TIAD - fiche 3

Exercice 1 (Modèle de régression à variables qualitatives)
Soit Y une variable observable prenant les valeurs 1, avec probabilité p et 0, avec probabilité $1-p$. Nous observons un échantillon $(Y_i)_{i\in[\![1,n]\!]}$ de variables aléatoires indépendantes identiquement distribuées de même loi que Y.

1.1. Quelles sont les lois de Y et $\sum_{i=1}^{n} Y_i$?

1.2. Ecrire la vraisemblance des observations. Calculer et étudier l'estimateur du maximum de vraisemblance $\widehat{p}$ de p.

Nous considérons maintenant un couple de variables aléatoire (Y, X) où X est une variable explicative prenant les valeurs 0 ou 1, et où les Y est une variable à expliquer, prenant elle aussi les valeurs 0 et 1, et pour laquelle on suppose qu'il existe une variable latente Y^* telle que

- conditionnellement à X, Y^* suit une loi $\mathcal{N}(aX+b, 1)$ avec a, $b \in \mathbb{R}$,
- $Y = \mathbf{1}_{Y^*>0}$.

2.1. Quelles sont les lois de Y et Y^* conditionnellement à X ? Montrer que $\mathbb{P}(Y=1\,|\,X=1) = \Phi(a+b)$ où Φ, la fonction de répartition de la loi normale $\mathcal{N}(0,1)$. Calculer les probabilités $\mathbb{P}(Y=1\,|\,X=0)$, $\mathbb{P}(Y=0\,|\,X=1)$ et $\mathbb{P}(Y=0\,|\,X=0)$.

Nous disposons d'un échantillon de 100 observations, $(y_i, x_i)_{i\in[\![1,100]\!]}$ réalisations des 100 couples $(Y_i, X_i)_{i\in[\![1,100]\!]}$ indépendants deux à deux et de même loi que (Y, X) étudié à la question 2.1. Ces observations sont résumées dans le Tableau 1.

		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

2.2. Ecrire la vraisemblance des observations $(Y_1, \cdots Y_{100})$ conditionnellement à $(X_1, \cdots X_{100})$ en utilisant les nouvelles variables $\alpha = \Phi(a+b)$ et $\beta = \Phi(b)$ ainsi que les effectifs n_{00}, n_{01}, n_{10} et n_{11} définis dans le Tableau 1.

2.3. Donner la log-vraisemblance. Puis, écrire les conditions du premier ordre du programme de maximisation de la log-vraisemblance par rapport à α et β.

2.4. En déduire les estimateurs du maximum de vraisemblance $\widehat{\alpha}$ et $\widehat{\beta}$ pour α et β. Faire l'application numérique avec les valeurs données dans le Tableau 1.

2.5. Donner la valeur de $\Phi(0)$. En déduire $\Phi^{-1}(1/2)$. Puis, en utilisant que $\Phi^{-1}(2/3) = 0.43$, déduire les valeurs numériques des EMV $\widehat{a}$ et $\widehat{b}$ pour les observations décrites au Tableau 1.

2.6. Les estimateurs $\widehat{\alpha}$ et $\widehat{\beta}$ sont-ils convergents lorsque $n_{00} + n_{01} \to +\infty$ et $n_{10} + n_{11} \to +\infty$? Qu'en déduire pour $\widehat{a}$ et $\widehat{b}$?

2.7. Qu'en est-il de la normalité asymptotique ?