DÉPARTEMENT DE MATHÉMATIQUES
ET DE GÉNIE INDUSTRIEL

MTH6312 - MÉTHODES STATISTIQUES D'APPRENTISSAGE

Devoir nº 1 - Automne 2024

Date de remise : 27 septembre avant 23h55 (en pdf dans Moodle)

DIRECTIVES :

- ✓ Remettre un rapport en PDF. Inclure dans votre rapport le code R ou Python que vous avez utilisé.
- ✓ Lors de la correction, il sera tenu compte de la clarté des démarches ainsi que la qualité de la présentation de votre rapport.
- ✓ Tout cas de soupçon de fraude sera automatiquement reporté à la Division académique du Comité de discipline étudiante. Une des sanctions minimales possibles est une note **F** à ce cours.

QUESTION Nº 1 (8 points)

On dispose de n observations indépendantes, de la forme $\{(x_i, y_i), i = 1, \ldots, n\}$. On considère ici que $p = 1$ (c-à-d x_i est un scalaire).

Les sous questions suivantes sont indépendantes l'une de l'autre. Veuillez répondre à chacune d'elles en justifiant votre réponse selon le contexte donné.

1.a) (2 points) Déterminer l'estimateur du maximum de vraisemblance (e.v.m) $\hat{\theta}_{evm}$ de θ si étant donné x_i, y_i est distribuée selon la fonction de densité

$$f(y_i|\theta, x_i) = \begin{cases} 2(x_i\theta)^2 y_i^3 \exp\left\{-\theta x_i y_i^2\right\} & \text{si } \; y_i \geq 0, \; x_i > 0 \\ 0 & \text{sinon,} \end{cases}$$

où $\theta > 0$ est un paramètre. Donner le nom et les paramètres de la distribution asymptotique de $\hat{\theta}_{evm}$.

1.b) (2 points) Déterminer l'estimateur du maximum de vraisemblance (e.v.m) $\hat{\theta}_{evm}$ de θ si, étant donné x_i, y_i est distribuée selon une loi de Pascal de paramètre x_i et θ. Précisément la fonction de masse de y_i est

$$p(y_i \mid \theta, x_i) = \begin{cases} \binom{y_i - 1}{x_i - 1} \theta^{x_i} (1-\theta)^{y_i - x_i} & \text{si } \; y_i = x_i, x_i + 1, \ldots \\ 0 & \text{sinon,} \end{cases}$$

où $0 < \theta < 1$ est un paramètre, et x_i est un entier tel que $x_i \geq 1$.
Préciser le nom et les paramètres de la distribution asymptotique de $\hat{\theta}_{evm}$.

1.c) (4 points) En reprenant le contexte de la question **1.b)** ci-dessus, on considère une distribution *a priori* de θ, de fonction de densité

$$\pi(\theta) = \begin{cases} 2(1-\theta) & \text{si } 0 < \theta < 1 \\ 0 & \text{sinon.} \end{cases}$$

1.c1) Déterminer l'estimateur MAP de θ, $\hat{\theta}_{Map}$.

1.c2) Déterminer l'estimateur de Bayes de θ, $\hat{\theta}_{Bayes}$, en considérant une fonction de perte quadratique. Comparer brièvement les deux estimateurs $\hat{\theta}_{Map}$ et $\hat{\theta}_{Bayes}$.

QUESTION N° 2 (12 points)

On dispose de n observations indépendantes, de la forme $\{(\mathbf{x}_i, y_i), i = 1, \dots, n\}$. On considère ici que $p = 1$ (c-à-d x_i est un scalaire). Les coefficients $w_1, \dots, w_n$ sont tels que $w_i \geq 0$ et on suppose que $\sum_{i=1}^n w_i = 1$.

Les sous questions 2.a) et 2.b) suivantes sont indépendantes l'une de l'autre.

2.a) (1 point) Déterminer l'estimateur des moindres carrés $\hat{\beta}$ défini par

$$\hat{\beta} = \text{argmin } RSS_1(\beta), \quad \text{où } RSS_1(\beta) = \sum_{i=1}^n (y_i - \beta x_i)^2.$$

2.b) (2 points) Déterminer les estimateurs des moindres carrés $\hat{\beta}_0$ et $\hat{\beta}_1$ définis par

$$(\hat{\beta}_0, \hat{\beta}_1) = \text{argmin } RSS_2(\beta_0, \beta_1), \text{ où } RSS_2(\beta_0, \beta_1) = \sum_{i=1}^n w_i (y_i - \beta_0 - \beta_1 x_i)^2.$$

On considère à présent les $n = 11$ observations suivantes :

i	1	2	3	4	5	6	7	8	9	10	11
x_i	2,12	2,25	4,23	4,98	4,76	4,22	3,09	2,73	2,44	4,32	3,41
y_i	20,3	27,6	35,4	39,2	32,7	36,5	26,7	28,8	23,9	37,3	25,8
w_i	1/24	1/12	1/48	7/48	1/24	1/12	1/6	5/48	1/12	7/48	1/12

2.c) (1 point) Produire le diagramme de dispersion (ou *Scatter Plot*) pour les 11 observations. Donner ensuite les valeurs numériques des estimateurs obtenus aux sous questions 2.a) et 2.b).

2.d) (2 points) Tracer le graphe de $RSS_1(\beta)$ en fonction de β (en considérant un intervalle approprié pour β). Comparer le point optimal $\hat{\beta}$ indiqué par le graphe avec la solution calculée algébriquement à la sous-question 2.c) et commenter brièvement.

2.e) (4 points) En considérant un intervalle approprié pour β_0 et un intervalle approprié pour β_1, tracer les courbes de niveau de la fonction $RSS_2(\beta_0, \beta_1)$ (voir un exemple à la page 46 de ISLr). Comparer le point optimal $(\hat{\beta}_0, \hat{\beta}_1)$ indiqué par les courbes de niveau avec la solution calculée algébriquement à la sous-question 2.c) et commenter brièvement.

2.f) (2 points) En utilisant une procédure appropriée d'optimisation de votre choix (de `R` ou `Python`), procéder à la minimisation numérique des sommes de carrés $RSS_1(\beta)$, $RSS_2(\beta_0, \beta_1)$ et $RSS_3(\beta_0, \beta_1, \beta_2)$, où $RSS_3(\beta_0, \beta_1, \beta_2) = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i - \beta_2 x_i^2)^2$.
Comparer la solution d'optimisation numérique avec votre solution exacte pour $RSS_1(\beta)$ et $RSS_2(\beta_0, \beta_1)$.