Site : ☐ Luminy ☒ St-Charles ☐ St-Jérôme ☐ Cht-Gombert ☐ Aix-Montperrin ☐ Aubagne-SATIS
Sujet Session de ☒ 1er semestre - ☐ 2ème semestre - ☒ Session unique Durée de l'épreuve : 3h
Examen de ☐ L1/☐ L2/☐ L3 - ☒ M1/☐ M2 - ☐ LP - ☐ DU Nom Diplôme : MAS
Code Apogée du module : SMSAU02C Libellé du module : **Statistique.**
Document autorisé : ☒ OUI-☐ NON Calculatrice autorisée ☒ OUI-☐ NON

Examen final du jeudi 21 décembre 2023.

Les documents, notes de cours et calculatrices sont interdits, à l'exception d'une feuille A4 recto-verso. Le soin apporté à la rédaction sera un élément d'appréciation. Le barême est indicatif.

Exercice 1 (4 points). Une classe de 29 étudiants vient de passer un examen de Logiciel R. Les notes des étudiants varient de 5/20 à 18/20, et on suppose qu'elles sont des réalisations d'une loi Normale. La moyenne de la classe vaut 11,5. La variance est égale à 9. Peut-on affirmer, avec un risque de 5%, que la moyenne de cet examen est strictement plus grande que 11,2 ?

N.B. : On rappelle les quantiles suivants, où $t_{n,\alpha}$ désigne le quantile d'ordre α de la loi de Student à n degrés de liberté, et $q_\alpha^{N(0,1)}$ celui de la loi gaussienne centrée réduite.

α	95%	97,5%
$q_\alpha^{N(0,1)}$	1,64	1,96
$t_{28,\alpha}$	1,701	2,048

Exercice 2 (10 points). Soit une variable aléatoire continue X, de densité :

$$f_\theta(x) = \begin{cases} b - \frac{\theta}{2} & \text{si} \quad -2 < x \leq 0 \\ b + \frac{\theta}{2} & \text{si} \quad 0 < x < 2 \end{cases}$$

avec b une constante à déterminer. Par ailleurs, le paramètre θ est inconnu, avec $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$.
On considère un n-échantillon $(X_1, \ldots, X_n)$ de cette loi.

1. (0,5 point) Trouver la valeur de b.
2. (0,5) Calculer l'espérance de la variable aléatoire X. En déduire un estimateur de θ par la méthode des moments, qui sera noté $\hat{\theta}_n^{(1)}$ dans la suite.
3. (1) Calculer $Var_\theta(X)$. Est-elle > 0 pour tout $\theta \in \left]-\frac{1}{2}, \frac{1}{2}\right[$? En déduire $Var_\theta(\hat{\theta}_n^{(1)})$.
4. (0,75) Etudier maintenant le biais et la convergence de $\hat{\theta}_n^{(1)}$.
5. (0,5) Calculer la probabilité $P_\theta(0 < X < 2)$.
6. (0,5) On introduit la variable aléatoire $Y = 1_{0<X<2}$, ainsi que le n-échantillon correspondant $Y_i = 1_{0<X_i<2}$ pour $i = 1, \ldots, n$. Donner la loi de Y.

7. (0,5) Calculer la probabilité $P_\theta(-2 < X < 2)$, et déduisez quelles sont les valeurs possibles pour la variable aléatoire X.
8. (0,5) En déduire la relation entre les variables aléatoires $1_{-2<X\leq 0}$ et $1_{0<X<2}$
9. (0,75) On remarque que la densité de la variable aléatoire X peut aussi s'écrire

$$f_\theta(x) = \left(\frac{1}{4} - \frac{\theta}{2}\right)^{1_{-2<X\leq 0}} \left(\frac{1}{4} + \frac{\theta}{2}\right)^{1_{0<X<2}}.$$

Utilisez cette expression pour écrire la vraisemblance du n-échantillon $X_1, \ldots, X_n$, en fonction de $S_n = \sum_{i=1}^n 1_{0<X_i<2}$.

10. (1) En déduire l'estimateur du maximum de vraisemblance de θ.
11. (1,5) Considérons l'estimateur : $\hat{\theta}_n^{(2)} = \dfrac{S_n}{n} - \dfrac{1}{2}$. En étudier biais, lconvergence et l'efficacité.
12. (0,5) Quel estimateur faut-il privilégier entre $\hat{\theta}_n^{(1)}$ et $\hat{\theta}_n^{(2)}$, et pourquoi ?
13. (1,5) Ecrire le théorème central limite pour la suite de variables aléatoires $Y_1, \ldots, Y_n$. En utilisant les résultats des questions précédentes, et en considérant $\hat{\theta}_n^{(2)}$ comme estimateur ponctuel de θ, proposer un intervalle de confiance de seuil asymptotique α (ou de niveau de confiance asymptotique $1-\alpha$) pour le paramètre θ.

Exercice 3 (6 points). On s'intéresse à un portefeuille d'assurance automobile, et plus particulièrement au nombre d'accidents de cette population assurée. On a observé $\mathbf{x} = (x_1, \cdots, x_n)$ le nombre d'accidents de n conducteurs, et on suppose que $\mathbf{x}$ est une réalisation d'un n-échantillon $\mathbf{X} = (X_1, \cdots, X_n)$ de loi de Poisson, dont le paramètre Θ est lui-même une variable aléatoire de loi Gamma $\Gamma(p, \lambda)$ avec $p, \lambda > 0$. Cela permet de représenter la diversité des comportements conducteur (et le risque associé). On rappelle que

— la fonction de masse d'une variable aléatoire X de loi Poisson $\mathcal{P}(\theta)$ est donnée par (pour x à valeurs dans $\mathbb{N}$)

$$P_\theta(X = x) = \exp(-\theta)\frac{\theta^x}{x!}.$$

On a alors $E_\theta(X) = Var_\theta(X) = \theta$.

— la densité d'une loi Gamma, $\Gamma(p, \lambda)$, est donnée par (pour θ à valeurs dans $\mathbb{R}^{+*}$)

$$f(\theta; p, \lambda) = \frac{\theta^{p-1}\exp(-\lambda\theta)}{\Gamma(p)\lambda^{-p}}, \text{avec } \Gamma(.) \text{ la fonction Gamma.}$$

On rappelle que $E[\Theta] = p/\lambda$ et que $Var(\Theta) = p/\lambda^2$.

1. (1) Quelle est la loi a posteriori de $\Theta \,|\, \mathbf{X}$?
2. (1,5) Donner un estimateur de Bayes associé à la perte quadratique, noté ensuite $T^\star(\mathbf{X})$.
3. (1,5) Montrer que pour tout $\theta > 0$, le risque quadratique de cet estimateur vaut

$$R(\theta, T^\star) = \frac{n\theta + \lambda^2(\theta - \frac{p}{\lambda})^2}{(n+\lambda)^2}.$$

4. (1) En déduire que le risque bayésien vaut $R^B = \dfrac{p}{\lambda(n+\lambda)}$.
5. (1) Montrer qu'un estimateur de Bayes associé à la perte $l(\theta, T) = \theta^3\exp(-2\theta)(\theta - T)^2$ vaut

$$\tilde{T}(\mathbf{X}) = \frac{n\bar{X}_n + p + 3}{\lambda + n + 2}.$$