La théorie autour du modèle logit : Calcul des estimés de β, $\hat{\beta}$, et Inférence.

1. Calcul des estimateurs $\hat{\beta}$ de β.

Les estimateurs $\hat{\beta}$ de β calculés par le maximum de vraisemblance sont solutions du système d'équations :

$$\sum_{i=1}^{n}\left(y_i - \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}\right) x_{i_t} = 0 \quad , \quad \forall t = 0, \dots, k$$

avec

n : nombre d'observations

k : nombre de covariables

$x_{i_0} = \mathbb{1}$ où $\mathbb{1}^T = (\underbrace{1, \dots, 1}_{n \text{ fois}})$.

Preuve : (Juste pour votre culture)

Fonction de vraisemblance des y_i, $i=1, \dots, n$, de de distribution binomiale, i.i.d :

$$L(\underset{\sim}{y}, \beta) = \prod_{i=1}^{n} P_i(\beta)^{y_i}\left(1-P_i(\beta)\right)^{1-y_i} \quad , \quad \underset{\sim}{y} = (y_1, \dots, y_n)$$

log-vraisemblance $\ell(\underset{\sim}{y}, \beta)$:

$$\ell(\underset{\sim}{y}, \beta) = \log\left(L(\underset{\sim}{y}, \beta)\right) = \sum_{i=1}^{n}\left\{ y_i \ln(P_i(\beta)) + (1-y_i)\ln(1-P_i(\beta))\right\}$$

$$= \sum_{i=1}^{n}\left\{ y_i \ln\left(\frac{P_i(\beta)}{1-P_i(\beta)}\right) + \ln\left(1-P_i(\beta)\right)\right\}$$

$$\frac{\partial \ell(\underline{y},\beta)}{\partial \beta_t} = \frac{\partial \ell(\underline{y},\beta)}{\partial p_i(\beta)} * \frac{\partial p_i(\beta)}{\partial \beta_t} \quad \text{(règle de la chaîne)}$$

$$= \sum_{i=1}^{n} \left(\frac{y_i - p_i(\beta)}{p_i(\beta)(1-p_i(\beta))} \right) \left(\frac{\partial p_i(\beta)}{\partial \beta_t} \right)$$

Modèle logit : $p_i(\beta) = e^{x_i^T\beta} / 1+e^{x_i^T\beta}$

$$\frac{\partial p_i(\beta)}{\partial \beta_t} = \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}} x_{i_t} = p_i(\beta)\left(1-p_i(\beta)\right) x_{i_t}$$

$$\frac{\partial \ell(\underline{y},\beta)}{\partial \beta_t} = 0 \iff \sum_{i=1}^{n} \frac{y_i - p_i(\beta)}{p_i(\beta)(1-p_i(\beta))} \cdot p_i(\beta)\left(1-p_i(\beta)\right) x_{i_t} = 0$$

$$\iff \sum_{i=1}^{n} \left(y_i - p_i(\beta)\right) x_{i_t} = 0 \qquad \forall i = 1, \ldots, k.$$

2. Inférence

2.1. Information de Fischer

L'information à l'aide de la même stratégie est :

$$I_{s,t}(\beta) = -E\left[\frac{\partial^2 \ell(\underline{y},\beta)}{\partial \beta_t \partial \beta_s}\right] = \sum_{i=1}^{n} p_i(\beta)\left(1-p_i(\beta)\right) x_{i_t} x_{i_s}$$

$$\forall s,t \in \{0, \ldots, k\}$$

2.2. Matrice de Variance - Covariance : Σ

Rappel : $\hat{\Sigma}(\beta) = I_{A,t}^{-1}(\beta)$

i.e. La matrice de var-cov. est l'inverse de l'information de Fisher.

2.3. Tests statistiques :

On cherche souvent à déterminer la significativité d'un ou de plusieurs paramètres dans un modèle i.e tester $H_0 : \beta_j = 0$ ou $H_0 : \beta_j = \dots = \beta_3 = 0$.

Généralement, on se base sur la normalité asymptotique des $\hat{\beta}_{EMV}$. On parle alors de test de Wald. La statistique d'un tel test est

$$W = \hat{\beta}_j^2 / \hat{\sigma}_{\hat{\beta}_j}^2 \quad \text{et} \quad \text{signe}(\hat{\beta}_j)\sqrt{W} \sim \mathcal{N}(0,1)$$

Mais en pratique, on utilise des tests t ou tests de student.

Exp : Test : $H_0 : \beta_j = 0$ vs $H_a : \beta_j \neq 0$.

$$t_{obs} = \hat{\beta}_j / \hat{\sigma}_{\hat{\beta}_j}$$

p-value : $p = \mathbb{P}\left[t_{n-(k+1)} \geq |t_{obs}|\right]$

$\begin{cases} n : \text{nombre d'observations} \\ k+1 : (\beta_0, \dots, \beta_k) : \text{nombre de paramètres} \end{cases}$

- Test du rapport de vraisemblance.

Il s'agit de comparer 2 modèles emboîtés i.e dont l'un est un cas particulier de l'autre. La statistique du test est

$$LR = -2\ln\left(\frac{L_{Simple}}{L_{Complet}}\right) \underset{H_0}{\sim} \chi^2_k$$

k = Différence entre le nombre de paramètres des 2 modèles.

- Qualité d'ajustement du modèle

→ $AIC = -2\ln(L) + 2d$

→ $BIC = -2\ln(L) + d\ln n$

→ Déviance

(Dépendamment du cours du prof.)

Exemple analytique (Estimation, Matrice de variance covariance, Test simple

Base de données

i	y_i	x_i
1	4	1
2	1	0
3	0	0
4	2	1
5	0	1

1) Calculer $\hat{\beta}$ du modèle de regression logistique associé à ces données

2) La variable x est-elle significative pour le modèle?

Solution

1. $\hat{\beta}$ est solution du système d'équation :

$$\sum_{i=1}^{5} (y_i - P_i(\beta)) x_{i_t} = 0 \quad , \quad t = 0, 1.$$

Pour $t=0$, $x_{i_0} = 1$ et on a :

$i=1$, $y_1 - P_1(\beta) = 4 - \dfrac{e^{\beta_0+\beta_1(1)}}{1+e^{\beta_0+\beta_1(1)}} = 4 - e^{\beta_0+\beta_1} / 1+e^{\beta_0+\beta_1}$

$i=2$, $y_2 - P_2(\beta) = 1 - e^{\beta_0+\beta_1(0)} / 1+e^{\beta_0+\beta_1(0)} = 1 - e^{\beta_0} / 1+e^{\beta_0}$

$i=3$, $y_3 - P_3(\beta) = 0 - \dfrac{e^{\beta_0+\beta_1(0)}}{1+e^{\beta_0+\beta_1(0)}} = - e^{\beta_0} / 1+e^{\beta_0}$

$i=4$, $y_4 - P_4(\beta) = 2 - \dfrac{e^{\beta_0+\beta_1(1)}}{1+e^{\beta_0+\beta_1(1)}} = 2 - e^{\beta_0+\beta_1} / 1+e^{\beta_0+\beta_1}$

$i=5$, $y_5 - P_5(\beta) = 0 - \dfrac{e^{\beta_0+\beta_1(1)}}{1+e^{\beta_0+\beta_1(1)}} = - e^{\beta_0+\beta_1} / 1+e^{\beta_0+\beta_1}$

$$\sum_{i=1}^{5} (y_i - P_i(\beta)) x_{i_0} = 0 \Leftrightarrow 7 - 3 \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} - 2 \frac{e^{\beta_0}}{1+e^{\beta_0}} = 0 \quad (1)$$

Pour $t=1$, x_{i_1} correspond au vecteur x_i du tableau.

$i=1$, $(y_1 - P_1(\beta))(1) = 4 - e^{\beta_0+\beta_1} / 1+e^{\beta_0+\beta_1}$

$i=2$, $(y_2 - P_2(\beta))(0) = 0$

$i=3$, $(y_3 - P_3(\beta))(0) = 0$

$i=4$, $(y_4 - P_4(\beta))(1) = 2 - e^{\beta_0+\beta_1} / 1+e^{\beta_0+\beta_1}$

$i=5$, $(y_5 - P_5(\beta))(1) = - e^{\beta_0+\beta_1} / 1+e^{\beta_0+\beta_1}$.

$$\sum_{i=1}^{5} (y_i - P_i(\beta)) x_{i_1} = 0 \Leftrightarrow 6 - 3 \frac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} = 0 \quad (2)$$

$(2) \Rightarrow \dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}} = 2$ dans $(1) \Rightarrow \dfrac{e^{\beta_0}}{1+e^{\beta_0}} = \dfrac{1}{2}$

D'où $\hat{\beta}_0 = 0$

Remplacer dans (1) ou (2) et obtenir $\hat{\beta}_1 \approx \ln 2$

Donc $\boxed{\hat{\beta}_0 = 0 \quad \text{et} \quad \hat{\beta}_1 = \ln 2}$

2] Significativité de la variable x

Test : $H_0 : \beta_1 = 0$ vs $H_a : \beta_1 \neq 0$.

(i) Calculons la matrice de variance-covariance.

→ Information de Fisher

$$I_{s,t}(\hat{\beta}) = \sum_{i=1}^{5} p_i(\hat{\beta})(1 - p_i(\hat{\beta})) x_{i_t} x_{i_s}, \quad s,t \in \{0,1\}.$$

Fixer $s = 0$.

• $t = 0$; $x_{i_t} = x_{i_s} = x_{i_0} = 1$

$$I_{0,0}(\hat{\beta}) = \sum_{i=1}^{5} p_i(\hat{\beta})(1 - p_i(\hat{\beta}))$$

$$= 3 \frac{e^{\hat{\beta}_0 + \hat{\beta}_1}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1}} \left(1 - \frac{e^{\hat{\beta}_0 + \hat{\beta}_1}}{1 + e^{\hat{\beta}_0 + \hat{\beta}_1}}\right) + 2 \frac{e^{\hat{\beta}_0}}{1 + e^{\hat{\beta}_0}} \left(1 - \frac{e^{\hat{\beta}_0}}{1 + e^{\hat{\beta}_0}}\right)$$

$$\left(\text{car } p_i(\hat{\beta}) = e^{\hat{\beta}_0 + x_i \hat{\beta}_1} / 1 + e^{\hat{\beta}_0 + x_i \hat{\beta}_1}\right)$$

$$= 3 \frac{e^{\ln 2}}{1 + e^{\ln 2}} \left(1 - \frac{e^{\ln 2}}{1 + e^{\ln 2}}\right) + 2 \frac{e^0}{1 + e^0} \left(1 - \frac{e^0}{1 + e^0}\right)$$

$$= 3 \cdot \frac{2}{1+2} \left(1 - \frac{2}{1+2}\right) + 2 \cdot \frac{1}{2}\left(1 - \frac{1}{2}\right)$$

$$= \frac{2}{3} + \frac{1}{2} = \frac{7}{6}$$

Procéder de même pour obtenir

$$I(\hat{\beta}) = \begin{bmatrix} 1.17 & 0.67 \\ 0.67 & 0.67 \end{bmatrix}$$

D'où

$$\text{Var}[\hat{\beta}] = I^{-1}(\hat{\beta}) = \begin{bmatrix} 2 & -2 \\ -2 & 3.5 \end{bmatrix}$$

La statistique du test est donc

$$T_{\hat{\beta}_1} = \hat{\beta}_1 / \sqrt{Var[\hat{\beta}_1]} = \ln 2 / \sqrt{3.5} \approx 0.371$$

$T_3(0.95) \in (2.353, 3.182)$

$T_{\hat{\beta}_1} < T_3(0.95) \Rightarrow$ On ne rejette pas H_0 et la variable x n'est pas significative pour le modèle.

7