

## • TEST FOR COMPARING NESTED MODELS

(test about a subset of coefficients)

We have a model  $Y_i \sim \text{Pois}(\lambda_i)$   $i=1, \dots, n$  indep.

with  $\log(\lambda_i) = \sum_{j=1}^p \beta_j x_{ij}$

$$= \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \beta_{p+1} x_{i,p+1} + \dots + \beta_p x_{in}$$

we call it the "full" model ( $\hat{\beta}$  is the proposed model).

we want to test

$$\begin{cases} H_0: \beta_{p+1} = \dots = \beta_p = 0 \\ H_1: \text{at least one } \beta_r \neq 0 \quad r \in \{p+1, \dots, p\} \end{cases}$$

We can partition the vector  $\hat{\beta} = \begin{bmatrix} \hat{\beta}^{(0)} \\ \hat{\beta}^{(1)} \\ \vdots \\ \hat{\beta}^{(p)} \end{bmatrix}$   $\hat{\beta}^{(0)} \in \mathbb{R}^{p_0}$   
 $\hat{\beta}^{(1)} \in \mathbb{R}^{p-p_0}$

with  $\hat{\beta}^{(0)} = [\beta_0, \beta_1, \dots, \beta_{p_0}]^T$

and  $\hat{\beta}^{(1)} = [\beta_{p_0+1}, \dots, \beta_p]^T$

$\begin{cases} H_0: \hat{\beta}^{(1)} = 0 \\ H_1: \hat{\beta}^{(1)} \neq 0 \end{cases}$

under  $H_0$  we have the "restricted" model

$Y_i \sim \text{Pois}(\lambda^{(0)})$  indep  $i=1, \dots, n$

with  $\log(\lambda^{(0)}) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_{p_0} x_{ip_0}$

To compare two nested models we use the LIKELIHOOD RATIO TEST : it compares the MAXIMUM of the likelihood under the full and under the restricted models.

- Remark : the likelihood is  $L(\beta)$ . The maximum is obtained at the MLE.

under  $H_1$ ,  $\hat{\beta} = (\hat{\beta}^{(0)}, \hat{\beta}^{(1)})^T$ , the MLE is  $\hat{\beta} = (\hat{\beta}^{(0)}, \hat{\beta}^{(1)})^T$

under  $H_0$ ,  $\hat{\beta} = (\hat{\beta}^{(0)}, 0)^T$ , the MLE is  $\hat{\beta} = (\hat{\beta}^{(0)}, 0)^T$

- likelihood under the full model:  $\hat{L}(\text{model})$ , with maximum  $\hat{L}(\text{model}) = L(\hat{\beta}^{(0)}, \hat{\beta}^{(1)})$

- likelihood under the restricted model:  $\hat{L}(\text{restricted})$ , with maximum  $\hat{L}(\text{restricted}) = L(\hat{\beta}^{(0)}, 0)$

When studied as functions of the ML estimator  $\hat{\beta}$ , they are random quantities (with a distribution).

### LIKELIHOOD RATIO TEST

$$W = 2 \frac{\log \hat{L}(\text{model})}{\log \hat{L}(\text{restricted})} = 2 \{ \hat{e}(\text{model}) - \hat{e}(\text{restricted}) \}$$

$$= 2 \{ e(\hat{\beta}^{(0)}, \hat{\beta}^{(1)}) - e(\hat{\beta}^{(0)}, 0) \} \sim \chi^2_{p-p_0} \quad \text{under } H_0$$

( # covariates under  $H_1$ ) - ( # covariates under  $H_0$ )

with the data we compute  $w_{\text{obs}} = 2(e(\hat{\beta}) - e(\hat{\beta}^{(0)}))$ .

What values do we expect under  $H_0$  and  $H_1$ ?

First, notice that  $\hat{\beta}$  is a CONSTRAINED estimate  $\Rightarrow \hat{e}(\text{model}) \geq \hat{e}(\text{restricted})$

$$\Rightarrow w_{\text{obs}} \geq 0$$

If  $H_0$  is true, the estimates under  $H_0$  and  $H_1$  will be similar

$$\Rightarrow \hat{e}(\text{model}) \approx \hat{e}(\text{restricted}) \Rightarrow w_{\text{obs}} \approx 0$$

If  $H_0$  is not true, removing  $\hat{\beta}^{(1)}$  will lead to a worse fit, hence

$$\Rightarrow \hat{e}(\text{model}) > \hat{e}(\text{restricted}) \Rightarrow w_{\text{obs}} \gg 0$$

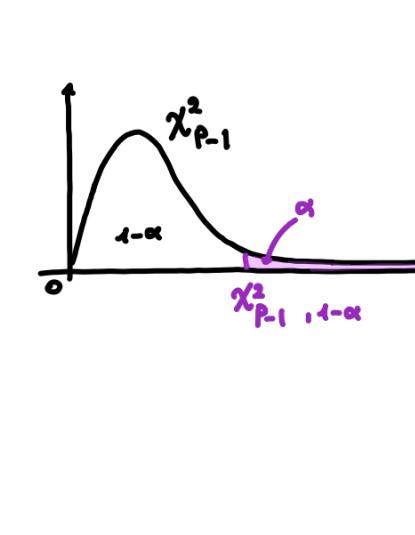
The reject region will comprise large values of the test

• fixed significance level  $\alpha$

$$\alpha = P_{H_0}(W > \chi^2_{p-p_0, 1-\alpha})$$

$$R_0 = (\chi^2_{p-p_0, 1-\alpha}; +\infty)$$

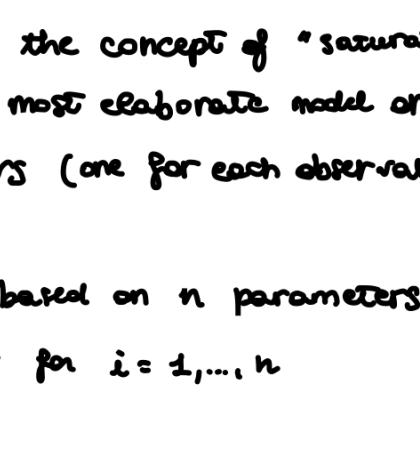
↳  $(1-\alpha)$ -quantile of a  $\chi^2$  distribution with  $p-p_0$  d.f.



• p-value

$$w_{\text{obs}} = P_{H_0}(W > w_{\text{obs}})$$

$$\text{with } W \stackrel{H_0}{\sim} \chi^2_{p-p_0}$$



## • TEST ABOUT THE OVERALL SIGNIFICANCE

Similarly to the LR, we can do:

$$\begin{cases} H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0 \\ H_1: \exists r \in \{2, \dots, p\}: \beta_r \neq 0 \end{cases}$$

We can use the test for nested models with  $\beta_0 = 1$ .

In this case we compare the full model with a model with only the intercept (null model).

Under  $H_0$ :

$Y_i \sim \text{Pois}(\lambda_i)$  independent for  $i=1, \dots, n$

with  $\lambda_i = e^{\beta_0} = \lambda$  equal for all  $i$

Remark:  $\lambda = e^{\beta_0} \Leftrightarrow \beta_0 = \log \lambda$  this is a reparametrization (one-to-one correspondence)

likelihood

$$L(\lambda) = \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{y_i}}{y_i!} \propto e^{-n\lambda} \lambda^{\sum y_i}$$

log-likelihood

$$e(\lambda) = -n\lambda + \sum_{i=1}^n y_i \cdot \log \lambda = -n\lambda + n\bar{y} \log \lambda$$

score function

$$e'_\lambda(\lambda) = -n + \frac{n\bar{y}}{\lambda}$$

likelihood equation

$$e'_\lambda(\lambda) = 0 \Rightarrow -n\lambda = -n\bar{y} \Rightarrow \hat{\lambda} = \bar{y} \quad \text{MLE of } \lambda \text{ under the null model}$$

we estimate the common mean using the sample mean

Moreover, since  $\log \lambda_i = \log \lambda = \beta_0$

we automatically obtain  $\hat{\beta}_0 = \log \hat{\lambda} = \log \bar{y}$

second derivative

$$e''_{\lambda\lambda}(\lambda) = -n\bar{y} \cdot \frac{1}{\lambda^2} \Rightarrow e''_{\lambda\lambda}(\hat{\lambda}) = -\frac{n}{\bar{y}} < 0 \quad \text{it's a max}$$

$$\text{Hence } \hat{e}(\text{restricted}) = e(\hat{\lambda}) = -n\bar{y} + n\bar{y} \cdot \log \bar{y} = n(\bar{y} \log \bar{y} - \bar{y})$$

Under  $H_1$  we have the model with p covariates

$$\lambda_i = \exp \{ \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} \}$$

we estimate  $\hat{\beta}$  numerically, and we compute

$$\hat{e}(\text{model}) = e(\hat{\beta}) = -\sum_{i=1}^n \hat{\lambda}_i + \sum_{i=1}^n y_i \cdot \log \hat{\lambda}_i = -\sum_{i=1}^n e^{\hat{\beta}^T \hat{x}_i} + \sum_{i=1}^n y_i e^{\hat{\beta}^T \hat{x}_i}$$

The likelihood ratio test in this case is

$$W = 2 \{ \hat{e}(\text{model}) - \hat{e}(\text{restricted}) \} = 2 \{ e(\hat{\beta}) - e(\hat{\lambda}) \} \sim \chi^2_{p-1} \quad \text{under } H_0$$

With the data:

$$w_{\text{obs}} = 2 \left\{ -\sum_{i=1}^n \hat{\lambda}_i + \sum_{i=1}^n y_i \cdot \log \hat{\lambda}_i - n(\bar{y} \log \bar{y} - \bar{y}) \right\}$$

$$= 2 \left\{ \sum_{i=1}^n y_i \cdot \log \hat{\lambda}_i - \sum_{i=1}^n \hat{\lambda}_i + n\bar{y} \right\}$$

$$= 2 \left\{ \sum_{i=1}^n y_i \cdot \log \frac{\hat{\lambda}_i}{\bar{y}} - \sum_{i=1}^n \hat{\lambda}_i + n\bar{y} \right\}$$

The reject region comprises large values of the test

• fixed significance level  $\alpha$

$$\alpha = P_{H_0}(W > \chi^2_{p-1, 1-\alpha})$$

$$R_0 = (\chi^2_{p-1, 1-\alpha}; +\infty)$$

↳  $(1-\alpha)$ -quantile of a  $\chi^2$  distribution with  $p-1$  d.f.



• p-value

$$w_{\text{obs}} = P_{H_0}(W > w_{\text{obs}})$$

$$\text{with } W \stackrel{H_0}{\sim} \chi^2_{p-1}$$



## • GOODNESS OF FIT OF THE MODEL

We first need to introduce the concept of "saturated model".

SATURATED MODEL : it is the most elaborate model one can estimate :

i.e., a model with  $n$  parameters (one for each observation).

→ model for  $n$  observations based on  $n$  parameters

$Y_i \sim \text{Pois}(\lambda_i)$  independent for  $i=1, \dots, n$

parameters  $(\lambda_1, \dots, \lambda_n)$

$$L(\lambda_1, \dots, \lambda_n) = \prod_{i=1}^n [e^{-\lambda_i} \lambda_i^{y_i}] = e^{-\sum \lambda_i} \prod_{i=1}^n \lambda_i^{y_i}$$

$$e(\lambda_1, \dots, \lambda_n) = -\sum \lambda_i + \sum_{i=1}^n y_i \cdot \log \lambda_i = -\sum \lambda_i + \sum_{i=1}^n y_i \log \lambda_i$$

$$e'_{\lambda_i}(\lambda_1, \dots, \lambda_n) = \left[ \frac{\partial}{\partial \lambda_1} e(\lambda_1, \dots, \lambda_n) \dots \frac{\partial}{\partial \lambda_n} e(\lambda_1, \dots, \lambda_n) \right]^T \quad \text{with } \frac{\partial}{\partial \lambda_i} e(\lambda_1, \dots, \lambda_n) = -1 + \frac{y_i}{\lambda_i}$$

$$\frac{\partial^2 e(\lambda_1, \dots, \lambda_n)}{\partial \lambda_i^2} = 0 \Rightarrow \frac{\lambda_i}{\lambda_i^2} = 1 \Rightarrow \lambda_i^2 = \lambda_i \quad \text{MLE of } \lambda_i \text{ under the saturated model}$$

The log-likelihood computed at  $(\lambda_1^*, \dots, \lambda_n^*)$  is

$$e(\lambda_1^*, \dots, \lambda_n^*) = -\sum \lambda_i^* + \sum_{i=1}^n y_i \cdot \log \lambda_i^*$$

$$= -\sum \lambda_i^* + \sum_{i=1}^n y_i \cdot \log y_i \quad (1)$$

Remark : the saturated model has a perfect fit since it models each observation with itself : it is INTERPOLATING THE DATA.

However, it is useless since it is not providing any simplification of the phenomenon, and it is not modeling its variability.

The residual deviance is more useful when used to compare different models (on the same data), with the same number of covariates  $p$  (but different covariates).

## • RELATIONSHIP BETWEEN SATURATED, PROPOSED AND NULL MODEL

Notice that the saturated model and the null model (with only the intercept) are the two extreme cases:

• SATURATED model :  $n$  parameters  $\uparrow$  nested

• proposed model :  $p$  parameters  $\downarrow$  nested

• null model : 1 parameter  $\uparrow$  nested

Let's analyze the different likelihood ratio tests that we can define with these models:

### I. between the SATURATED AND PROPOSED

The proposed model can be seen as a restricted model w.r.t. the saturated model

$$\text{LR test: } 2 \{ \hat{e}(\text{saturated}) - \hat{e}(\text{model}) \} = D(\text{model})$$

this is the DEVIANCET OF THE MODEL (RESIDUAL DEVIANCE)

### II. between the SATURATED AND THE NULL

Also the null model can be seen as a restricted model w.r.t. the saturated model

$$\text{LR test: } 2 \{ \hat{e}(\text{saturated}) - \hat{e}(\text{null}) \} = D(\text{$$