Individual Assignment - Optimization Techniques

By: Marie Reine Obama

April 2020

Estimating a Poisson Regression using optimization methods

The goal of this project is to fit a Poisson regression using optimization techniques. For this purpose, we created a dataset with random integer. To make it realistic, we simulated a case of a breakdown in a factory.

Therefore, the goal of this simulated case is to determine the average number of breakdowns that can occur in a factory. We assume that this number is dependent of the number of verification tests performed by the engineers. To do so, we will use a probabilistic Poisson model.

The Poisson regression is given by the formula below:

$$\lambda = \exp(\beta 0 + \beta 1 xi)$$

λ is the average number of motor breakdowns we can estimate for a year in a factory, and we state that it depends on the number of verification test to be performed.

Description of the dataset

Given **x**, the number of verification tests performed by the engineers. In our case study, x is represented by a random sample from a discrete uniform distribution, with 1500 samples to draw, and range of distribution from 1 to 10. We use the rdunif() function in R from the tidyverse package to perform this task. The code is showed below:

x = rdunif (1500, 10, a=1)

Given y, the number of motor breakdowns during a year in a factory. We established **y** as a random sample from a discrete uniform distribution as well, with 1500 samples to draw, and range of distribution from 1 to 5:

y = rdunif(1500, 5, a=1)

Maximum Likelihood

Poisson regression assumes the response variable Y has a Poisson distribution, and assumes the logarithm of its expected value can be modeled by a linear combination of unknown parameters. The probability density function of the Poisson distribution is given by:

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$$

The Maximum likelihood estimation (MLE) is a method of estimating the beta parameters of a probability distribution by maximizing a likelihood function, so that under the assumed statistical model the observed data is most probable. The point in the parameter space that maximizes the likelihood function is called the maximum likelihood estimate. We can present the formula as:

$$\log L\,(\beta_0, \beta) = \sum(y_i(\beta 0 + \ \beta 1 xi) - \exp(\beta 0 + \ \beta 1 xi) - \log\,(yi!))$$

To fit a Poisson distribution to *x* we do not minimize the residual sum of squares, instead we maximize the likelihood by varying its parameter.

Results

For finding the optimal values that maximize the likelihood function, we can use the optim() function in R.

By default the optim() function searches for parameters, which minimize the function fn. In order to find a maximum, all we have to do is to change the sign of the function, hence the likelihood poisson function will return a result with -sum(…)

```
#Define the likelihood_poisson function
likelihood_poisson = function(param,dataset){

  b0 = param[1]
  b1 = param[2]

  x = dataset[[1]]
  y = dataset[[2]]

  return(-sum(y*(b0 + b1*x) - exp(b0+b1*x) - lfactorial(y)))

}

dataset = list(x,y)

#The paramaters will be optimized over 2 values
```

```
optimPoisson <-
optim(par=c(1.0,1.0),fn=likelihood_poisson,dataset=list(x,y),method="L-BFGS-B")

optimPoisson$par
```

After performance the optimization technique, the optimal values for β_0 & β_1 are:

```
[1] 1.081554458 0.001183291
```

We can go further by computing lamba:

```
#we set beta 0 and beta 1 final
b0_final = optimPoisson$par[1]
b1_final = optimPoisson$par[2]

#we compute lambda
lambda = mean(exp(b0_final + b1_final*x))

> print(lambda)
[1] 2.96866
```

We set β0 as the first value of the optimization function and β1 as the second. The output of lambda is *2.96866 ~ 3* which means that 3 motor breakdowns can occur in average in a facto ry.

References

Nasini, Stefano. Binomial Regression. April 2020.

- https://www.statology.org/mle-poisson-distribution/
- http://eric.univ-lyon2.fr/~ricco/cours/slides/regression_poisson.pdf
- https://data.princeton.edu/wws509/notes/c4.pdf
- https://www.magesblog.com/post/2013-03-12-how-to-use-optim-in-r/
- https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/optim
- https://www.dataquest.io/blog/tutorial-poisson-regression-in-r/