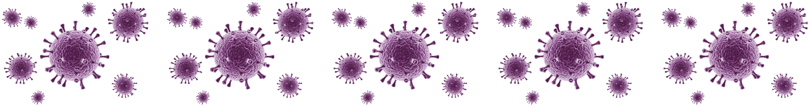
Modèles Linéaires Appliqués / Régression
Famille Exponentielle & Régression

Arthur Charpentier

UQAM

Hiver 2020 - COVID-19 # 13



Régression GLM

$$f(y_i|\theta_i,\varphi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)} + c(y_i,\varphi)\right\}$$

La log-vraisemblance est alors

$$\log\mathcal{L}(\boldsymbol{\theta},\varphi|\boldsymbol{y}) = \frac{\sum_{i=1}^n y_i\theta_i - \sum_{i=1}^n b(\theta_i)}{a(\varphi)} + \sum_{i=1}^n c(y_i,\varphi).$$

Paramètre naturel pour y_i : θ_i
Prédiction pour y_i : $\mu_i = \mathbb{E}(Y_i) = b'(\theta_i)$
Score associé à y_i : $\eta_i = \boldsymbol{x}_i^\top\beta$

Fonction de lien : g telle que $\eta_i = g(\mu_i) = g(b'(\theta_i))$
(g doit être bijective, et croissante...)

Fonctions Liens : Canonique

Cas particulier : lien canonique $g = b'^{-1}$ i.e. $\eta_i = \theta_i$

$$\mu(\boldsymbol{x}) = b'(\boldsymbol{x}^\top\beta)$$

Exemple Pour la loi normale, $\mu(\boldsymbol{x}) = \boldsymbol{x}^\top\beta$ (identité)

Exemple Pour la loi de Poisson, $\mu(\boldsymbol{x}) = \exp(\boldsymbol{x}^\top\beta)$ (logarithmique)

Exemple Pour la loi de Bernoulli, $\mu(\boldsymbol{x}) = \dfrac{\exp\left(\boldsymbol{x}^\top\beta\right)}{1+\exp\left(\boldsymbol{x}^\top\beta\right)}$ (logit)

Exemple Pour la loi Tweedie $V(\mu) = \mu^\gamma$, $\boldsymbol{x}^\top\beta = \dfrac{\mu(\boldsymbol{x})^{1-\gamma}}{1-\gamma}$ (power)

Fonctions Liens : Identité

$$\mu(\boldsymbol{x}) = \boldsymbol{x}^\top \beta$$

Soit $\boldsymbol{x}_j = (x_1, \cdots, x_j, \cdots, x_k)$ et $\boldsymbol{x}_{j+1} = (x_1, \cdots, x_j + 1, \cdots, x_k)$

$$\beta_j = \mu(\boldsymbol{x}_{j+1}) - \mu(\boldsymbol{x}_j)$$

ou

$$\mu(\boldsymbol{x}_{j+1}) = \mu(\boldsymbol{x}_j) + \beta_j$$

Si x_j est une variable indicatrice

$$\beta_j = \left(\frac{1}{n_1} \sum_{i:x_{j,i}=1} y_i\right) - \left(\frac{1}{n_0} \sum_{i:x_{j,i}=1} y_i\right)$$

Fonctions Liens : Logarithmique

$$\mu(\boldsymbol{x}) = \exp\left(\boldsymbol{x}^\top \boldsymbol{\beta}\right)$$

Soit $\boldsymbol{x}_j = (x_1, \cdots, x_j, \cdots, x_k)$ et $\boldsymbol{x}_{j+1} = (x_1, \cdots, x_j + 1, \cdots, x_k)$

$$\beta_j = \log \frac{\mu(\boldsymbol{x}_{j+1})}{\mu(\boldsymbol{x}_j)}$$

ou

$$\mu(\boldsymbol{x}_{j+1}) = \mu(\boldsymbol{x}_j) \cdot e^{\beta_j}$$

Si x_j est une variable indicatrice

$$\beta_j = \log\left(\frac{1}{n_1} \sum_{i:x_{j,i}=1} y_i\right) - \log\left(\frac{1}{n_0} \sum_{i:x_{j,i}=1} y_i\right)$$

Fonctions Liens : Logit & Probit

Gaddum (1933), Bliss (1935), ou Berkson (1944)

$$\mu(\boldsymbol{x}) = \frac{\exp\left(\boldsymbol{x}^\top\boldsymbol{\beta}\right)}{1+\exp\left(\boldsymbol{x}^\top\boldsymbol{\beta}\right)}$$

ou

$$\text{cote}(\boldsymbol{x}) = \frac{\mu(\boldsymbol{x})}{1-\mu(\boldsymbol{x})} = e^{\boldsymbol{x}^\top\boldsymbol{\beta}}$$

$$\text{cote}(\boldsymbol{x}_{j+1}) = \text{cote}(\boldsymbol{x}_j)\cdot e^{\beta_j}$$

i.e. e^{β_j} est un rapport de cote, *odds-ratio*.
Pour le modèle Probit, Bliss(1934), pour '*probability unit*'

$$\mu(\boldsymbol{x}) = \Phi(\boldsymbol{x}^\top\boldsymbol{\beta})$$

où Φ est la fonction de répartition de la loi $\mathcal{N}(0,1)$.

Fonctions Liens : cloglog

Supposons que $N|\boldsymbol{X}=\boldsymbol{x}\sim\mathcal{P}(\exp[\boldsymbol{x}^\top\beta])$, alors

$$\mathbb{P}[N=0|\boldsymbol{X}=\boldsymbol{x}]=\exp\big(-\exp[\boldsymbol{x}^\top\beta]\big)$$

$$\mathbb{P}[N>0|\boldsymbol{X}=\boldsymbol{x}]=1-\exp\big(-\exp[\boldsymbol{x}^\top\beta]\big)$$

Si $Y=\mathbf{1}_{\boldsymbol{N}_\star}(N)$,

$$Y\sim\mathcal{B}(p_{\boldsymbol{x}})\text{ où }p_{\boldsymbol{x}}=1-\exp\big(-\exp[\boldsymbol{x}^\top\beta]\big)$$

correspondant au lien cloglog,

$$\mu(\boldsymbol{x})=1-\exp\big(-\exp[\boldsymbol{x}^\top\beta]\big)$$

Modèles Latents

On suppose que $y_i = \mathbf{1}(\tilde{y}_i > 0)$ où $\tilde{y}_i$ est une variable non-observée telle que

$$\tilde{y}_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} - \varepsilon_i$$

Alors

$$\mu_i = \mathbb{P}[Y_i = 1] = \mathbb{P}[\tilde{Y}_i > s] = \mathbb{P}[\boldsymbol{x}_i^\top \boldsymbol{\beta} - \varepsilon_i > 0] = F(\boldsymbol{x}_i^\top \boldsymbol{\beta}) = F(\eta_i)$$

- si $F(x) = \dfrac{e^x}{1+e^x}$ - loi logistique : modèle logistique
- si $F(x) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$ - loi $\mathcal{N}(0,1)$: probit
- si loi de Gumbel : modèle cloglog

Fonctions Liens : puissance

Plus généralement, on peut envisager un lien puissance

$$\boldsymbol{x}^\top\beta = \begin{cases} \mu(\boldsymbol{x})^\lambda & \text{si } \lambda \neq 0 \\ \log \mu(\boldsymbol{x}) & \text{si } \lambda = 0 \end{cases}$$

Sous R, la synthaxe est

```
> glm(Y~X1+X2,family=poisson(link=power(lambda)))
```

Régression GLM

$$f(y_i|\theta_i,\varphi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)} + c(y_i,\varphi)\right\}$$

La log-vraisemblance est alors

$$\log\mathcal{L}(\boldsymbol{\theta},\varphi|\boldsymbol{y}) = \frac{\sum_{i=1}^n y_i\theta_i - \sum_{i=1}^n b(\theta_i)}{a(\varphi)} + \sum_{i=1}^n c(y_i,\varphi).$$

On peut remplacer $a(\varphi)$ par φ/w_i, w_i est un poids individuel.

Hypothèses

- g est suffisement régulière (au moins $\mathcal{C}^2$
- φ est supposé connu (pour l'instant)
- $n > p = k+1$ et $\boldsymbol{X}$ est de plein rang
  (et donc $\boldsymbol{X}^\top\boldsymbol{X}$ est définie positive)

$$\log \mathcal{L}(\boldsymbol{\theta}, \varphi, \boldsymbol{y}) = \sum_{i=1}^{n} \Big[\underbrace{\frac{y_i\theta_i - b(\theta_i)}{\varphi} + c(y_i, \varphi)}_{\log \mathcal{L}_i} \Big].$$

$$\frac{\partial \log \mathcal{L}_i}{\partial \beta_j} = \frac{\partial \log \mathcal{L}_i}{\partial \theta_i} \cdot \frac{\partial \theta_i}{\partial \mu_i} \cdot \frac{\partial \mu_i}{\partial \eta_i} \cdot \frac{\partial \eta_i}{\partial \beta_j}$$

$$\frac{\partial \log \mathcal{L}_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{\varphi} = \frac{y_i - \mu_i}{\varphi}$$

$$\frac{\partial \theta_i}{\partial \mu_i} = \left(\frac{\partial \mu_i}{\partial \theta_i}\right)^{-1} = \frac{1}{b''(\theta_i)} = \frac{1}{V(\mu_i)}$$

$$\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \boldsymbol{x}_i^\top \boldsymbol{\beta}}{\partial \beta_j} = x_{i,j}$$

Les conditions du premier ordre s'écrivent

$$\sum_{i=1}^{n} \frac{\partial \log \mathcal{L}_i}{\partial \beta_j} = \sum_i \frac{\partial \mu_i}{\partial \eta_i} \times \frac{y_i - \mu_i}{\mathrm{V}(\mu_i)} x_{ij} = 0, \ \forall j$$

Pour simplifier, notons que

$$\frac{\partial \mu_i}{\partial \eta_i} = \left(\frac{\partial \eta_i}{\partial \mu_i}\right)^{-1} = (g'(\mu_i))^{-1}$$

Les conditions du premier ordre s'écrivent

$$\boldsymbol{X}^\top \boldsymbol{\Omega}(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}$$

$$\boldsymbol{\Omega} = \mathrm{diag}\big((V(\mu_i)g'(\mu_i))^{-1}\big)$$

mais on utilisera une autre écriture...

Régression GLM

Les conditions du premier ordre s'écrivent

$$\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{\Delta}(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}$$

$$\boldsymbol{W} = \text{diag}((V(\mu_i)g'(\mu_i)^2)^{-1}) \text{ et } \boldsymbol{\Delta} = \text{diag}(g'(\mu_i))$$

Remarque: la matrice d'information de Fisher est $\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}$

Remarque: avec la fonction de lien canonique $\boldsymbol{X}^\top(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{0}$

Partir d'une valeur initiale β^0

$\beta^0 \;\rightarrow\; \boldsymbol{\eta}^0 = \boldsymbol{X}\beta^0 \;\rightarrow\; \boldsymbol{\mu}^0 = g^{-1}(\boldsymbol{\eta}^0)$

Poser $\boldsymbol{z}^j = \boldsymbol{\eta}^j + (\boldsymbol{y} - \boldsymbol{\mu}^j) \cdot \left.\dfrac{\partial \boldsymbol{\eta}^j}{\partial \boldsymbol{\mu}}\right|_{\boldsymbol{\mu}=\boldsymbol{\mu}^j}$

Régresser (WLS) $\boldsymbol{z}^j$ sur $\boldsymbol{X}$ avec les poids $\boldsymbol{W}^j$ (évalués en $\boldsymbol{\mu}^j$)

Récupérer β^{j+1}, et itérer

$$\beta^{j+1} = \left(\boldsymbol{X}^\top \boldsymbol{W}^j \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{W}^j \left(\boldsymbol{X}\beta^j + \boldsymbol{\Delta}^j(\boldsymbol{y} - \boldsymbol{\mu}^j)\right)$$

$$\beta^{j+1} = \beta^j + \left(\boldsymbol{X}^\top \boldsymbol{W}^j \boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{W}^j \boldsymbol{\Delta}^j(\boldsymbol{y} - \boldsymbol{\mu}^j)$$

Lorsque $\|\beta^{j+1} - \beta^j\| < \epsilon$, on arrête, et $\widehat{\beta} = \beta^{j+1}$.

On pose $\boldsymbol{\Sigma} = \left(\boldsymbol{X}^\top \boldsymbol{W}^j \boldsymbol{X}\right)^{-1}$

Proposition sous les hypothèses mentionnées auparavant

- l'estimateur du maximum de vraisemblance $\widehat{\beta}$ existe et est unique,
- $\widehat{\beta} \overset{p.s.}{\to} \beta$ (fortement consistant)
- lorsque $n \to \infty$, $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta} \overset{\mathcal{L}}{\to} \mathcal{N}(\mathbf{0}, \varphi \boldsymbol{\Sigma})$

où $\boldsymbol{\Sigma} = \big(\boldsymbol{X}^\top \boldsymbol{W} \boldsymbol{X}\big)^{-1}$ et $\boldsymbol{W} = \text{diag}\big((V(\mu_i)g'(\mu_i)^2)^{-1}\big)$.

Proposition sous les hypothèses mentionnées auparavant

- lorsque $n \to \infty$, $\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta} \overset{\mathcal{L}}{\to} \mathcal{N}(\mathbf{0}, \varphi \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^\top)$
- lorsque $n \to \infty$, $\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu} \overset{\mathcal{L}}{\to} \mathcal{N}(\mathbf{0}, \varphi \boldsymbol{\Delta}^{-2}\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{X}^\top)$