3.3 Poisson Regression

1. Introduction

Poisson regression and Log-Linear regression

Poisson distribution is used for **count data**.

If a random variable Y is **Poisson** distributed then it has probability distribution

$$f(y) = \frac{\mu^y e^{-\mu}}{y!} \qquad y = 0, 1, \ldots \tag{3.3.1}$$

where μ is a parameter such that $\mathbb{E}(\mathrm{Y}) = \mu$ and which is often called "rate".

Example:

The number of tropical cyclones crossing the North Queensland coast can be represented as a Poisson random variable. Here, μ is the rate of tropical cyclones crossing the North Queensland coast in the cyclone season, from November to April.

The effect of explanatory variables on the response Y is modelled through the parameter μ.

- **Poisson regression:** the events relate to varying amounts of exposure which need to be taken into account when modelling the rate of events (explanatory variables are usually continuous or categorical)
- **Log-linear regression:** exposure is constant (explanatory variables are usually categorical)

Poisson regression

Let $\mathrm{Y}_1, \ldots, \mathrm{Y}_N$ be independent random variables, with Y_i denoting the *number of events observed from exposure* n_i for the i-th covariate pattern.

The expected value of Y_i is

$$\mathbb{E}(\mathrm{Y}_i) = \mu_i = n_i \theta_i. \tag{3.3.2}$$

The parameter θ_i depends on a set of explanatory variables $\mathbf{x}_i$ and this dependence is usually modelled as

$$\theta_i = e^{\mathbf{x}_i^\top \boldsymbol{\beta}} \tag{3.3.3}$$

The **natural link function** is the **logarithmic function**

$$\log \mu_i = \log n_i + \mathbf{x}_i^\top \boldsymbol{\beta} \tag{3.3.4}$$

The term $\log n_i$ is called the **offset**, a known constant.

This model has an interpretation in terms of **rate ratio (RR)**. Suppose we have a dummy variable

$$x_j = \begin{cases} 0 & \text{if factor is absent} \\ 1 & \text{if factor is present} \end{cases}$$

Then the rate ratio for presence versus absence is

$$\mathrm{RR} = \frac{\mathbb{E}(\mathrm{Y}_i|present)}{\mathbb{E}(\mathrm{Y}_i|absent)} = e^{\beta_j} \tag{3.3.5}$$

provided all the other explanatory variables stay the same.

Similarly, for a continuous explanatory variable x_l, e^{β_l} represents the multiplicative effect on the rate μ.

2. Residuals

Residuals for the Poisson model

Once the regression coefficients are estimated through $\hat{\boldsymbol{\beta}}$, the fitted values are given by

$$\hat{\mathrm{Y}}_i = \hat{\mu} = n_i e^{\mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}, \qquad i = 1, \ldots, N \tag{3.3.6}$$

Similarly to the case of linear regression, we can call these **fitted values** e_i since they estimate the expected values $\mathbb{E}(\mathrm{Y}_i) = \mu_i$. Then using the fact that $\mathbb{V}\mathrm{ar}(\mathrm{Y}_i) = \mathbb{E}(\mathrm{Y}_i)$, the **Pearson Residuals** are

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

where o_i (or Y_i) denotes the observed count and e_i (or $\hat{\mathrm{Y}}_i$) the expected count.

The Pearson residuals are used to compute the *Pearson chi-squared goodness of fit statistic*

$$P^2 = \sum_{i=1}^{N} P_i^2 = \sum_{i=1}^{N} \frac{(o_i - e_i)^2}{e_i} \tag{3.3.7}$$

The **deviance residuals** are

$$d_i = \mathrm{sign}(o_i - e_i)\sqrt{2[o_i \log(o_i/e_i) - (o_i - e_i)]}, \quad i = 1, \ldots, N$$

Again, $D = \sum d_i^2$ is the **deviance**

$$D = 2\sum_{i=1}^{N} [o_i \log(o_i/e_i) - (o_i - e_i)] \tag{3.3.8}$$

Since in many cases $\sum_{i=1}^{N} o_i = \sum_{i=1}^{N} e_i$,

$$D = 2\sum_{i=1}^{N} [o_i \log(o_i/e_i)]$$

Now, if we use a **Taylor expansion** of $o_i \log(o_i/e_i)$:

$$o_i \log(o_i/e_i) = (o_i - e_i) + \frac{1}{2}\frac{(o_i - e_i)^2}{e_i} + \cdots,$$

then

$$
\begin{aligned}
D &= 2\sum_{i=1}^{N}\left[(o_i - e_i) + \frac{1}{2}\frac{(o_i - e_i)^2}{e_i} - (o_i - e_i)\right] \\
&= \sum_{i=1}^{N}\left[\frac{(o_i - e_i)^2}{e_i}\right] = P^2
\end{aligned}
$$

An important aspect of D and P^2 is that they depend on the fitted values and the observations, and do not depend on any nuisance parameters (like σ^2 for the Normal).

3. Modelling example

Example (Poisson regression)

Consider the artificial dataset **`poisson`** from the **`dobson`** package in **`R`** and which represents counts **`y`** observed at various values of a covariate **`x`**.

3.1 The data

The data are plotted using the code below.

```
library(dobson)
data("poisson")
attach(poisson)

plot(x, y, pch=16, xaxt="n", xlim=c(-1, 1), ylim=c(0, 15), xlab="X", ylab="Y")
axis(1, at=c(-1, 0, 1))
```

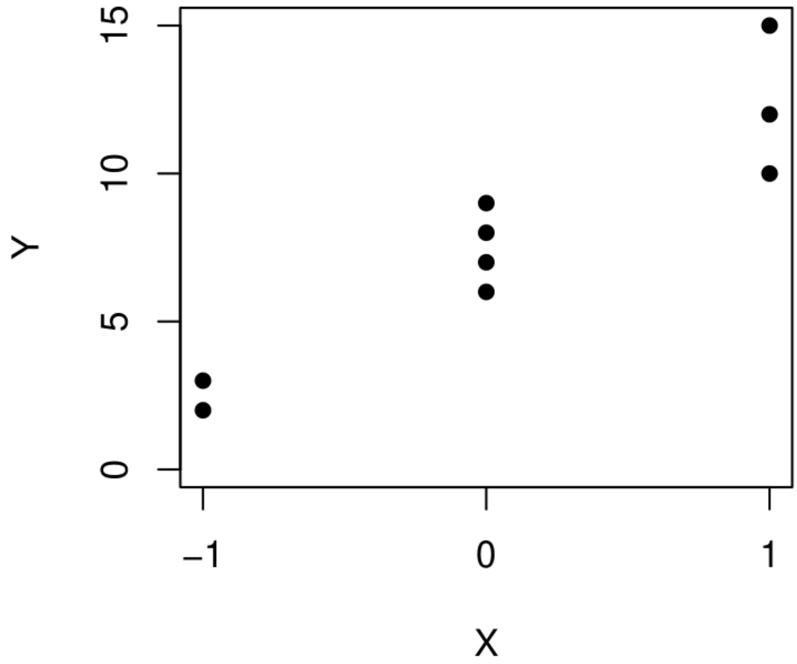

Figure 3.3.1: Illustration of the data.

We assume that the responses Y_i *are Poisson random variables*. In practice, such assumptions would be made noticing that the variability increases with Y. If Y_i is Poisson distributed then $\mathbb{E}(\mathrm{Y}_i) = \mathbb{V}\mathrm{ar}(\mathrm{Y}_i)$.

Let us model the relationship between Y_i and X_i by the straight line

$$\begin{aligned}\mathbb{E}(\mathrm{Y}_i) = \mu_i &= \beta_1 + \beta_2 \mathrm{X}_i \\ &= \mathbf{x}_i^\top \boldsymbol{\beta}\end{aligned}$$

for $i = 1, \ldots, N$, where

$$\boldsymbol{\beta} = (\beta_1, \beta_2)^\top, \quad \text{and} \quad \mathbf{x}_i = (1, \mathrm{X}_i)^\top.$$

3.2 Modelling with the identity link function

We take the *link function to be the identity function*

$$g(\mu_i) = \mu_i = \mathbf{x}_i^\top \boldsymbol{\beta} = \eta_i.$$

Therefore, $d\eta_i/d\mu_i = 1$. In this case we have

$$w_{ii} = \frac{1}{\mathbb{V}\mathrm{ar}(\mathrm{Y}_i)} = \frac{1}{\beta_1 + \beta_2 \mathrm{X}_i}.$$

Using an estimate $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2)^\top$ for β, we obtain

$$\begin{aligned}
\mathrm{Z}_i &= \mathbf{x}_i^\top \hat{\boldsymbol{\beta}} + (\mathrm{Y}_i - \mu_i)\left(\frac{d\eta_i}{d\mu_i}\right) \\
&= \hat{\beta}_1 + \hat{\beta}_2 x_i + (\mathrm{Y}_i - \hat{\beta}_1 - \hat{\beta}_2 \mathrm{X}_i) \\
&= \mathrm{Y}_i
\end{aligned}$$

Additionally,

$$\mathcal{I} = \mathbf{X}^\top \mathbf{W} \mathbf{X} = \begin{pmatrix} \sum_{i=1}^N \frac{1}{\hat{\beta}_1 + \hat{\beta}_2 x_i} & \sum_{i=1}^N \frac{x_i}{\hat{\beta}_1 + \hat{\beta}_2 x_i} \\ \sum_{i=1}^N \frac{x_i}{\hat{\beta}_1 + \hat{\beta}_2 x_i} & \sum_{i=1}^N \frac{x_i^2}{\hat{\beta}_1 + \hat{\beta}_2 x_i} \end{pmatrix}.$$

and

$$\mathbf{X}^\top \mathbf{W} \mathbf{z} = \begin{pmatrix} \sum_{i=1}^N \frac{y_i}{\hat{\beta}_1 + \hat{\beta}_2 x_i} \\ \sum_{i=1}^N \frac{x_i y_i}{\hat{\beta}_1 + \hat{\beta}_2 x_i} \end{pmatrix}.$$

The maximum likelihood estimates are obtained iteratively from the equations

$$(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{(m-1)} \hat{\boldsymbol{\beta}}^{(m)} = \mathbf{X}^\top \mathbf{W} \mathbf{z}^{(m-1)}.$$

Example:

The maximum likelihood estimates are

$$\hat{\boldsymbol{\beta}} = (7.45163, 4.93530)^\top.$$

At these values the inverse of the information matrix $\mathcal{I}^{-1} = \left(\mathbf{X}^\top \mathbf{W} \mathbf{X}\right)^{-1}$ is

$$\begin{pmatrix} 0.7816754 & 0.4165551 \\ 0.4165551 & 1.1863043 \end{pmatrix}.$$

This is the variance-covariance matrix for $\hat{\boldsymbol{\beta}}$, giving the standard errors of the estimates $\sqrt{0.7817} = 0.8841$ for $\hat{\beta}_1$ and $\sqrt{1.1863} = 1.0892$ for $\hat{\beta}_2$.

```
library(dobson)
data("poisson")
attach(poisson)

# Iterative Least Squares procedure
ILS <- function(niter, b, x, y){

  N <- length(y)
  y <- z <- as.matrix(y)
  X <- matrix( c( rep(1,N), x ), N, 2)
  bs <- matrix(nrow=niter, ncol=2)

  for(i in 1:niter){
    w <- 1/(b[1] + b[2]*x)
    W <- diag(w, N, N)
    b <- bs[i, ] <- solve( t(X) %*% W %*% X) %*% t(X) %*% W %*% z
  }
  inv.inform <- solve(t(X) %*% W %*% X)

  return(list(bs=bs, inv.inform=inv.inform, sd=sqrt(diag(inv.inform))) )
}

est <- ILS(5, b=c(7,5), x=x, y=y )
est
```

Example:

The 95% confidence intervals for $\hat{\beta}_1$ and $\hat{\beta}_2$ are respectively $(5.7188, 9.1845)$ and $(2.8005, 7.0700)$.

```
library(dobson)
data("poisson")
attach(poisson)

# Iterative Least Squares procedure
ILS <- function(niter, b, x, y){

  N <- length(y)
  y <- z <- as.matrix(y)
  X <- matrix( c( rep(1,N), x ), N, 2)
  bs <- matrix(nrow=niter, ncol=2)

  for(i in 1:niter){
    w <- 1/(b[1] + b[2]*x)
    W <- diag(w, N, N)
```

```r
    b <- bs[i, ] <- solve( t(X) %*% W %*% X) %*% t(X) %*% W %*% z
  }
  inv.inform <- solve(t(X) %*% W %*% X)

  return(list(bs=bs, inv.inform=inv.inform, sd=sqrt(diag(inv.inform))) )
}

est <- ILS(5, b=c(7,5), x=x, y=y )

UCI <- tail(est$bs,1) + qnorm(0.975)*est$sd
LCI <- tail(est$bs,1) - qnorm(0.975)*est$sd

LCI
UCI
```

The results can be obtained using the `glm()` function with **identity link** function as follows

```r
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(link="identity"),data=poisson)
summary(res.p)
```

3.3 Modelling with the canonical link function

Now let's use the **canonical link** function (logarithm) since it seems from Figure 3.3.1 that the data increase more than linearly or, at least, the variance increases as the value of the predictor increases.

```
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)
summary(res.p)
```

The output from the `resp.p` object indicates that both the intercept and the regression coefficient are strongly significant.

The **rate ratio** of increasing the predictor of 1 unit is $e^{0.6698} = 1.953818$ times higher.

```
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)

RR <- exp(res.p$coefficients[2])
RR
```

The **Pearson residuals** are given by

```
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)

fit_p <- fitted(res.p)
pearRes <- c(y - fit_p) / sqrt(fit_p)
pearRes
```

The **Pearson chi-squared goodness of fit statistic** is

```
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)

fit_p <- fitted(res.p)
pearRes <- c(y - fit_p) / sqrt(fit_p)
```

```
chisq <- sum(pearRes^2)
chisq
```

For the **deviance** statistic we get

```
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)

fit_p <- fitted(res.p)

devRes <- sign(y - fit_p) * sqrt(2*(y * log(y/fit_p) - (y - fit_p)))
Dstat <- sum(devRes^2)
Dstat
```

The values of the D and P^2 statistics are small compared with the chi-squared distribution of $9-2$ degrees of freedom.

```
library(dobson)
data("poisson")
attach(poisson)

qchisq(0.95,nrow(poisson)-2)
```

The difference ΔD in deviance with the null model is 15.48186, larger than the $1-\alpha$ quantile of the χ^2_{p-1} meaning that we reject the null hypothesis.

```
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)
res.0 <- glm(y ~ 1 , family=poisson(),data=poisson)

delta.D <- 2*(logLik(res.p) - logLik(res.0))
delta.D
# 'log Lik.' 15.48186 (df=2)
if(delta.D>qchisq(0.95, df=1)){
  cat("There is evidence against H0")
}else{
  cat("There is not enough evidence against H0")
}
# There is evidence against H0
```

If we compute the **pseudo** R^2 statistic we get it is relatively small but we should be alarmed since the pseudo R^2 is a measure of the predictability of the individual outcome Y_i rather than the predictability of all the event rates (Mittlbock and Heinzl, 2001). (See also Dobson & Barnett, p.164 and example 9.2.1, p.201-204).

```r
library(dobson)
data("poisson")
attach(poisson)

res.p <- glm(y ~ x , family=poisson(),data=poisson)
res.0 <- glm(y ~ 1 , family=poisson(),data=poisson)

pseudoR2 <- (logLik(res.0) - logLik(res.p)) / logLik(res.0)
pseudoR2
```

Activity

The following questions are taken from Dobson & Barnett (Exercise 9.1).

Let $\mathrm{Y}_1, \ldots, \mathrm{Y}_N$ be independent random variables with $\mathrm{Y}_i \sim \mathrm{Pois}(\mu_i)$ and $\log \mu_i = \beta_1 + \sum_{j=2}^{J} x_{ij}\beta_j$, $i = 1, \ldots, N$.

Question 1 *Submitted Mar 16th 2023 at 10:22:05 pm*

Show that the score statistic for β_1 is $U_1 = \sum_{i=1}^{N}(\mathrm{Y}_i - \mu_i)$.

asdf

Question 2 *Submitted Mar 16th 2023 at 10:22:49 pm*

Hence, show that for the maximum likelihood estimates $\hat{\mu}_i$, $\sum \hat{\mu}_i = \sum y_i$

asdf

Question 3 *Submitted Mar 16th 2023 at 10:23:48 pm*

Deduce that the deviance can be written as $D = \sum_{i=1}^{N} o_i \log\left(\frac{o_i}{e_i}\right)$ where o_i denotes the observed value y_i and e_i is used to denote the estimated expected value $\hat{y}_i$.

asdf