TISD M1 Pro - DM4 (d'après l'examen de 2009)

Exercice 1 (Maximum de vraisemblance et régression logistique (théorique))
On considère un échantillon de couples *i.i.d.*$(Y_i, X_i)_{i\in[\![1,n]\!]}$, où $Y_i \in \{0,1\}$ et $X_i \in \mathbb{R}$ est une variable explicative. On suppose que :

$$Y_i = \mathbb{1}_{Y_i^*>0} \quad \text{avec} \quad Y_i^* = \alpha + \beta X_i + \varepsilon_i, \tag{1}$$

où les $(\varepsilon_i)_{i\in[\![1,n]\!]}$ sont des résidus *i.i.d.*, indépendants des X_i, centrés et à densité. On notera f leur densité (supposée continue) et F leur fonction de répartition. On supposera de plus que leur distribution est symétrique, ce qui se traduit par : f paire ou de façon équivalente $F(u) = 1 - F(-u)$.

1. Quelle est la loi de Y_i conditionnellement à $X_i = x_i$? Montrer que $\mathbb{P}(Y_i = 1 \,|\, X_i = x_i) = F(\alpha + \beta\, x_i)$. En déduire l'espérance et la variance conditionnelles de Y_i.

2. On pose $p_i = F(\alpha + \beta x_i)$. Ecrire la vraisemblance des $Y_1, \cdots Y_n$ conditionnellement aux $X_1, \cdots X_n$.

3. Montrer que :

$$\frac{\partial \ln\, p_i}{\partial \alpha} = \frac{f(\alpha + \beta x_i)}{F(\alpha + \beta x_i)}$$

faire de même pour la dérivée partielle par rapport à β et pour les dérivées partielles de $\ln(1-p_i)$.

4. En déduire que l'estimateur du maximum de vraisemblance $(\widehat{\alpha}, \widehat{\beta})$ satisfait les équations suivantes :

$$\begin{aligned}
&\sum_{i=1}^{n} Y_i \frac{f(\widehat{\alpha} + \widehat{\beta} X_i)}{F(\widehat{\alpha} + \widehat{\beta} X_i)} - (1 - Y_i) \frac{f(\widehat{\alpha} + \widehat{\beta} X_i)}{1 - F(\widehat{\alpha} + \widehat{\beta} X_i)} = 0 \\
&\sum_{i=1}^{n} X_i Y_i \frac{f(\widehat{\alpha} + \widehat{\beta} X_i)}{F(\widehat{\alpha} + \widehat{\beta} X_i)} - X_i (1 - Y_i) \frac{f(\widehat{\alpha} + \widehat{\beta} X_i)}{1 - F(\widehat{\alpha} + \widehat{\beta} X_i)} = 0.
\end{aligned} \tag{2}$$

Dans la suite du problème, on s'intéresse à la régression logistique, obtenue pour des résidus ε_i tels que

$$F(x) = \frac{1}{1 + e^{-x}}. \tag{3}$$

5. Montrer que cette distribution est symétrique.

6. Montrer que F est strictement croissante et en déduire que les paramètres (α, β) sont identifiables, c'est à dire que si l'on connaît $\mathbb{P}(Y_i = 1 \,|\, X_i = x)$ pour tout x, alors α et β sont

déterminés de façon unique.

7. Réécrire dans le cas de la régression logistique les dérivées partielles de la question **3** en fonction de F la fonction de répartition des ε_i.

8. Montrer que dans le cas de la régression logistique, le système (2) se ré-écrit :

$$\begin{aligned}\sum_{i=1}^{n} \big(Y_i - F(\widehat{\alpha} + \widehat{\beta} X_i)\big) &= 0 \\ \sum_{i=1}^{n} \big(Y_i - F(\widehat{\alpha} + \widehat{\beta} X_i)\big) X_i &= 0. \end{aligned} \tag{4}$$

Jusqu'à la fin du problème, on se place dans le cas d'un modèle logistique où X est également une variable dichotomique, c'est-à-dire qui prend les valeurs 0 ou 1. On se donne le tableau de contingence suivant :

		Y	
		0	1
X	0	$n_{00} = 26$	$n_{01} = 26$
	1	$n_{10} = 16$	$n_{11} = 32$

TABLE 1 – Tableau de contingence de X et Y

9. Réécrire les équations (4) avec uniquement n_{00}, n_{01}, n_{10}, n_{11} (sans les X_i ou Y_i).

10. Résoudre les équations obtenues à la question **11** pour obtenir l'estimateur du maximum de vraisemblance (Indication : on pourra poser $a = F(\widehat{\alpha})$ et $b = F(\widehat{\alpha} + \widehat{\beta})$).

Exercice 2 (Pression et température du mercure, R)
Nous analysons des données pour obtenir des relations entre la température en degrés Celsius et la pression en millimètre de mercure du mercure gazeux. La base de données, sous **R**, s'appelle `pressure`. Elle contient deux variables, T (`temperature`) et P (`pressure`).
On s'intéresse à la régression de $\log(P)$ sur $\log(T)$:

$$\log(P) = a \log(T) + b + \varepsilon \tag{5}$$

1. Tracer la courbe reliant le log de la pression en fonction du log de la température.

2. Sur la courbe de la question **1**, on peut voir une légère inflexion au niveau de la 5^{e}valeur. On considère les observations 2 à 5 d'une part, 6 à 19 d'autre part. Faire le test de Chow pour savoir s'il y a deux comportements différents. Ne pas oublier de vérifier les hypothèses du test de Chow.