Examen de Statistique Inférentielle

Durée : 2h00.
La qualité de la rédaction sera prise en compte dans la notation.

Exercice 1 : IC et tests pour la loi exponentielle

On considère un échantillon $Y_1, \ldots, Y_n$ tiré selon une loi Exponentielle $\mathcal{E}(\lambda)$, d'intensité $\lambda > 0$ inconnue (Rappel : pour tout $y \geq 0$, $\mathbb{P}(Y \leq y) = \int_0^y \lambda e^{-\lambda t} dt$). Soit $\lambda_0 > 0$.

1. Calculer $\hat{\lambda}$ l'estimateur du maximum de vraisemblance de λ. Construire un intervalle (bilatéral) de confiance asymptotique $1 - \alpha$, pour λ.
2. En déduire une région critique (de niveau asymptotique $1 - \alpha$) pour le test d'hypothèses
$$H_0 : \ \lambda = \lambda_0 \quad \text{contre} \quad H_1 : \lambda \neq \lambda_0$$
3. S'agit-il d'un test convergent ?
4. Application numérique. On observe sur un échantillon de taille $n = 200$, $\sum_{i=1}^n y_i = 94$. Construire l'intervalle de confiance 0.95 pour λ et tester l'hypothèse $H_0 : \lambda = 2$ contre $H_1 : \lambda \neq 2$, avec un risque asymptotique de première espèce 0.05
5. Montrer qu'un test basé sur le principe du rapport de vraisemblance aurait conduit à une région critique de forme différente (et qu'on ne cherchera pas à expliciter complètement). Comment pourrait-on déterminer pratiquement la région critique de ce test pour un niveau α et une taille n d'échantillon donnés (expliquer la méthode de la manière la plus précise possible) ?

Exercice 2 : Ablation des amygdales et maladie de Hodgkin

Une étude médicale menée en 1971 a comparé le nombre d'individus ayant subis une ablation des amygdales pour des patients souffrant de la maladie de Hodgkin et des patients issus d'un groupe de "controle" (et qui ne sont pas porteurs de la maladie). Les résultats sont les suivants :

	Ablation des amygdales	Pas d'ablation
Maladie de Hodgkin	67	34
Controle	43	64

1. Avec un test de comparaison de proportions, peut-on dire que les personnes atteintes de la maladie de Hodgkin ont une probabilité plus grande de subir une ablation des amygdales ? (on décrira le modèle statistique, les hypothèses, la statistique de test, la région critique et la décision).
2. Une autre étude sur la même question a été menée en 1975. Elle portait sur des enfants de mêmes parents (avec une différence d'age de moins de 5 ans), dont un seul des deux était porteur de la maladie de Hodgkin. Les résultats sur 85 paires d'enfants de mêmes parents sont les suivants :

	Ablation des amygdales	Pas d'ablation
Enfant atteint de la maladie de Hodgkin	41	44
Controle (enfant indemne)	33	52

Il avait alors été conclu par cette étude (de 1975) que les résultats de l'étude précédente (de 1971) devaient être remis en cause (ou approfondis).
Qu'en pensez vous ? Quels sont les calculs qui ont été effectués ? Quelle est l'erreur qui a été commise ?

3. Les données brutes de l'étude précédente ont été reprises et sont présentées dans le tableau suivant :

		(Frère ou soeur non malade) Ablation des amygdales	Pas d'ablation
(Frère ou soeur atteint de la maladie)	Ablation des amygdales	26	15
	Pas d'ablation	7	37

Ces données peuvent être perçues comme $n = 85$ réalisations indépendantes d'une loi multinomiale avec 4 modalités, de probabilités :

		(Frère ou soeur non malade) Ablation des amygdales	Pas d'ablation
(Frère ou soeur atteint de la maladie)	Ablation des amygdales	π_{11}	π_{12}
	Pas d'ablation	π_{21}	π_{22}

Expliquez pourquoi tester l'hypothèse *"la probabilité d'avoir subi une ablation des amygdales est la même chez les personnes atteintes de la maladie que chez leur frère (ou soeur) indemne"* se traduit par le test de l'hypothèse $H_0 : \ \pi_{12} = \pi_{21}$.
On considérera l'hypothèse alternative $H_1 : \ \pi_{12} \neq \pi_{21}$

4. Construisez le test du Chi-2 correspondant. Quel est le nombre de degrés de liberté de la statistique de test ? Le résultat de ce test vous semble-t-il toujours aussi contradictoire avec le résultat du test réalisé dans la question 1 ?

Exercice 3 : coût moyen d'un patient à l'hôpital

Un hôpital souhaite estimer le coût moyen d'un patient. Le coût par jour d'un patient est de 200 €. Pour un échantillon de 500 patients on a observé une durée de séjour moyenne de 5.4 jours et un écart-type de 3.1 jours.

1. Donner un intervalle de confiance à 99% pour la durée moyenne de séjour d'un patient et en déduire un intervalle pour le coût moyen d'un patient.
2. Un autre hôpital effectue la même étude et mesure, sur un échantillon de 400 patients, une durée de séjour moyenne de 5.9 jours et un écart type de 2.7 jours. Peut-on dire que le coût moyen d'un patient est plus élevé dans cet autre hôpital (on précisera les hypothèses statistiques testées, ainsi que les hypothèses mathématiques nécessaires à la validité du test).

Quelques valeurs qui pourront être utiles

Quantiles d'une loi normale centrée réduite $\mathcal{N}(0,1)$
$q_{0.95} = 1.650 \quad q_{0.975} = 1.960 \quad q_{0.995} = 2.576$
Quantiles d'une loi du Chi-2
1 ddl : $q_{0.025} = 0.001 \quad q_{0.05} = 0.004 \quad q_{0.10} = 0.016 \quad q_{0.90} = 2.70 \quad q_{0.95} = 3.84 \quad q_{0.975} = 5.02$
2 ddl : $q_{0.025} = 0.05 \quad q_{0.05} = 0.10 \quad q_{0.10} = 0.21 \quad q_{0.90} = 4.61 \quad q_{0.95} = 5.99 \quad q_{0.975} = 7.38$
Quantiles d'une loi de Fisher-Snedecor à (399,499) ddl : $q_{0.025} = 0.83$ et $q_{0.975} = 1.20$