La durée du partiel est 2h00. Le seul document autorisé est le formulaire de statistique. L'utilisation d'une calculatrice scientifique est autorisée si elle est non programmable. L'usage des téléphones portables est interdit.

Question de cours (6 pts)

Soit $X_1, \ldots, X_n$ un échantillon d'une loi de densité $f(x, \theta)$ où θ est un paramètre inconnu appartenant à un ensemble $\Theta \subset \mathbb{R}$ et x appartient à l'ensemble des valeurs possible de la loi.

1. Ecrire la vraisemblance et rappeler le principe du maximum de vraisemblance.
2. Définir l'information de Fisher et rappeler la loi asymptotique de l'estimateur du maximum de vraisemblance $\hat{\theta}$ du paramètre θ (on supposera que les conditions de dérivabilité sur la densité $f(x, \theta)$ sont vérifiées) .
3. L'échantillon $X_1, \ldots, X_n$ est un échantillon issu d'une loi exponentielle de paramètre λ (de densité $f(x, \lambda) = \lambda \exp(-\lambda x)$ si $x \geq 0$ et $f(x, \lambda) = 0$ si $x < 0$).
   Calculer l'estimateur du maximum de vraisemblance de λ et donner une approximation de sa loi de probabilité lorsque n est grand.

Problème (8 pts)

On souhaite étudier la part du budget des ménages qui est utilisée pour le logement. Cette proportion aléatoire, notée X, est modélisée par une loi de probabilité de densité définie par

$$f(x, \theta) = \theta x^{\theta - 1}, \text{ si } x \in [0, 1] \quad \text{ et } \quad f(x, \theta) = 0 \text{ si } x \notin [0, 1]$$

où le paramètre θ est un paramètre strictement positif, $\theta > 0$.

On dispose d'un échantillon de $n = 100$ ménages sur lequel on mesure $X_1, \ldots, X_n$. L'objectif est d'estimer le paramètre θ.

1. (a) Vérifier que $f(x, \theta)$ est bien une densité de probabilité.
   (b) Pourquoi cette famille de loi est-elle adaptée à la question posée ? A quelle loi correspond le cas particulier où $\theta = 1$?
2. (a) Montrer que $\mathbb{E}(X) = \theta/(\theta + 1)$.
   (b) En déduire un estimateur de θ basé sur la méthode des moments.
   (c) On observe $\sum_{i=1}^{100} x_i = 48$. Proposez une estimation de θ.
3. (a) Ecrire la vraisemblance et calculer l'estimateur du maximum de vraisemblance.
   (b) On observe $\sum_{i=1}^{100} \ln(x_i) = -112$. Proposez une autre estimation de θ.
4. Parmi les deux estimations obtenues, laquelle a votre préférence (justifier votre réponse).
5. Calculer l'information de Fisher et en déduire la loi asymptotique de l'estimateur du maximum de vraisemblance $\hat{\theta}$.
6. Calculer une approximation de la probabilité $\mathbb{P}[|\hat{\theta} - \theta| > 0.2]$.

Exercice (6 pts)

Les diagnostiques de performance énergétique permettent de classer les logements en 7 groupes (lettres `A` à `G`) selon qu'ils sont économes (catégorie `A`) ou très énergivores (catégorie `G`). L'unité de mesure est le $Kwh/an/m^2$. Il est donc supposé que la surface entre en compte de manière proportionnelle pour la mesure de la consommation annuelle d'un logement.

On a constitué un échantillon de $n = 12$ logements de la catégorie `B` pour lesquels on a mesuré la consommation annuelle (notée Y) et la surface (notée s).
On considère le modèle suivant

$$Y_i = \beta s_i + \epsilon_i, \quad i = 1, \ldots, 12 \qquad (*)$$

où ϵ_i est un résidu.

1. A partir du nuage de points présenté dans la Figure 1, vous semble-t-il raisonnable de considérer une relation proportionnelle entre la consommation annuelle et la la surface ?
2. En quoi serait-il naturel de considérer l'estimateur suivant de β : $\widehat{\beta} = \dfrac{\bar{Y}_n}{\bar{s}_n}$?
3. Calculer l'expression de l'estimateur noté $\widetilde{\beta}$ obtenu par la méthode des moindres carrés.
4. On a mesuré sur l'échantillon $\sum_{i=1}^{12} s_i = 1318$ et $\sum_{i=1}^{12} y_i = 10508$, $\sum_{i=1}^{12} y_i s_i = 1351219$ et $\sum_{i=1}^{12} s_i^2 = 170596$.
   (a) Donner les estimations de β obtenues par les deux méthodes,
   (b) tracer, sur la Figure 1 ci-dessous, la droite de régression estimée pour la première méthode.
5. On suppose maintenant que dans le modèle $(*)$ les résidus ϵ_i sont des variables normales indépendantes, de loi $\mathcal{N}(0, \sigma^2)$ et que, comme dans le cours, les surfaces s_i ne sont pas aléatoires.
   (a) Quelle est la loi de probabilités de Y_i ?
   (b) Ecrire la vraisemblance de l'échantillon,
   (c) Calculer l'estimateur du maximum de vraisemblance du paramètre β.
   (d) Que remarque-t-on ?

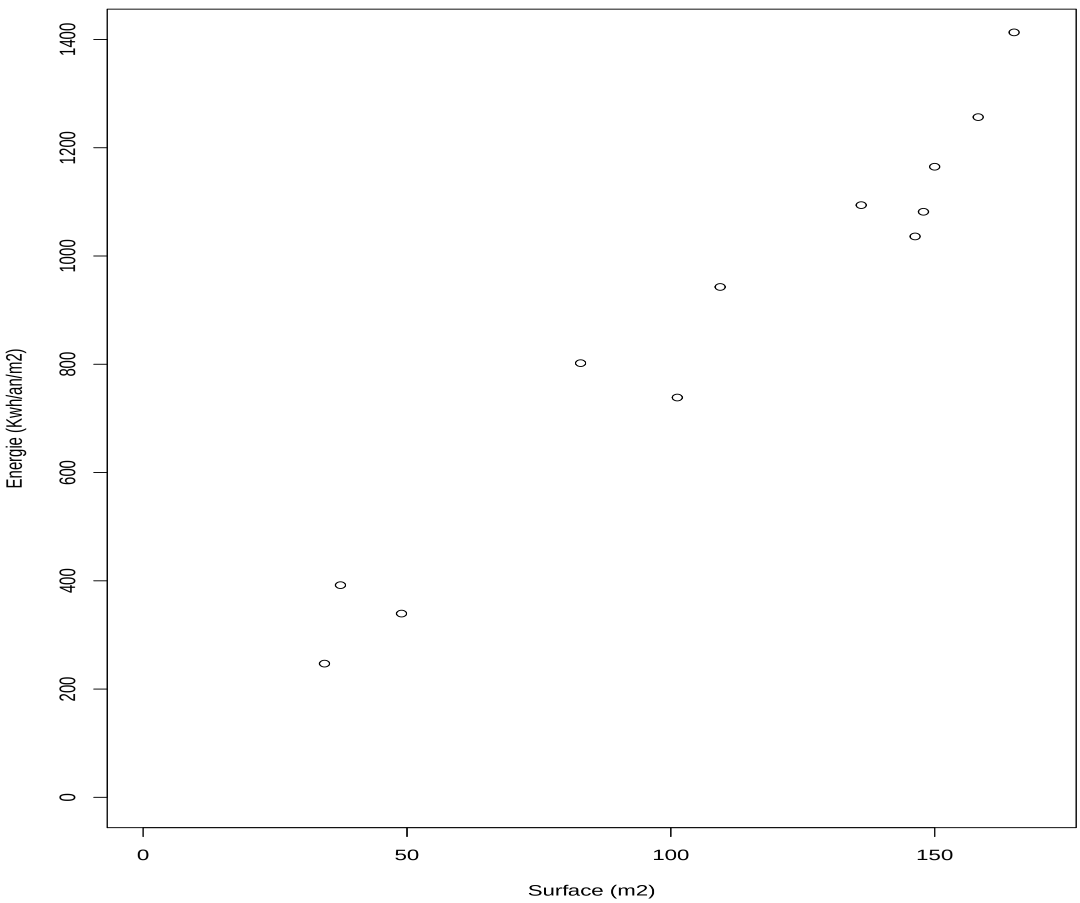

FIGURE 1 – Consommation d'énergie en fonction de la surface pour un échantillon de $n = 12$ appartements appartenants à la catégorie énergétique B.