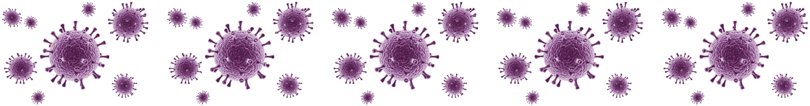
Modèles Linéaires Appliqués / Régression GLM & Sur-dispersion

Arthur Charpentier

UQAM

Hiver 2020 - COVID-19 # 17



Poisson Mélange

L'équidispersion est une propriété de la loi de Poisson: si $Y \sim \mathcal{P}(\lambda)$, $\mathbb{E}(Y) = \text{Var}(Y) = \lambda$.
Si on suppose maintenant que $Y|\Theta \sim \mathcal{P}(\lambda\Theta)$, où Θ suit une loi Gamma de paramètres identiques α (de telle sorte que $\mathbb{E}(\Theta) = 1$), on obtient la loi binomiale négative,

$$\mathbb{P}(Y = k) = \frac{\Gamma(k + \alpha^{-1})}{\Gamma(k+1)\Gamma(\alpha^{-1})} \left(\frac{1}{1 + \lambda/\alpha}\right)^{\alpha^{-1}} \left(1 - \frac{1}{1 + \lambda/\alpha}\right)^{k}, \forall k \in \mathbb{N}$$

On peut réécrire cette loi, en posant $r = \alpha^{-1}$ et $p = \dfrac{1}{1 + \alpha\lambda}$

$$f(y) = \binom{y}{y + r - 1} p^r [1 - p]^y, \forall k \in \mathbb{N}$$

ou encore

$$f(y) = \exp\left[y\log(1-p) + r\log p + \log\binom{y}{y+r-1}\right], \forall k \in \mathbb{N}$$

qui est une loi de la famille exponentielle, en posant $\theta = \log[1-p]$, $b(\theta) = -r\log(p)$ et $a(\varphi) = 1$.
Si on calcule la moyenne, on obtient

$$\mathbb{E}(Y) = b'(\theta) = \frac{\partial b}{\partial p}\frac{\partial p}{\partial \theta} = \frac{r(1-p)}{p} = \lambda,$$

et si on calcule la variance

$$\text{Var}(Y) = b''(\theta) = \frac{\partial^2 b}{\partial p^2}\left(\frac{\partial p}{\partial \theta}\right)^2 + \frac{\partial b}{\partial p}\frac{\partial^2 p}{\partial \theta^2} = \frac{r(1-p)}{p^2}$$

Poisson Mélange

Autrement dit

$$\text{Var}(Y) = \frac{1}{p}\mathbb{E}(Y) = [1 + \alpha \cdot \lambda] \cdot \lambda$$

Pour une régression binomiale négative de type 2 (NB2),

$$\mathbb{E}(Y) = \lambda = \mu \text{ et } \text{Var}(Y) = \lambda + \alpha\lambda^2.$$

Le lien canonique est $g(\lambda) = \theta$, i.e. $g(\mu) = \log\left(\frac{\alpha\mu}{1+\alpha\mu}\right)$

Remarque: si $\alpha = 0$, on a une loi de Poisson; si $\alpha = 1$, on a une loi géométrique.

Poisson Mélange

- Poisson, fonction variance $V(\mu) = \mu$, `ml.pois`
- quasiPoisson, fonction variance $V(\mu) = \varphi\mu$,
- géométrique, fonction variance $V(\mu) = \mu + \mu^2$,
- négative binomiale (NB1), fonction variance $V(\mu) = \mu + \alpha\mu$, `ml.nb1`
- négative binomiale (NB2), fonction variance $V(\mu) = \mu + \alpha\mu^2$, `ml.nb2`

on note que la surdispersion correspond à une hétérogénéité résiduelle, c'est à dire un effet aléatoire. Par exemple on peut supposer que

$$(Y|\boldsymbol{X}=\boldsymbol{X},\boldsymbol{Z}=\boldsymbol{z})\sim\mathcal{P}(\exp[\boldsymbol{X}^\top\beta+\boldsymbol{z}^\top\boldsymbol{\alpha}])$$

de telle sorte que si $u=\boldsymbol{z}^\top\boldsymbol{\alpha}-\mathbb{E}(\boldsymbol{Z}^\top\boldsymbol{\alpha}|\boldsymbol{X}=\boldsymbol{X})$, alors

$$(Y|\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z})\sim\mathcal{P}(\exp[\boldsymbol{x}^\top\gamma+u])$$

On a un modèle dit à effets fixes, au sens où

$$(Y|\boldsymbol{X}=\boldsymbol{x})\sim\mathcal{P}(\exp[\boldsymbol{x}^\top\gamma+U]),$$

où $U=\boldsymbol{z}^\top\boldsymbol{\alpha}-\mathbb{E}(\boldsymbol{Z}^\top\boldsymbol{\alpha}|\boldsymbol{X}=\boldsymbol{x})$. Par exemple, si on suppose que $U\sim\gamma(\alpha,\alpha)$, i.e. d'espérance 1 et de variance $\sigma^2=\alpha^{-1}$, alors

$$(Y|U=u)\sim\mathcal{P}(\lambda u)\text{ où }\lambda=\exp[\boldsymbol{x}^\top\gamma],$$

de telle sorte que $\mathbb{E}(Y|U=u)=\text{Var}(Y|U=u)$. Mais si on regarde la loi nonconditionnelle, $\mathbb{E}(Y)=\lambda$ alors que

$$\text{Var}(Y)=\text{Var}(\mathbb{E}[Y|U])+\mathbb{E}(\text{Var}(Y|))=\lambda+\lambda^2\sigma^2.$$

On peut alors proposer un test de la forme suivante: on suppose que

$$\text{Var}(Y|\boldsymbol{X}=\boldsymbol{x})=\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})+\tau\cdot\mathbb{E}(Y|\boldsymbol{X}=\boldsymbol{x})^2,$$

on on cherche à tester

$$H_0:\tau=0 \text{ contre } \tau>0.$$

Parmi les statistiques de test classique, on pourra considérer

$$T=\frac{\sum_{i=1}^n[(Y_i-\widehat{\mu}_i)^2-Y_i]}{\sqrt{2\sum_{i=1}^n\widehat{\mu}_i^2}}$$

qui suit, sous H_0, une loi normale centrée réduite.