Contrôle Continu de Statistique Inférentielle

Durée : 2h00.
La qualité de la rédaction sera prise en compte dans la notation. Les documents, téléphones mobiles et calculatrices sont interdits.

Exercice 1.

On considère la famille de lois $\{\mathbb{P}_\theta, \theta \in]0, +\infty[\}$, continues par rapport à la mesure de Lebesgue et de densité

$$f(x, \theta) = \theta x^{\theta-1} \mathbb{1}_{]0,1[}(x).$$

On dispose d'un échantillon $X_1, \dots, X_n$ de loi $\mathbb{P}_\theta$ et on souhaite estimer le paramètre θ.

1. Vérifier que $f(x, \theta)$ est bien une densité de probabilité, calculer $\mathbb{E}_\theta(X)$ et en déduire un estimateur de θ basé sur la méthode des moments.
2. Montrer que cet estimateur est biaisé (on pourra utiliser l'inégalité de Jensen).
3. Montrer que cet estimateur est convergent en probabilité.
4. Donner sa loi asymptotique et en déduire l'expression d'un intervalle de confiance asymptotique $1 - \alpha$ pour le paramètre θ.
5. Montrer que cette famille de lois appartient à la famille exponentielle et calculer l'information de Fisher.
6. Calculer l'estimateur du maximum de vraisemblance de θ.
7. Donner sa loi asymptotique et en déduire l'expression d'un intervalle de confiance asymptotique $1 - \alpha$ pour le paramètre θ.
8. Comparer les deux intervalles de confiance.
9. On s'intéresse au pourcentage de matières grasses dans certains aliments. On note X la variable aléatoire correspondante. On constitue un échantillon de taille $n = 100$ que l'on suppose de loi $\mathbb{P}_\theta$ et on mesure le pourcentage de matière grasse. On obtient les valeurs suivantes
$$\frac{1}{100}\sum_{i=1}^{100} x_i = 0.48 \qquad \text{et} \qquad \sum_{i=1}^{100} \ln(x_i) = -112.$$
Donner les estimations de θ obtenues par les deux estimateurs.
10. En utilisant l'estimateur du maximum de vraisemblance, calculer l'IC (les approximations sont autorisées) pour θ de confiance asymptotique $1 - \alpha = 0.95$. Pourrait-on supposer que $\theta = 1$? Quelle serait alors la loi de X ?

Exercice 2

On mesure la température (en degrés Celsius) à 10h du matin dans une station météorologique pendant un mois au printemps. Les températures suivantes ont été relevées :

`0.60 0.19 6.44 2.92 ... 3.22`

Les valeurs des statistiques descriptives sont les suivantes :

$$\sum_{i=1}^{30} x_i = 66.3 \qquad \sum_{i=1}^{30} x_i^2 = 180.27$$

On modélise ces $n = 30$ mesures comme une réalisation d'un échantillon de taille n d'une loi normale $\mathcal{N}(\mu, \sigma^2)$

1. Statistiques descriptives : calculer la moyenne et la variance observée au cours du mois échantillonné.
2. Il est généralement admis qu'au cours de ce mois la variance est connue, de valeur $\sigma^2 = 1$. Construisez un intervalle de confiance $1 - \alpha = 0.95$ pour μ.
3. On ne suppose plus que la valeur de σ^2 est connue. Quelle impact cela a-t-il si on souhaite construire un intervalle de confiance pour μ. Construisez ce nouvel intervalle de confiance ($1 - \alpha = 0.95$).
4. Construisez un intervalle de confiance pour σ^2 (les approximations sont autorisées). L'hypothèse posée dans la question 2) vous semble-t-elle plausible ?

Quelques valeurs parfois utiles pour les deux exercices

— **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 de Student**
29 ddl : $q_{0.95} = 1.700 \quad q_{0.975} = 2.045 \quad q_{0.995} = 2.756$
30 ddl : $q_{0.95} = 1.697 \quad q_{0.975} = 2.042 \quad q_{0.995} = 2.750$

— **Quantiles d'une loi du** χ^2
29 ddl : $q_{0.025} = 16.05 \quad q_{0.05} = 17.71 \quad q_{0.95} = 42.56 \quad q_{0.975} = 45.72$
30 ddl : $q_{0.025} = 16.79 \quad q_{0.05} = 18.49 \quad q_{0.95} = 43.78 \quad q_{0.975} = 46.98$

— $\sqrt{29} \approx 5.39$ et $1/\sqrt{29} \approx 0.19$
— $\sqrt{30} \approx 5.48$ et $1/\sqrt{30} \approx 0.18$