SUP'COM
2024/2025
INDP2
Mme. F. Abdelkefi, Mme A. Benazza, Mme F. Chaabane et Mme. L. Najjar
10/01/2025

EXAMEN Inférence Statistique (session principale)

- *La durée de l'épreuve est de 1h30.*
- *Aucun document n'est autorisé.*
- *Seul l'usage d'une calculatrice basique et non connectée est autorisé.*

Pour certains exercices, on pourrait s'aider des rappels et tables donnés à la fin de l'énoncé.

1 Estimateur linéaire en moyenne quadratique

On considère le modèle de mesure

$$y_k = \theta + b_k, \tag{1}$$

où les b_k sont des échantillons i.i.d. issus de la réalisation d'un bruit blanc gaussien centré de variance σ^2 supposée connue. θ est l'inconnue à estimer. θ est supposée décorréleé du bruit et uniformément distribuée sur $[-\theta_0, \theta_0]$ où θ_0 est un réel strictement positif. On dispose de N observations y_i, $i = 1, \ldots, N$. L'objectif est d'estimer θ à partir des observations y_i, $i = 1, \ldots, N$ par un estimateur linéaire en moyenne quadratique (ELMQ).

1. Formuler l'ELMQ en tant qu'un problème de projection orthogonale.
2. Calculer l'ELMQ dans le cas d'une observation $N = 1$.
3. En déduire la valeur minimale de l'EQM correspondante.

2 Intervalle de confiance et test d'hypothèse

Deux modèles de batterie sont considérés avec x modélisant la durée de l'autonomie correspondante, supposées suivre, selon le modèle de batterie, deux lois gaussiennes de moyenne m_i et de variance σ_i^2, $i = 1, 2$.

— M1 : $x \sim \mathcal{N}(m_1 = 6, \sigma_1^2 = 1)$ heures
— M2 : $x \sim \mathcal{N}(m_2 = 4, \sigma_2^2 = 2)$ heures

1. Un échantillon de N batteries de type M1 est considéré. Déterminer l'expression analytique de l'intervalle de confiance sur m_1, noté $IC_{1-\alpha} = [m_{1,min}, m_{1,max}]$.
2. Préciser cet intevalle pour $N = 100$, $\alpha = 5\%$.
3. Pour valider qu'un lot de $N = 4$ batteries du même type sont du type M1, on utilise un test d'hypothèse unilatéral avec rejet à gauche. Formuler les hypothèses H_0 et H_1.
4. On ne valide pas le type M1, si la moyenne des autonomies sur les N échantillons est inférieure à un seuil de décision S heures sur la moyenne empirique $\bar{m}_1$. Définir l'erreur de première espèce. Calculer sa valeur pour $S = 5$ heures.
5. Pour un risque de première espèce de 5%, que devient la valeur du seuil S sur la moyenne empirique $\bar{m}_1$.
6. Quand peut-on juger que le test est significatif ?

3 Détecteur binaire

On formule les hypothèses binaires sur deux modèles de batteries, où l'autonomie x en heures est modélisée selon chacune des deux hypothèses comme suit

— H_0 : $x \sim \mathcal{N}(m_0 = 6, \sigma_0^2 = 1)$ heures
— H_1 : $x \sim \mathcal{N}(m_1 = 4, \sigma_1^2 = 2)$ heures

1. Donner la règle de décision optimale au sens du risque bayésien. On considère le coût tout ou rien dans le cas d'hypothèses équiprobables et N mesures x_i, $i = 1, \dots, N$ de x supposées indépendantes identiquement distribuées.
2. Pour ce problème, discuter s'il est plus critique de faire une fausse alarme ou une détection manquée (missing-Misdetection).
3. Déterminer les régions de décision pour $N = 1$.
4. Calculer la probabilité de détection manquée et de fausse alarme pour $N = 1$. Utiliser la fonction $erf(.)$.
5. En déduire l'expression de la probabilité d'erreur. Est-elle minimale ? Justifier.

On donne

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt. \tag{2}$$

et

$$\lim_{x \to +\infty} erf(x) = 1. \tag{3}$$

4 Test de comparaison

Deux méthodes de dosage d'une substance chimique sont considérées. Les résultats des deux méthodes de dosage sont modélisées par deux lois gaussiennes indépendantes dont les paramètres (m_1, σ_1^2) et (m_2, σ_2^2) sont inconnus. L'objectif est de comparer si les procédés de dosage peuvent être considérés comme équivalents. Les résultats des dosages par chacune des méthodes, mesurés sur $N = 8$ échantillons du mélange (pour chaque procédé) donnent les moyennes et variances empiriques (non biaisées) ci-après $\bar{m}_1 = 1.2775$, $\tilde{\sigma_1}^2 = 5.664.10^{-3}$, $\bar{m}_2 = 1.2$, $\tilde{\sigma_2}^2 = 1.38.10^{-2}$.

1. Le but est de conclure sur l'équivalence ou non des deux protocoles de dosage. Indiquer -dans l'ordre- les deux tests de comparaison à effectuer. Indiquer pour chaque test, les hypothèses à tester et les issues et décisions envisageables.
2. On se restreint dans ce qui suit au premier test de comparaison. Indiquer la statistique de décision.
3. Déterminer en le justifiant sa distribution de probabilité et ses paramètres.
4. Déterminer les régions de décision pour une erreur de première espèce $\alpha = 5\%$.
5. Indiquer le résultat du test au vu des mesures effectuées.
6. Reprendre ce test dans le cas unilatéral avec rejet à droite.
   (a) Déterminer les régions de décision pour une erreur de première espèce $\alpha = 5\%$.
   (b) Indiquer le résultat du test au vu des mesures effectuées.

Rappels

• La somme des carrés de N variables aléatoire gaussiennes centrées réduites indépendantes suit une loi de Khi 2 à $N - 1$ degrés de liberté.
• Soient U et V deux variables indépendantes qui suivent des lois de Khi 2 de degrés de libertés respectifs N_1 et N_2, alors la variable aléatoire $F = \frac{U}{V}$ suit une loi de Fisher-Snedecor $\mathcal{F}(N_1, N_2)$ de degrés de liberté N_1 et N_2.

ANNEXE

On fournit les tables de la loi Normale et de Fisher-Snedecor.