0$) asymptotiquement obtenue pour R (répondre grâce à Matlab) ? On admettra que

$$\mathbb{E}[R] = 1 + \frac{2mn}{m+n} \qquad Var(R) = \frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}$$

3. Définir un test de niveau ϵ de rejet de $\mathcal{H}_0$ à partir de R.
4. Écrire une fonction Matlab qui calcule R pour les échantillons X et Y.
5. Écrire une fonction Matlab qui réalise le test des longueurs au niveau asymptotique α pour m et n grands.
6. Écrire une fonction Matlab qui calcule la fonction de répartition de R. La comparer avec la fonction de répartition de la loi gaussienne correspondant à l'approximation utilisée ci-dessus.

On verra dans les tps suivants d'autres tests asymptotiques ou non donnant des règles d'acceptation ou de rejet d'hypothèse d'adéquation à la même loi pour deux populations.