

Exercise 7.30

Per August Jarval Moen

26/10/2023

Let Y_i denote the annual number of shark attacks in Florida, for $i = 1, \dots, n = 13$. We want to test $H_0 : Y_i \stackrel{\text{ind}}{\sim} \text{Poisson}(\mu)$ for some $\mu > 0$ versus the alternative hypothesis $H_1 : Y_i \stackrel{\text{ind}}{\sim} \text{NegBin}(\mu, \gamma)$ ¹ for some $\mu > 0$ and $\gamma > 0$.

The pmf of the negative binomial distribution $\text{NegBin}(\mu, \gamma)$ will converge to the pmf of a $\text{Poisson}(\mu)$ as $\gamma \rightarrow 0$. Informally, we have that the Poisson distribution corresponds to a negative binomial distribution with $\gamma = 0$. Therefore, the model implied by the null hypothesis H_0 is in some sense nested in the model implied by the alternative hypothesis H_1 . It is therefore reasonable to test the null hypothesis by a likelihood ratio test, as the models are nested in some sense. The likelihood ratio statistic is given by

$$Z_{\text{LR}} = -2 \left\{ \sup_{\mu > 0} \ell_{\text{Poisson}}(\mu) - \sup_{\mu > 0, \gamma > 0} \ell_{\text{NegBin}}(\mu, \gamma) \right\},$$

where $\ell_{\text{Poisson}}(\mu)$ is the log-likelihood of the Poisson model given the observed data, and $\ell_{\text{NegBin}}(\mu, \gamma)$ is the log-likelihood of the Negative Binomial model given the observed data.

However, the Negative Binomial distribution is only defined for $\gamma > 0$. Hence, $\gamma = 0$ lies in the boundary of the parameter space of the Negative Binomial model (the alternative hypothesis), which violates the assumptions of Wilk's Theorem. It turns out that this is not a problem: It can be shown that under H_0 ,

$$Z_{\text{LR}} \xrightarrow{d} \frac{1}{2}\delta_0 + \frac{1}{2}\chi_1^2 \quad \text{as } n \rightarrow \infty, \quad (1)$$

where δ_0 indicates the distribution with a point-mass in zero (it will be zero with probability 1). So under H_0 , the likelihood ratio statistic should be 0 with probability $\frac{1}{2}$ and be distributed as a chi square with probability $\frac{1}{2}$. To obtain a correct p-value, we can perform the test as if Z_{LR} is a chi square, and then divide the p-value by two.

So let's try it out. First we load the data and fit the Poisson model:

```
Y = c(33, 29, 29, 12, 17, 21, 31, 28, 19, 14, 11, 26, 23)
```

We use the R function `glm.nb` from library `MASS` to fit the Negative Binomial model. Note that R uses the parameterization $\theta = 1/\gamma$ for the overdispersion parameter.

```
library(MASS)
negbin = glm.nb(Y~1)
```

We can test H_0 versus H_1 by using the function `odTest` in the `pscl` library:

```
library(pscl)
odTest(negbin)
```

```
## Likelihood ratio test of H0: Poisson, as restricted NB model:
```

¹Here the negative binomial is parametrized by the mean $\mu > 0$ and the "overdispersion" parameter $\gamma > 0$, as on page 248 in the book.

```
## n.b., the distribution of the test-statistic under  $H_0$  is non-standard
## e.g., see help(odTest) for details/references
##
## Critical value of test statistic at the alpha= 0.05 level: 2.7055
## Chi-Square Test Statistic = 6.5206 p-value = 0.005332
```

The p-value of the likelihood ratio test using the result in (1) is 0.005332. This is very small, so we reject H_0 in favor of H_1 .

Lastly, we check if there is any evidence of a positive linear trend over time. We use a negative binomial model:

```
Y = c(33,29,29,12,17,21,31,28,19,14,11,26,23)
time = 1:length(Y)
data = data.frame(Y,time)
negbin2 = glm.nb(Y~time,data=data)
summary(negbin2)

##
## Call:
## glm.nb(formula = Y ~ time, data = data, init.theta = 20.15274625,
##       link = log)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.0006  -1.0233   0.3683   0.7313   1.1484
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.32176    0.17685  18.783  <2e-16 ***
## time        -0.03047    0.02280  -1.337   0.181
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(20.1527) family taken to be 1)
##
##      Null deviance: 15.391  on 12  degrees of freedom
## Residual deviance: 13.522  on 11  degrees of freedom
## AIC: 92.879
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta:  20.2
##             Std. Err.: 15.7
##
## 2 x log-likelihood:  -86.879
```

The estimated coefficient of time is not significantly different from zero, even for one-sided tests. Hence we do not have any evidence that there is any linear trend over time at all.