TD 3 STATISTIQUE - 1SN

Exercice 1.
On veut tester l'hypothèse H_0 qu'une pièce de monnaie est parfaite contre l'hypothèse H_1 qu'elle est truquée. On jette 100 fois cette pièce et on élabore la règle de décision suivante :

On rejette H_0 si le nombre de faces obtenu lors des 100 *lancers est inférieur strictement à* 48 *ou supérieur strictement à* 52.

1. Calculer le risque de première espèce α.
2. Calculer une valeur approchée de ce risque en utilisant la convergence en loi de la loi binomiale vers la loi normale (la loi normale étant une variable aléatoire absolument continue, on prendra soin de rectifier l'intervalle $[48; 52]$ par $[47.5, 52.5]$).
3. On réalise l'expérience consistant à jeter 100 fois la pièce et on observe 47 fois face. Que conclure avec ce test ? Que pensez vous de cette décision ?

Exercice 2. On étudie une population dans laquelle les tirages sont supposés indépendants et de loi normale $N(m, 1)$. On souhaite tester l'hypothèse $H_0 : m = 0$ contre l'hypothèse $H_1 : m = 1$ au moyen d'un échantillon de taille $N = 2$, et prendre une décision avec un risque de première espèce $\alpha = 5\%$.

1. On considère le test T_1 défini par la règle de décision suivante :

   $$\text{Rejet de } H_0 \text{ si } x_1 + x_2 > k$$

   - Déterminer la loi de $X_1 + X_2$ sous l'hypothèse H_0.
   - En déduire la valeur de k sachant que $\alpha = 5\%$.
   - Déterminer la région critique du test et représenter la graphiquement dans le plan.
   - Calculer le risque de seconde espèce et la puissance du test.
2. On considère un second test T_2 défini par la règle de décision :

   $$\text{Rejet de } H_0 \text{ si } \inf(x_1, x_2) > l$$

   - Déterminer la valeur de l pour $\alpha = 5\%$.
   - Déterminer la région critique du test et représenter la graphiquement dans le plan.
   - Calculer la puissance du test T_2.

Exercice 3.

Un radar de surveillance envoie un train d'impulsions (1200 impulsions par seconde) à l'aide d'une antenne en rotation uniforme (15 tours par minute) avec une largeur de lobe principal de $1,5°$.

1. Quel est le temps pendant lequel le lobe principal de l'antenne attaque la cible pendant un seul balayage et le nombre d'impulsions envoyées pendant ce temps noté N ?

2. A l'aide d'un prétraitement adéquat, ces N impulsions réfléchies lors de la présence de la cible fournissent un vecteur d'observation à N composantes $z = (z_1, ..., z_N)$ avec

$$\begin{aligned} z_i &= A + b_i \text{ en présence de cible (hypothèse } H_1) \\ z_i &= b_i \text{ en l'absence de cible (hypothèse } H_0) \end{aligned}$$

où les B_i sont des va gaussiennes centrées indépendantes de variance σ^2 modélisant les divers bruits. On désire réaliser la détection d'une cible à l'aide du détecteur de Neyman-Pearson. Expliciter ce détecteur et déterminer la puissance du test. Pour l'application numérique, on prendra $A = 1, \sigma = 0.6, N = 20$, $\alpha = 10^{-10}$ et on admettra que pour $x > 5$, on a :

$$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} < \frac{1}{2} e^{-\frac{x^2}{2}}$$

et que $\frac{1}{2}e^{-\frac{x^2}{2}}$ donne une bonne approximation de l'ordre de grandeur de $Q(x)$.

Remarque

La forte valeur de β peut paraître disproportionnée par rapport à celle de la probabilité de fausse alarme. Cependant, il ne faut pas oublier que l'opération de détection s'effectue à chaque balayage (chaque tour de l'antenne i.e. toutes les 4s) et que celle-ci s'accompagne d'une estimation des paramètres de la cible permettant de la poursuivre (d'où l'importance donnée à la minimisation des fausses détections, celles-ci pouvant rapidement saturer les algorithmes de poursuite).

Exercice 4.

On dispose de n observations $x_1, ..., x_n$ issues d'un échantillon $(X_1, ..., X_n)$ de loi normale $\mathcal{N}\left(0, \sigma^2\right)$.

1. Déterminer suivant les valeurs de σ_0^2 et de σ_1^2, la statistique $T(X_1, ..., X_n)$ et la région critique du test de Neyman-Pearson associé au problème suivant :

$$\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 = \sigma_1^2 \end{cases}$$

   Commenter la forme de la statistique de test. Dans la suite de ce problème, on supposera $\sigma_1 > \sigma_0$.

2. Déterminer la valeur du risque de seconde espèce β lorsque

$$\alpha = 0.01, \sigma_0^2 = 1, \sigma_1^2 = 2 > 0 \text{ et } n = 60.$$

3. Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) associées au problème précédent. On posera

$$\Phi_n(x) = \int_x^{+\infty} f_n(u)\, du$$

   où $f_n(u)$ est la densité d'une loi du χ_n^2 et on notera $\Phi_n^{-1}(x)$ son inverse. Comment la puissance du test de Neyman Pearson dépend-elle de σ_0 et de σ_1 ? Représenter la forme approximative des courbes COR pour différentes valeurs de ces paramètres.

4. Quelle est la loi asymptotique de la statistique $T(X_1, ..., X_n)$ sous les deux hypothèses H_0 et H_1 ? En utilisant cette loi asymptotique, déterminer la valeur du risque de seconde espèce β lorsque $\alpha = 0.01, \sigma_0^2 = 1, \sigma_1^2 = 2 > 0$ et $n = 100$.

Exercice 1

1)

$$\begin{aligned} \alpha &= 1 - 2P_{48} - 2P_{49} - P_{50} \\ P_k &= \left(\frac{1}{2}\right)^{100} \mathcal{C}_{100}^{k} \end{aligned}$$

AN : $\alpha \sim 0.6173$

2) Si N_f désigne le nombre de faces, on a :

$$\begin{aligned} \alpha &= 1 - P\left[-0.40 \leq \frac{N_f - 50}{5} \leq 0.40\right] \\ &= 1 - F_{0,1}(0.40) + F_{0,1}(-0.40) \sim 0.6892 \end{aligned}$$

3) En rectifiant l'intervalle $[48; 52]$ par $[47.5, 52.5]$, on obtient :

$$\begin{aligned} \alpha &= 1 - P\left[-0.50 \leq \frac{N_f - 50}{5} \leq 0.50\right] \\ &= 1 - F_{0,1}(0.50) + F_{0,1}(-0.50) \sim 0.6170 \end{aligned}$$

et donc on voit que l'approximation est meilleure.

Exercice 2

1) On considère le test T_1 défini par la règle de décision suivante :

Rejet de H_0 si $x_1 + x_2 > k$

a) La loi de $X_1 + X_2$ sous l'hypothèse H_0 est une loi normale $\mathcal{N}(0, 2)$

b)

$$\begin{aligned} \alpha &= 0.05 = P\left[\text{Rejeter } H_0 \middle| H_0 \text{ vraie}\right] \\ &= P\left[X_1 + X_2 > k \middle| X_1 + X_2 \sim \mathcal{N}(0, 2)\right] \\ &= P\left[\frac{X_1 + X_2}{\sqrt{2}} > \frac{k}{\sqrt{2}} \middle| \frac{X_1 + X_2}{\sqrt{2}} \sim \mathcal{N}(0, 1)\right] \end{aligned}$$

Les tables donnent

$$\frac{k}{\sqrt{2}} \approx 1.64 \text{ d'où } k \approx 2.32$$

c) La région critique du test est le demi plan supérieur délimité par la droite d'équation

$$x_1 + x_2 = k$$

d) Le risque de seconde espèce est

$$\begin{aligned} \beta &= P\left[\text{Rejeter } H_1 \middle| H_1 \text{ vraie}\right] \\ &= P\left[X_1 + X_2 \leq k \middle| X_1 + X_2 \sim \mathcal{N}(2, 2)\right] \\ &= P\left[\frac{X_1 + X_2 - 2}{\sqrt{2}} > \frac{k - 2}{\sqrt{2}} \middle| \frac{X_1 + X_2 - 2}{\sqrt{2}} \sim \mathcal{N}(0, 1)\right] \end{aligned}$$

Les tables donnent

$$\beta \approx 0.59$$

d'où la puissance du test

$$\pi = 1 - \beta \approx 0.41$$

2) On considère un second test T_2 défini par la règle de décision :

$$\text{Rejet de } H_0 \text{ si } \inf(x_1, x_2) > l$$

a)

$$\begin{aligned}
\alpha &= 0.05 = P\left[\text{Rejeter } H_0 |\, H_0 \text{ vraie}\right] \\
&= P\left[\inf(X_1, X_2) > l |\, \mu = 0\right] \\
&= P\left[X_1 > l \text{ et } X_2 > l |\, \mu = 0\right] \\
&= P\left[X_1 > l |\, \mu = 0\right] P\left[X_2 > l |\, \mu = 0\right] \\
&= P\left[X_1 > l |\, X_1 \sim \mathcal{N}(0,1)\right] P\left[X_2 > l |\, X_2 \sim \mathcal{N}(0,1)\right] \\
&= \left[\int_l^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du\right]^2
\end{aligned}$$

d'où

$$\int_l^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = \sqrt{0.05}$$

Les tables donnent

$$l \approx 0.76$$

b) La région critique du test est définie par $x_1 > l$ et $x_2 > l$.

c) Le risque de seconde espèce est

$$\begin{aligned}
\beta &= P\left[\text{Rejeter } H_1 |\, H_1 \text{ vraie}\right] \\
&= 1 - P\left[\text{Rejeter } H_0 |\, H_1 \text{ vraie}\right] \\
&= 1 - P\left[X_1 > l \text{ et } X_2 > l |\, \mu = 1\right] \\
&= 1 - \left[\int_{l-1}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du\right]^2
\end{aligned}$$

d'où la puissance du test

$$\begin{aligned}
\pi &= 1 - \beta \\
&= \left[\int_{0.76-1}^\infty \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du\right]^2 \\
&\approx 0.354
\end{aligned}$$

Exercice 3

1) Lors d'un tour, l'antenne est en vis à vis avec la cible pendant $\frac{1{,}5}{360} \times \frac{60}{15} = \frac{1}{60} s$. Le nombre d'impulsions réflèchies pendant cette période est $N = 1200 \times \frac{1}{60} = 200$.

2) Le détecteur de Neyman-Pearson pour une telle situation a été vu en cours :

$$\text{Rejet de } H_0 \text{ si } \overline{Z}_n = \frac{1}{n}\sum_{i=1}^{n} Z_i > K_\alpha$$

où K_α est un seuil dépendant de la probabilité de fausse alarme α. La détermination de K_α se fait à l'aide de l'équation

$$\alpha = Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du$$

avec $x = \frac{K_\alpha \sqrt{n}}{\sigma}$. On montre que pour $x > 5$, on a

$$Q(x) = \int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} < \frac{1}{2} e^{-\frac{x^2}{2}}$$

et que $\frac{1}{2} e^{-\frac{x^2}{2}}$ donne une bonne approximation de l'ordre de grandeur de $Q(x)$. On en déduit alors

$$\frac{K_\alpha \sqrt{n}}{\sigma} \simeq \sqrt{-2\ln(2\alpha)}$$

soit

$$K_\alpha \simeq \sigma \sqrt{\frac{-2\ln(2\alpha)}{n}}$$

Une application numérique donne $K_\alpha \simeq 0.89$. Vous vérifierez alors aisément que $\beta \simeq 0.20$.

Exercice 4

1) Des calculs classiques permettent d'obtenir la règle de décision suivante :

$$\begin{aligned}
\text{Si } \sigma_1 &> \sigma_0, \text{ rejet de } H_0 \text{ si } T = \sum_{i=1}^{n} X_i^2 > S_\alpha \\
\text{Si } \sigma_1 &< \sigma_0, \text{ rejet de } H_0 \text{ si } T = \sum_{i=1}^{n} X_i^2 < S_\alpha.
\end{aligned}$$

La statistique du test est donc $T = T(X_1, ..., X_n) = \sum_{i=1}^{n} X_i^2$, c'est-à-dire n fois le moment d'ordre 2 empirique des variables X_i. Ceci semble naturel puisque $E\left[X_i^2\right] = \sigma^2$ varie sous les deux hypothèses H_0 et H_1.

2) Le risque de première espèce est :

$$\begin{aligned}
\alpha &= P\left[\text{rejeter } H_0 \mid H_0 \text{ vraie}\right] \\
&= P\left[T > S_\alpha \mid \sigma = \sigma_0\right] \\
&= P\left[\frac{1}{\sigma_0^2} T > \frac{S_\alpha}{\sigma_0^2} \middle| \sigma = \sigma_0\right].
\end{aligned}$$

On sait que $\frac{1}{\sigma_0^2} T$ suit une loi du chi2 à n degrés de liberté sous l'hypothèse H_0, donc

$$\alpha = \Phi_n\left(\frac{S_\alpha}{\sigma_0^2}\right) \text{ i.e. } S_\alpha = \sigma_0^2 \Phi_n^{-1}(\alpha).$$

Pour $\sigma_0 = 1$ et $n = 60$, on obtient $S_\alpha \simeq 88.38$. Le risque β peut alors se calculer comme suit :

$$\begin{aligned}
\beta &= P\left[\text{rejeter } H_1 | \, H_1 \text{ vraie}\right] \\
&= P\left[T \leq S_\alpha | \, \sigma = \sigma_1\right] \\
&= 1 - P\left[T > S_\alpha | \, \sigma = \sigma_1\right] \\
&= 1 - \Phi_n\left(\frac{S_\alpha}{\sigma_1^2}\right).
\end{aligned}$$

Les tables de la loi du chi2 permettent d'obtenir $\beta \in [0.05, 0.1]$.
3) Les courbes COR sont définies par

$$\begin{aligned}
\text{PD} &= 1 - \beta = \Phi_n\left(\frac{S_\alpha}{\sigma_1^2}\right) \\
\text{PD} &= \Phi_n\left[\frac{\sigma_0^2}{\sigma_1^2}\Phi_n^{-1}(\alpha)\right].
\end{aligned}$$

On voit donc que PD est une fonction décroissante de $\frac{\sigma_0^2}{\sigma_1^2}$, c'est-à-dire une fonction croissante de $\frac{\sigma_1^2}{\sigma_0^2}$, ce qui est normal.
4) Le théorème de la limite centrale nous permet d'approcher, pour n "grand", la loi de $T(X_1, ..., X_n) = \sum_{i=1}^n X_i^2$ par une loi normale $\mathcal{N}\left(nE\left[X_1^2\right], n\text{var}\left[X_1^2\right]\right)$. Mais

$$\begin{aligned}
E\left[X_1^2\right] &= \sigma^2 \\
\text{var}\left[X_1^2\right] &= E\left[X_1^4\right] - E\left[X_1^2\right]^2 = 3\sigma^4 - \sigma^4 = 2\sigma^4.
\end{aligned}$$

Donc, on peut approcher la loi de T par une loi normale $\mathcal{N}\left(n\sigma^2, n2\sigma^4\right)$. En posant $\Psi(u) = \int_u^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$, on en déduit :

$$\begin{aligned}
\alpha &= P\left[\text{rejeter } H_0 | \, H_0 \text{ vraie}\right] \\
&= P\left[\frac{T - n\sigma^2}{\sqrt{2n}\sigma^2} > \frac{S_\alpha - n\sigma^2}{\sqrt{2n}\sigma^2} \middle| \, \sigma = \sigma_0\right] \\
&= \Psi\left(\frac{S_\alpha - n\sigma_0^2}{\sqrt{2n}\sigma_0^2}\right)
\end{aligned}$$

d'où $S_\alpha = \Psi^{-1}(\alpha)\sqrt{2n}\sigma_0^2 + n\sigma_0^2$ et

$$\begin{aligned}
\beta &= P\left[\text{rejeter } H_1 | \, H_1 \text{ vraie}\right] \\
&= P\left[T \leq S_\alpha | \, \sigma = \sigma_1\right] \\
&= 1 - P\left[T > S_\alpha | \, \sigma = \sigma_1\right] \\
&= 1 - \Phi_n\left(\frac{S_\alpha - n\sigma_1^2}{\sqrt{2n}\sigma_1^2}\right).
\end{aligned}$$

Pour $n = 100$, on obtient

$$\beta \simeq 8.9 \; 10^{-3}.$$