HAX912X – 2023/2024
TD2 – Correction

1

Exercice 1

$y|x \sim \mathcal{N}(\exp(\beta_0+\beta_1 x), \sigma^2)$

1) $f(y;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(y-\mu)^2\right) \nu(dy)$

$\Leftrightarrow f(y;\mu,\sigma^2) = \exp\left(\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log(\sigma)\right) \frac{\nu(dy)}{\sqrt{2\pi}}$

La famille gaussienne appartient à la famille exponentielle avec un paramètre de nuisance :

$\theta = \mu, \; \phi = \sigma^2$

$b(\theta) = \frac{\mu^2}{2} = \frac{\theta^2}{2}$

$c(y,\phi) = -\frac{y^2}{2\sigma^2} - \log(\sigma)$

$r(dy) = \frac{\nu(dy)}{\sqrt{2\pi}}$

Par ailleurs,

$$\mathbb{E}(y|x) = \exp(\beta_0+\beta_1 x)$$

$$\Rightarrow \log(\mathbb{E}(y|x)) = \beta_0+\beta_1 x$$

Modèle linéaire généralisé avec pour fonction de lien $\log(\cdot)$ $(g(x) = \log(x))$

Deux regresseurs : constante + variable x

2] $\theta = \mu = \exp(\beta_0+\beta_1 x) \neq \beta_1+\beta_2 x$

ce n'est pas le lien canonique qui a été utilisé.

3] $$V(\beta_0,\beta_1) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\sum_{i=1}^{n}\left(y_i - e^{\beta_0+\beta_1 x_i}\right)^2\right)$$

$$L(\beta_0,\beta_1) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{n}\left(y_i - e^{\beta_0+\beta_1 x_i}\right)^2$$

$$\frac{\partial L}{\partial \beta_0}(\beta_0,\beta_1) = \sum_{i=1}^{n} e^{\beta_0+\beta_1 x_i}\left(y_i - e^{\beta_0+\beta_1 x_i}\right)$$

$$\frac{\partial L}{\partial \beta_1}(\beta_0,\beta_1) = \sum_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i}\left(y_i - e^{\beta_0+\beta_1 x_i}\right)$$

Équations de vraisemblance :

$$\sum_{i=1}^{n} e^{2\beta_0+2\beta_1 x_i} = \sum_{i=1}^{n} y_i e^{\beta_0+\beta_1 x_i}$$

$$\sum_{i=1}^{n} x_i e^{2\beta_0+2\beta_1 x_i} = \sum_{i=1}^{n} y_i x_i e^{\beta_0+\beta_1 x_i}$$

Impossible de résoudre ce système de 2 équations à 2 inconnues explicitement.

Exercice 2

$$y \mid x \sim \mathcal{P}\left(e^{\beta_0+\beta_1 x}\right)$$

1) $f(y;\mu) = e^{-\mu} \dfrac{\mu^{y}}{y!} \mathbb{1}_{\mathbb{N}}(y)\, \delta(dy)$

$f(y;\mu) = \exp\left(y \log(\mu) - \mu\right) \dfrac{\mathbb{1}_{\mathbb{N}}(y)}{y!} \delta(dy)$

La famille poissonienne appartient à la famille exponentielle scalaire

$\theta = \log(\mu)$ $\quad b(\theta) = \mu = \exp(\theta)$

$\nu(dy) = \dfrac{\mathbb{1}_{\mathbb{N}}(y)}{y!} \delta(dy)$

Par ailleurs,

$$\mathbb{E}(y|x) = \exp(\beta_0 + \beta_1 x)$$

$$\Rightarrow \log(\mathbb{E}(y|x)) = \beta_0 + \beta_1 x$$

Modèle linéaire généralisé avec une fonction de lien $g(x) = \log(x)$

$$\theta = \log(\mu) = \beta_0 + \beta_1 x$$

C'est le lien canonique qui a été utilisé.

2) $$V(\beta_0, \beta_1) = \prod_{i=1}^{n} \frac{e^{-e^{\beta_0 + \beta_1 x_i}}}{y_i!} e^{y_i(\beta_0 + \beta_1 x_i)}$$

$$L(\beta_0, \beta_1) = -\sum_{i=1}^{n} \log(y_i!) - \sum_{i=1}^{n} \left[e^{\beta_0 + \beta_1 x_i} + y_i(\beta_0 + \beta_1 x_i) \right]$$

$$\frac{dL}{d\beta_0}(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} e^{\beta_0 + \beta_1 x_i}$$

$$\frac{dL}{d\beta_1}(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i e^{\beta_0 + \beta_1 x_i}$$

Équations de vraisemblance :

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} e^{\beta_0+\beta_1 x_i}$$

$$\sum_{i=1}^{n} y_i x_i = \sum_{i=1}^{n} x_i e^{\beta_0+\beta_1 x_i}$$

Impossible de résoudre ce système de 2 équations à 2 inconnues explicitement.

Exercice 3

$$y|x \sim NB(h, \pi)$$

1) h fixé $\quad \pi = \left[\dfrac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}\right]$

$$f(y;\pi) = C_{y-1}^{h-1}\, \pi^{h} (1-\pi)^{y-h}\, \mathbb{1}_{\{h,\dots\}}(y)\, \delta(dy)$$

$$f(y;\pi) = \exp\left(y\log(1-\pi) + h\log(\pi) - h\log(1-\pi)\right)\nu(dy)$$

La famille negative binomiale appartient à la famille exponentielle scalaire

$$\theta = \log(1-\pi) \Rightarrow \pi = 1-e^{\theta}$$

$$b(\theta) = -h\log\left(\frac{\pi}{1-\pi}\right) = h\log\left(\frac{1-e^{\theta}}{e^{\theta}}\right) = -h\left(\log(1-e^{\theta})-\theta\right)$$

$$\nu(dy) = C_{y-1}^{h-1}\, \mathbb{1}_{\{h,\dots\}}(y)\, \delta(dy)$$

Par ailleurs, (6)

$$\mathbb{E}(y|x) = b'(\theta) = \frac{h e^{\theta}}{1-e^{\theta}} + h$$

$$\mathbb{E}(y|x) = \frac{h}{1-e^{\theta}} = \frac{h}{\pi} = h\left[\frac{1+e^{\beta_0+\beta_1 x}}{e^{\beta_0+\beta_1 x}}\right]$$

$$-\log\left[\frac{\mathbb{E}(y|x)}{h} - 1\right] = \beta_0 + \beta_1 x$$

Modèle linéaire généralisé avec pour fonction de lien $g(x) = -\log\left[\frac{\mathbb{E}(y|x)}{h} - 1\right]$

$$\theta = \log(1-\pi) = \log\left(\frac{1}{1+e^{\beta_0+\beta_1 x}}\right) \neq \beta_0 + \beta_1 x$$

Ce n'est pas le lien canonique qui a été utilisé.

2] $$V(\beta_0,\beta_1) = \prod_{i=1}^{n} C_{y_i-1}^{h-1} \left[\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right]^{h} \frac{1}{(1+e^{\beta_0+\beta_1 x_i})^{y_i-h}}$$

$$L(\beta_0,\beta_1) = \sum_{i=1}^{n} \log\left(C_{y_i-1}^{h-1}\right) + h\sum_{i=1}^{n}(\beta_0+\beta_1 x_i) - \sum_{i=1}^{n} y_i \log\left(1+e^{\beta_0+\beta_1 x_i}\right)$$

$$\frac{dL}{d\beta_0}(\beta_0,\beta_1) = nh - \sum_{i=1}^{n} y_i \left[\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right]$$

$$\frac{dL}{d\beta_1}(\beta_0,\beta_1) = h\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i x_i \left[\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right]$$

Équations de vraisemblance :

$$nh = \sum_{i=1}^{n} y_i \left[\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right]$$

$$h\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i x_i \left[\frac{e^{\beta_0+\beta_1 x_i}}{1+e^{\beta_0+\beta_1 x_i}}\right]$$

Impossible de résoudre ce système de 2 équations à 2 inconnues explicitement

3] Le lien canonique est tel que

$$\theta = \log(1-\pi) = \beta_0+\beta_1 x$$

$$\Leftrightarrow \pi = 1 - e^{\beta_0+\beta_1 x}$$

$$V(\beta_1,\beta_1) = \prod_{i=1}^{n} C_{y_i-1}^{h-1}\left(1-e^{\beta_0+\beta_1 x_i}\right)^{h} e^{(y_i-h)(\beta_0+\beta_1 x_i)}$$

$$L(\beta_1,\beta_1) = \sum_{i=1}^{n} \log\left(C_{y_i-1}^{h-1}\right) + h\sum_{i=1}^{n} \log\left(1-e^{\beta_0+\beta_1 x_i}\right) + \sum_{i=1}^{n} (y_i-h)(\beta_0+\beta_1 x_i)$$

$$\frac{dL}{d\beta_0}(\beta_0,\beta_1) = h\sum_{i=1}^{n} \frac{-e^{\beta_0+\beta_1 x_i}}{1-e^{\beta_0+\beta_1 x_i}} + \sum_{i=1}^{n} y_i - nh$$

$$\frac{dL}{d\beta_1}(\beta_0,\beta_1) = h\sum_{i=1}^{n} \frac{-x_i e^{\beta_0+\beta_1 x_i}}{1-e^{\beta_0+\beta_1 x_i}} + \sum_{i=1}^{n} y_i x_i - h\sum_{i=1}^{n} x_i$$

Équations de vraisemblance :

$$\sum_{i=1}^{n} y_i - nh = h\sum_{i=1}^{n} \frac{e^{\beta_0+\beta_1 x_i}}{1-e^{\beta_0+\beta_1 x_i}}$$

$$\sum_{i=1}^{n} y_i x_i - h\sum_{i=1}^{n} x_i = h\sum_{i=1}^{n} \frac{x_i e^{\beta_0+\beta_1 x_i}}{1-e^{\beta_0+\beta_1 x_i}}$$

Pas de solution explicite.

Exercice 4

$$\mathbb{P}(y=1|x) = \mathbb{P}(y^* \leq 0|x) = \mathbb{P}(N(0,1) \leq -\beta_0-\beta_1 x|x)$$

$$\Leftrightarrow \mathbb{P}(y=1|x) = \Phi(-\beta_0-\beta_1 x) = 1-\Phi(\beta_0+\beta_1 x)$$

$$y|x \sim \mathcal{B}(1-\Phi(\beta_0+\beta_1 x))$$

1) $f(y;\pi) = \pi^{y}(1-\pi)^{1-y}\,\mathbb{1}_{\{0,1\}}(y)\,\delta(dy)$

$$f(y;\pi) = \exp\left(y\log\left(\frac{\pi}{1-\pi}\right) + \log(1-\pi)\right)\nu(dy)$$

La famille de Bernoulli appartient à la famille exponentielle scalaire avec : $\theta = \log\left(\frac{\pi}{1-\pi}\right) \Longrightarrow \pi = \frac{e^\theta}{1+e^\theta}$

$b(\theta) = -\log(1-\pi) = \log(1+e^\theta)$

$\nu(dy) = \mathbb{1}_{\{0,1\}}(y)\,\delta(dy)$

Par ailleurs,

$$\mathbb{E}(y|x) = 1 - \Phi(\beta_0+\beta_1 x)$$

$$\Longleftrightarrow \Phi^{-1}(1-\mathbb{E}(y|x)) = \beta_0+\beta_1 x$$

Modèle linéaire généralisé avec pour fonction de lien $g(x) = \Phi^{-1}(1-x)$

Modèle probit

2) $\theta = \log\left(\frac{\pi}{1-\pi}\right) = \log\left(\frac{1-\Phi(\beta_0+\beta_1 x)}{\Phi(\beta_0+\beta_1 x)}\right) \neq \beta_0+\beta_1 x$

Ce n'est pas le lien canonique qui a été utilisé.

3) $\beta_0 = \Phi(\beta_0)$ et $\alpha_1 = \Phi(\beta_0+\beta_1)$

$$V(\beta_0,\beta_1) = \prod_{i=1}^{n} \left(1-\Phi(\beta_0+\beta_1 x_i)\right)^{y_i} \left(\Phi(\beta_0+\beta_1 x_i)\right)^{1-y_i}$$

$$L(\beta_0,\beta_1) = \sum_{i=1}^{n} y_i \log\left(1-\Phi(\beta_0+\beta_1 x_i)\right) + \sum_{i=1}^{n} (1-y_i)\log\left(\Phi(\beta_0+\beta_1 x_i)\right)$$

(10)

$$L(\beta_0,\beta_1) = 30\log(1-\Phi(\beta_0)) + 40\log(1-\Phi(\beta_0+\beta_1)) + 10\log(\Phi(\beta_0)) + 20\log(\Phi(\beta_0+\beta_1))$$

$$L(\alpha_0,\alpha_1) = 30\log(1-\alpha_0) + 40\log(1-\alpha_1) + 10\log(\alpha_0) + 20\log(\alpha_1)$$

$$\frac{dL}{d\alpha_0}(\alpha_0,\alpha_1) = \frac{-30}{1-\alpha_0} + \frac{10}{\alpha_0}$$

$$\frac{dL}{d\alpha_1}(\alpha_0,\alpha_1) = \frac{-40}{1-\alpha_1} + \frac{20}{\alpha_1}$$

$$\begin{cases} -\dfrac{30}{1-\alpha_0} + \dfrac{10}{\alpha_0} = 0 \\ \dfrac{-40}{1-\alpha_1} + \dfrac{20}{\alpha_1} = 0 \end{cases} \Leftrightarrow \begin{cases} -30\alpha_0 + 10(1-\alpha_0) = 0 \\ -40\alpha_1 + 20(1-\alpha_1) = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} \alpha_0 = 1/4 \\ \alpha_1 = 2/3 \end{cases}$$