M4: Poisson and gamma regression

W1: Poisson 02.10.2017

Poisson distribution $Y \sim \text{Poisson}(\lambda)$

$$f(y) = \frac{\lambda^y}{y!} e^{-\lambda} \qquad y = 0, 1, 2, \ldots$$

$E(Y) = \lambda$, $Var(Y) = \lambda$

Exponential family: $\theta = \ln(\lambda)$

$$\exp\left(\frac{y\theta - b(\theta)}{\varphi} w + c\right).$$

The log-linear model

1. $Y_i \sim \text{Poisson}(\lambda_i)$, $E(Y_i) = \lambda_i$, $Var(Y_i) = \lambda_i$
2. $\eta_i = x_i^T \beta$
3. $\eta_i = \ln(\lambda_i) \Leftrightarrow \lambda_i = \exp(\eta_i)$

$\Downarrow$

Interpret β's

$$\underbrace{\lambda_i}_{\text{mean of response } E(Y_i)} = \exp(\underbrace{x_i^T \beta}_{\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}}) = e^{\beta_0} \cdot (e^{\beta_1})^{x_{i1}} (e^{\beta_2})^{x_{i2}} \cdots (e^{\beta_k})^{x_{ik}}$$

What if $x_{i1} \to x_{i1} + 1$

$$\lambda_i(\dots x_{i1}\dots) = \exp(\beta_0+\beta_2 x_{i2}+\dots+\beta_k x_{ik})\cdot\left(e^{\beta_1}\right)^{x_{i1}}$$

$$\lambda_i(\dots x_{i1}+1\dots) = \text{— " —}\; e^{\beta_1}\cdot\left(e^{\beta_1}\right)^{x_{i1}}$$

one unit increase in x_{i1} has impact e^{β_1} on $E(Y_i)$

$\beta_1 = 0$: $e^{\beta_1} = 1$ → x_{i1} increase does not change $E(Y_i)$

$\beta_1 < 0$: $e^{\beta_1} < 1$ → $E(Y_i)$ decrease when x_{i1} increase

$\beta_1 > 0$: $e^{\beta_1} > 1$ → $E(Y_i)$ increase — " —

Parameter estimation

maximum likelihood

1) Likelihood: $L(\beta) = \prod_{i=1}^{n} \frac{\lambda_i^{y_i}}{y_i!} e^{-\lambda_i}$

$$\eta_i = x_i^T\beta,\quad \eta_i = \ln(\lambda_i)$$

2) Log-likelihood: $l(\beta) = \sum_{i=1}^{n} \underbrace{\left(y_i \ln\lambda_i - \lambda_i - \overbrace{\ln(y_i!)}^{c_i}\right)}_{l_i}$

$$= \sum_{i=1}^{n}\left(y_i\eta_i - \exp(\eta_i) - c_i\right)$$

$$= \sum_{i=1}^{n}\left(y_i x_i^T\beta - \exp(x_i^T\beta) - c_i\right)$$

3) Score function: $\underset{p\times 1}{s(\beta)} = \sum_{i=1}^{n} s_i(\beta)$

$$s_i(\beta) = \frac{\partial l_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta} = \Big(y_i - \underbrace{\exp(\eta_i)}_{\lambda}\Big)\cdot \overset{p\times 1}{x_i}$$

$$e^{x_i^T\beta} = \lambda_i$$

$$s_i(\beta) = (y_i - \lambda_i)x_i$$

$$E(s_i(\beta)) = E\Big((Y_i - \lambda_i)\underset{p\times 1}{x_i}\Big) = \Big(\overbrace{E(Y_i)}^{\lambda_i} - \lambda_i\Big)x_i$$

$$= 0$$

$$S(\beta) = \sum_{i=1}^{n}(y_i - \overset{\exp(x_i^T\beta)}{\lambda_i})x_i$$

$S(\hat{\beta}) = 0 \rightarrow$ to find MLE

4) Expected Fisher information matrix

$$F(\beta) = Cov(S(\beta)) = \sum_{i=1}^{n} Cov(s_i(\beta)) = \sum_{i=1}^{n} \overbrace{E\big(\underset{p\times 1}{s_i(\beta)}\cdot \underset{1\times p}{s_i(\beta)^T}\big)}^{F_i(\beta)\; p\times p}$$

$$E(s_i(\beta)) = 0$$

$$Cov(s_i(\beta)) = E\big((s_i(\beta) - E(s_i(\beta)))(s_i(\beta) - E(s_i(\beta)))^T\big)$$

$$F_i(\beta) = E\big((Y_i - \lambda_i)x_i\,(Y - \lambda_i)x_i^T\big)$$

$$= x_i x_i^T \underbrace{E\big((Y_i - \lambda_i)^2\big)}_{Var(Y_i) = \lambda_i} = \lambda_i x_i x_i^T$$

$$F(\beta) = \sum_{i=1}^{n} \lambda_i x_i x_i^T = X^T \Lambda X$$

$\Lambda = diag(\lambda_i)$

Find $\hat{\beta}_{MLE}$: $S(\hat{\beta}) = 0$ p eq.

use Fisher scoring: $\beta^{(t+1)} = \beta^{(t)} + \left(F(\beta^{(t)})\right)^{-1} S(\beta^{(t)})$

Asymptotic properties of MLE → same as for binomial

$$\hat{\beta} \approx N_p(\beta, F^{-1}(\hat{\beta})) \Rightarrow$$ CI & Hypotesis tests as before

$$(X^T \Lambda X)^{-1}$$

RESIDUALS AND GOODNESS OF FIT (GOF)

Saturated model = "our perfect model"

what is this for the Poisson case?

$Y_i \sim \text{Poisson}(\lambda_i)$

$\tilde{\lambda}_i = y_i$

our perfect fit

Candidate model: $\hat{\lambda}_i$

$$ML\ \hat{\beta} \rightarrow \hat{\lambda}_i = \exp(x_i^T \hat{\beta})$$

Deviance: $D = -2(\underset{\text{model}}{l(\text{candidate})} - \underset{\text{model}}{l(\text{saturated})})$

Homework

$$l_i(\beta) = y_i \ln \lambda_i - \lambda_i - C_i$$

$$D = 2\sum_{i=1}^{n} \left[y_i \ln\left(\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i) \right]$$

$$\hat{y}_i = \hat{\lambda}_i = \exp(x_i^T \hat{\beta})$$

Deviance residuals

$$d_i = \text{sign}(y_i - \hat{y}_i) \cdot \sqrt{2\left(y_i - \ln\frac{y_i}{\hat{y}_i}\right) - (y_i - \hat{y}_i)}$$

<u>Pearson residuals</u>

$$r_i = \frac{o_i - e_i}{\sqrt{e_i}}$$

$$= \frac{y_i - \hat{y}_i}{\sqrt{\hat{y}_i}}$$

$o_i = y_i$

$e_i = \hat{y}_i = \exp(x_i^T \hat{\beta})$

o = observed

e = "expected"

← more when we work with contingency table

Goodness of fit

$$D = \sum_{i=1}^{n} d_i^2 \quad , \quad X_p^2 = \sum_{i=1}^{n} r_i^2$$

↑ pearson

under H_0

$\approx \chi^2_{n-p}$

H_0: our model is good (compared to saturated model)

When D or X_p^2 is big → not good

Reject H_0 when $P(\chi^2_{n-p} > \underset{X_p^2}{D}) \leq \underset{0.05}{\alpha}$