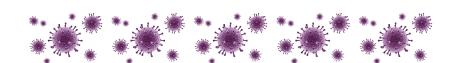
Modèles Linéaires Appliqués / Régression Régression GLM : Déviance & Résidus

Arthur Charpentier

UQAM

Hiver 2020 - COVID-19 # 15







Déviance

La déviance est l'écart entre la log-vraisemblance obtenue en β , et celle obtenue avec un modèle parfait (dit saturé),

$$D = 2\varphi \times [\log \mathcal{L}(\mathbf{y}) - \log \mathcal{L}(\widehat{\boldsymbol{\mu}})]$$

où $\widehat{\mu} = g^{-1}(X\widehat{\beta})$. On peut aussi définir la scaled deviance,

$$D^{\star} = \frac{D}{\varphi} = 2 \times [\log \mathcal{L}(\mathbf{y}) - \log \mathcal{L}(\widehat{\boldsymbol{\mu}})]$$

et la null deviance.

$$D_0^{\star} = 2 \times [\log \mathcal{L}(\mathbf{y}) - \log \mathcal{L}(\overline{\mathbf{y}})]$$

Normale :
$$D = \sum_{i=1}^{n} (y_i - \mu_i)^2$$

Poisson :
$$D = 2 \sum_{i=1}^{n} \left\{ y_i \ln \frac{y_i}{\mu_i} - (y_i - \mu_i) \right\}$$

Gamma:
$$D = 2\sum_{i=1}^{n} \left\{ -\ln \frac{y_i}{\mu_i} + \frac{y_i - \mu_i}{\mu_i} \right\}$$



Estimation de φ

Posons $\sigma^2 = a(\varphi)$

L'estimation de σ^2 est basée sur la déviance donnée par

$$D = \frac{1}{\sigma^2} \left\{ \sum_{i=1}^n y_i \widehat{\theta}_i - \sum_{i=1}^n b(\widehat{\theta}_i) \right\}.$$

Comme $\mathbb{E}(D) \approx n - p$, on pourrait estimer σ^2 par

$$\widetilde{\sigma}^2 = \frac{1}{n-p}D.$$

très instable. Afin d'éviter ces désagréments, on a recours à un développement de Taylor à l'ordre 2 de $L(\mathbf{y}|\mathbf{y},\sigma^2)$ qui nous donne

$$\widehat{\sigma}^2 = \frac{1}{n-p} (\mathbf{y} - \widehat{\boldsymbol{\mu}})' I_n(\widehat{\boldsymbol{\mu}}) (\mathbf{y} - \widehat{\boldsymbol{\mu}});$$

cette dernière estimation est souvent appelée estimation du χ^2 de Pearson.



Résidus

Les résidus de Pearson sont

$$\widehat{\varepsilon}_{P,i} = \frac{Y_i - \widehat{\mu}_i}{\sqrt{\mathsf{V}(\widehat{\mu}_i)}}$$

Les résidus de déviance sont

$$\widehat{\varepsilon}_{D,i} = \pm \sqrt{d_i}$$
, où $D = \sum_{i=1}^n d_i$.

Example: pour un modèle Gaussien, $\widehat{\varepsilon}_{P,i} = \widehat{\varepsilon}_{D,i} = Y_i - \widehat{\mu}_i$.

Example: pour un modèle de Poisson, $\widehat{\varepsilon}_{P,i} = \frac{Y_i - \widehat{\mu}_i}{\sqrt{\widehat{n}_i}}$ et

$$\widehat{\varepsilon}_{D,i} = \pm \sqrt{|Y_i \log[Y_i/\widehat{\mu}_i] - [Y_i - \widehat{\mu}_i]|}.$$

Example: pour un modèle Gamma, $\widehat{\varepsilon}_{P,i} = \frac{Y_i - \widehat{\mu}_i}{\widehat{n}_i}$ et

$$\widehat{\varepsilon}_{D,i} = \pm \sqrt{|\log[Y_i/\widehat{\mu}_i] - [Y_i - \widehat{\mu}_i]/\widehat{\mu}_i|}.$$



Résidus

> reg = glm(Y~YEARMARRIAGE, data=base, family=poisson)
Les résidus de Pearson sont

$$\widehat{\varepsilon}_{P,i} = \frac{Y_i - \widehat{\mu}_i}{\sqrt{\mathsf{V}(\widehat{\mu}_i)}}$$

1 > residuP = residuals(reg, type = "pearson")

Les résidus de déviance sont

$$\widehat{\varepsilon}_{D,i} = \pm \sqrt{d_i}$$
, où $D = \sum_{i=1}^n d_i$.

1 > residuP = residuals(reg, type = "deviance")