

# 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  $\beta$ , et celle obtenue avec un modèle parfait (dit saturé),

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

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

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

et la **null deviance**,

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

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 $\varphi$

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

L'estimation de  $\sigma^2$  est basée sur la déviance donnée par

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

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

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

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

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

cette dernière estimation est souvent appelée estimation du  $\chi^2$  de Pearson.

Les **résidus de Pearson** sont

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

Les **résidus de déviance** sont

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

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

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

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

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

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

```
1 > reg = glm(Y~YEARMARRIAGE, data=base, family=poisson)
```

Les **résidus de Pearson** sont

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

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

Les **résidus de déviance** sont

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

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





