

Contrôle continu du 9 Mai 2023
Durée 1h30-Notes de cours et calculatrices autorisées

1 Logistique

On suppose avoir collecté auprès d'un groupe d'étudiants ayant suivi un module "Régression linéaire et logistique" les variables :

- X_1 nombre d'heures à travailler le module (hors enseignements),
- X_2 moyenne obtenue au premier semestre,
- Y module validé (Y vaut 1 si le module est validé et 0 sinon).

On souhaite expliquer Y en fonction de X_1 et X_2 . Pour cela, on va utiliser un modèle logistique. On définit la fonction

$$\text{logit} : \begin{array}{ll}]0, 1[& \longrightarrow \mathbb{R} \\ u & \longmapsto \ln\left(\frac{u}{1-u}\right) \end{array},$$

et on note, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\pi(x_1, x_2) = \mathbb{P}(Y = 1 \mid X_1 = x_1, X_2 = x_2).$$

On fait l'hypothèse qu'il existe $(\beta_0^*, \beta_1^*, \beta_2^*) \in \mathbb{R}^3$ tels que, pour tout $(x_1, x_2) \in \mathbb{R}^2$,

$$\text{logit}(\pi(x_1, x_2)) = \beta_0^* + \beta_1^* x_1 + \beta_2^* x_2. \quad (1)$$

1. Montrer que l'équation (1) précédente est équivalente à l'équation suivante :

$$\pi(x_1, x_2) = \frac{e^{\beta_0^* + \beta_1^* x_1 + \beta_2^* x_2}}{1 + e^{\beta_0^* + \beta_1^* x_1 + \beta_2^* x_2}}. \quad (2)$$

On s'est servi des données des années précédentes pour estimer les paramètres. On obtient les estimations $\hat{\beta}_0 = -6$, $\hat{\beta}_1 = 0.05$ et $\hat{\beta}_2 = 0.4$.

2. À l'aide de ces paramètres, estimer la probabilité qu'un étudiant ayant obtenu 12 de moyenne générale au premier semestre et travaillé 40 h son module puisse le valider.
3. Montrer que la fonction logit est croissante.
4. On suppose $\beta_1^* > 0$. Soit $\lambda \in \mathbb{R}$ vérifiant $1 < \lambda < 1$. Montrer que

$$\pi(x_1, x_2) \geq \lambda \quad \Longleftrightarrow \quad x_1 \geq \frac{1}{\beta_1^*} (\text{logit}(\lambda) - \beta_0^* - \beta_2^* x_2).$$

5. Combien d'heures l'étudiant de la question 2 aurait-il dû travailler pour espérer obtenir son module avec une probabilité (estimée) supérieure à 0.8 ?

2 Quantification dans le plan et Voronoï

On considère deux points distincts $M_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ et $M_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ du plan \mathcal{P} . On note $\overline{M_1 M_2}$ le milieu du segment $M_1 M_2$. Soit Δ la droite perpendiculaire au segment $M_1 M_2$ qui passe par $\overline{M_1 M_2}$. On rappelle que Δ s'appelle la médiatrice du segment $M_1 M_2$. On appelle D la droite qui passe par $M_1 M_2$.

1. Faire un schéma. Montrer qu'un point $M = \begin{pmatrix} x \\ y \end{pmatrix}$ appartient à D si, et seulement si,

$$\det \begin{pmatrix} x - x_1 & x_2 - x_1 \\ y - y_1 & y_2 - y_1 \end{pmatrix} = 0.$$

En déduire que

$$D = \left\{ M = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{P} : (y_2 - y_1)x + (x_1 - x_2)y = (y_2 - y_1)x_1 + (x_1 - x_2)y_1 \right\}.$$

2. Montrer que le vecteur $\vec{u} := \begin{pmatrix} y_1 - y_2 \\ x_2 - x_1 \end{pmatrix}$ est orthogonal à la droite D . Déterminer, en fonction de x_1, x_2, y_1, y_2 , des réels a, b, c tels que $ax + by = c$ soit l'équation de la droite Δ .
3. On pose

$$\Omega = \left\{ M = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{P} : ax + by < c \right\},$$

et

$$\Omega' = \left\{ M = \begin{pmatrix} x \\ y \end{pmatrix} \in \mathcal{P} : ax + by > c \right\}.$$

Montrer que $\mathcal{P} = \Omega \cup \Omega' \cup \Delta$. Montrer ou admettre que Ω et Ω' sont les cellules de Voronoï ouvertes associées à M_1, M_2 .

4. Soit T le triangle isocèle rectangle à l'origine et de côté 2. C'est-à-dire le triangle défini par les points $O, \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}$. Soit X de loi uniforme sur T . Montrer que X a pour moyenne $\begin{pmatrix} \frac{2}{3} \\ \frac{2}{3} \end{pmatrix}$.
5. On considère le carré \mathcal{C} centré à l'origine et de côté 2. On appelle $\alpha, \beta, \gamma, \delta$ les sommets de \mathcal{C} avec $\alpha = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \beta = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \gamma = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$ et $\delta = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Dessiner \mathcal{C} .
6. On appelle :
- $T1$ le triangle $\gamma\beta\alpha$,
 - $T2$ le triangle $\alpha\delta\gamma$,
 - $T3$ le triangle $\beta\alpha\delta$,
 - $T4$ le triangle $\delta\gamma\beta$.

Soient A, B, C, D les quatre points du plan définis par $A = \begin{pmatrix} \frac{1}{3} \\ -\frac{1}{3} \end{pmatrix}, B = \begin{pmatrix} -\frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, C = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \end{pmatrix}, D = \begin{pmatrix} -\frac{1}{3} \\ -\frac{1}{3} \end{pmatrix}$.

Placer ces quatre points sur le schéma dessiné précédemment.

7. On considère, pour $i = 1, \dots, 4$, le vecteur aléatoire X_i de loi uniforme sur T_i . Quelles sont les moyennes de X_1, X_2, X_3, X_4 ?
8. On rappelle que la condition à vérifier pour pouvoir qualifier une quantification d'optimale est que les centres de masse des cellules de Voronoï soient les points définissant ces cellules (voir le cours). Montrer que (A, B) muni des poids $(\frac{1}{2}, \frac{1}{2})$ constituent une quantification optimale à deux points de la probabilité uniforme sur \mathcal{C} .
9. Montrer que (C, D) muni des poids $(\frac{1}{2}, \frac{1}{2})$ constituent également une quantification optimale à deux points de la probabilité uniforme sur \mathcal{C} .
10. Quelles sont à votre avis toutes les quantifications optimales à deux points de la probabilité uniforme sur \mathcal{C} ?