Insolvency - (Quasi-)Poisson Model and Negative Binomial Model

January 25, 2024

First the insolvency data are loaded:

```
library(catdata)
data(insolvency)
attach(insolvency)

## Das folgende Objekt ist maskiert encephalitis:
##
##     year
```

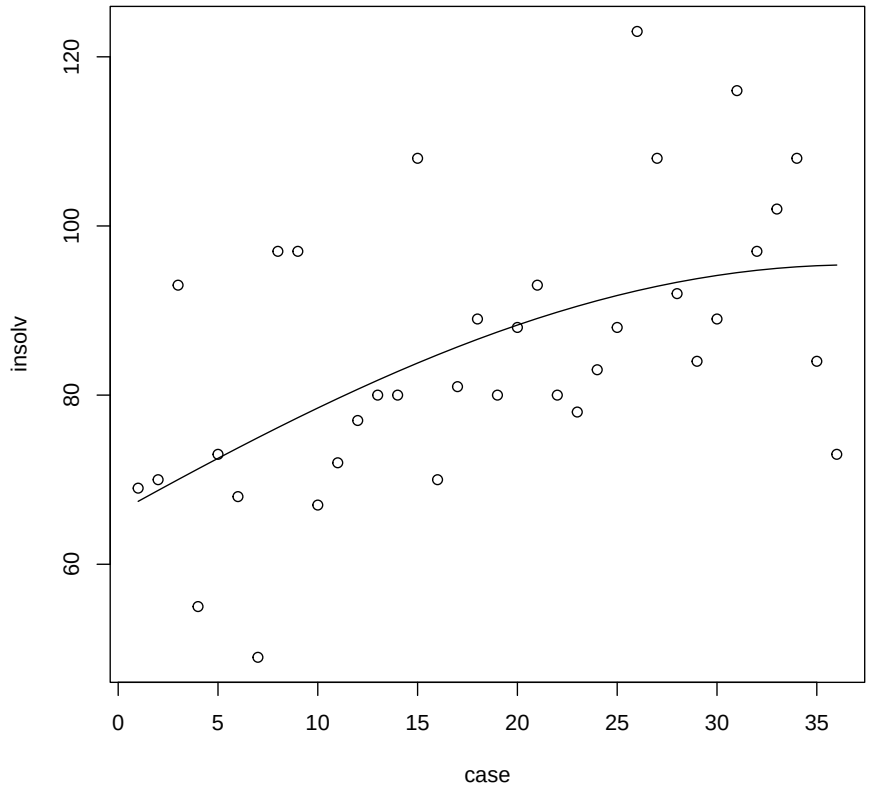
For the number of insolvent firms between 1994 and 1996 a Poisson model is fitted with time as predictor. Time is considered as a number from 1 to 36, denoting the month from January 1994 to December 1996.

```
ins1 <- glm(insolv ~ case + I(case^2), family=poisson(link=log), data=insolvency)
summary(ins1)

##
## Call:
## glm(formula = insolv ~ case + I(case^2), family = poisson(link = log),
##     data = insolvency)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2037  -0.9083  -0.2517   0.4880   3.0340
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.1916952  0.0617994  67.827  < 2e-16 ***
## case         0.0197825  0.0073901   2.677  0.00743 **
## I(case^2)   -0.0002670  0.0001896  -1.409  0.15897
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 108.128  on 35  degrees of freedom
```

```
## Residual deviance:  75.287  on 33  degrees of freedom
## AIC: 306.82
##
## Number of Fisher Scoring iterations: 4

# plot(ins1)
```

Scatter-Plot of number of insolvent firms dependent of the month (1-36). With estimated curve of the log-linear model.

```
plot(case, insolv)
points(ins1$fitted.values, type="l")
```



In many real-world datasets the variance of count-data is higher than predicted by the Poisson distribution. So next a Poisson model with disperison parameter is fitted (Quasi-Poisson model).

```
ins2 <- glm(insolv ~ case + I(case^2), family=quasipoisson, data=insolvency)
summary(ins2)

##
```

```
## Call:
## glm(formula = insolv ~ case + I(case^2), family = quasipoisson,
##     data = insolvency)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.2037  -0.9083  -0.2517   0.4880   3.0340
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.1916952  0.0939826  44.601   <2e-16 ***
## case         0.0197825  0.0112387   1.760   0.0876 .
## I(case^2)   -0.0002670  0.0002883  -0.926   0.3611
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 2.312738)
##
##     Null deviance: 108.128  on 35  degrees of freedom
## Residual deviance:  75.287  on 33  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 4

# plot(ins2)
```

An alternative to a quasi-poisson model is to use the negative binomial distribution.

```
library(MASS)
ins3 <- glm.nb(insolv ~ case + I(case^2),data=insolvency)
summary(ins3)

##
## Call:
## glm.nb(formula = insolv ~ case + I(case^2), data = insolvency,
##     init.theta = 77.92952593, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.3666  -0.6333  -0.1710   0.3350   2.0042
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  4.1953863  0.0861256  48.712   <2e-16 ***
## case         0.0192833  0.0105170   1.834   0.0667 .
## I(case^2)   -0.0002546  0.0002728  -0.933   0.3506
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## (Dispersion parameter for Negative Binomial(77.9295) family taken to be 1)
##
##     Null deviance: 52.104  on 35  degrees of freedom
## Residual deviance: 36.312  on 33  degrees of freedom
## AIC: 296.27
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  77.9
##           Std. Err.:  35.5
##
##  2 x log-likelihood:  -288.269
```

Since counts are rather large in addition a normal distribution model is fitted.

```
ins4 <- glm(insolv ~ case + I(case^2), family=gaussian(link=log), data=insolvency)
summary(ins4)
```

```
##
## Call:
## glm(formula = insolv ~ case + I(case^2), family = gaussian(link = log),
##     data = insolvency)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -25.809   -8.744   -2.374    4.560   30.480
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.1836089  0.1005663  41.600   <2e-16 ***
## case         0.0208026  0.0115423   1.802   0.0806 .
## I(case^2)   -0.0002915  0.0002896  -1.007   0.3214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 193.2793)
##
##     Null deviance: 9147.0  on 35  degrees of freedom
## Residual deviance: 6378.1  on 33  degrees of freedom
## AIC: 296.54
##
## Number of Fisher Scoring iterations: 4
```