TD 3 STATISTIQUE - 1HY

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
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 $T\left(X_1, ..., X_n\right)$ 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 problè me, on supposera $\sigma_1 > \sigma_0$.

2. Déterminer la valeur des risques de première et seconde espèce α et β en fonction du seuil du test noté S, de la fonction de réparition d'une loi du χ_n^2 notée F_n et de σ_0 et σ_1.

3. Déterminer les courbes COR (caractéristiques opérationnelles du récepteur) associées au problème précédent. 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 de $T\left(X_1, ..., X_n\right)$ sous les deux hypothèses H_0 et H_1 issue de l'application du théorème de la limite centrale ? En utilisant cette loi asymptotique, déterminer les courbes COR du problème et montrer que leur expression en fonction de σ_0^2 et σ_1^2 est en accord avec la question précédente.

Applications : à faire

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)

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

où F est la fonction de répartition d'une loi normale $\mathcal{N}(0, 1)$. On en déduit

$$k = \sqrt{2}F^{-1}(1 - \alpha).$$

Pour $\alpha = 5\%$, les tables donnent

$$k \approx 2.32.$$

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

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

c'est-à-dire

$$\beta = F\left(\frac{k-2}{\sqrt{2}}\right).$$

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)

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

d'où

$$l = F^{-1}\left(1 - \sqrt{\alpha}\right).$$

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

$$\begin{aligned}
\pi &= 1 - \beta \\
&= \left[1 - F(l-1)\right]^2 \\
&\approx 0.354.
\end{aligned}$$

Exercice 3

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

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

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

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

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

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

Le risque β peut se calculer de la même façon

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

3) Si on appelle π la puissance du test, les courbes COR sont définies par

$$\begin{aligned} \pi &= 1 - \beta = 1 - F_n\left(\frac{S}{\sigma_1^2}\right) \\ \pi &= 1 - F_n\left(\frac{\sigma_0^2}{\sigma_1^2} F_n^{-1}(1-\alpha)\right) \end{aligned}$$

On voit donc que π 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 $\mathcal{N}\left(n\sigma^2, 2n\sigma^4\right)$. En notant G la fonction de répartition d'une loi normale $\mathcal{N}(0,1)$, on en déduit

$$\begin{aligned} \alpha &= P\left[\text{rejeter } H_0 \,|\, H_0 \text{ vraie}\right] \\ &= P\left[\frac{T - n\sigma^2}{\sqrt{2n}\sigma^2} > \frac{S - n\sigma^2}{\sqrt{2n}\sigma^2} \middle| \sigma = \sigma_0\right] \\ &= 1 - G\left(\frac{S - n\sigma_0^2}{\sqrt{2n}\sigma_0^2}\right) \end{aligned}$$

d'où

$$S = n\sigma_0^2 + \sqrt{2n}\sigma_0^2 G^{-1}\left(1-\alpha\right).$$

et

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

La puissance du test s'écrit donc

$$\pi = 1 - \beta = 1 - G\left(\frac{S - n\sigma_1^2}{\sqrt{2n}\sigma_1^2}\right).$$

Après avoir remplacé S par son expression trouvée ci-dessus, on obtient

$$\pi = 1 - \beta = 1 - G\left(-\sqrt{\frac{n}{2}} + \frac{\sigma_0^2}{\sigma_1^2}\left[\sqrt{\frac{n}{2}} + G^{-1}(1-\alpha)\right]\right).$$

qui est une fonction croissante de $\frac{\sigma_1^2}{\sigma_0^2}$, comme dans la question précédente.