Negative binomial regression model report sample

Technology

Introduction

Data is normally collected, recorded and analyzed to be presented, published or stored over time. It is targeted to being used in the future for reference purposes or act as the basic, element and foundation of other research studies thereafter. For these reasons, the data collection, presentation and storage methods have revolutionized over the ears as the computer information and technology continues to advance, and many new innovations made each day. In the essay herewith, we present a detailed comparison of the negative binomial regression and the integer valued AR model for the dataset zeger. txt. The process will involve fitting the two models on the dataset provided, obtaining the maximum likelihood estimates of the two distributions and also obtaining the posterior distribution as required. The anova table will be obtained for the results to establish which model best fits this data.

Suppose that the observed data provided i. e x1, x2, x3,.. xn are independent and mutually independent for some values of p where p is the probability of a success in the rth trial such that
xi~NB(ri, p) where r1. rn are known and will be observed.

The probability density function will be defined as

Prxi= j: p= j-1ri-1 1-pj-riprifor j= ri, ri+1

Assuming a beta distribution prior which is proportional to:

∝pa-1*1-pb-1where 0

The posterior distribution of our data will be proportional to:

$\propto p^{\sum r_i + a - 1} * 1 - p^{\sum x_i + b - 1}$

This is proportional to a beta distribution with parameters $(\sum r_i + a, \sum x_i + b)$

Following the above fit, the p is estimated as 4. 2687e+138 i. e the mle of

p= 4. 2687e+138

Integer Valued AR Model

Essentially, the previous model comprised of binomial thinning terms.

However integer valued AR model of the new application is independent of

thinning in the new counts of data. The new Poisson method requires that all

values are observable (Hilbe, 2011). However to represent the new

occurrences and event in the economics portfolio which are in many cases

unobservable but need to be catered for all the same whether something

happens or does not happen.

The integer count models are explained as follows.

- Non negative integers should be represented by a series of counts in the

variables of x. The values will be x_0, x_1 and the value x_t for the time count.

Therefore, the equation to show this will be

$x_t = \alpha \circ x_{t-1} + e_t$

Where $\alpha \circ x_{t-1}$ represents survival of counts from the previous period and it

represents the arrival of new counts (Hilbe, 2011). The survival process is

generic and is described by a binomial process one count at a time (t -1)

survives to time t with a known probability $\alpha \in [0, 1]$ independent of all the

other counts this means that if x is the number of counts per time, then the

equation below will work.

$\alpha \circ x = x x_j = 1 y_j$

The well known INAR process is shown as

$\{x_t; t = 0, \pm 1, \pm 2, \}$ defined on the discrete support N0 by the equation

$x_t = \alpha \circ x_{t-1} + \varrho_t,$

Where $0 < \alpha < 1, \{\varrho_t\}$ is a sequence of independent and identically

distributed integer-valued random variables, with $E[\varrho_t] = \varrho$ and $Var[\varrho_t] = \sigma^2$

For the Bayesian methodology, it is assumed that future x_{n+n} observations

and the known parameters $\theta = (\alpha, \lambda)$ are as random. The following functions

explain it (Hilbe, 2011).

The inverse probability

$p(label| x, \theta) = f(x; \theta)$

The Prior Probability

$p(label| \theta)$

The Bayes Rule Function is as follows.

$p(label| x, \theta) = p(x| labe) \, p(label| \theta)/((p(x| L)p(L| \theta))$

$\alpha = 0.3053$

Determining the Model that best fits the Data

This is done by carrying out a model fit test. The two models are tested for

the dataset zeger. txt. The output is an analysis of the variance of the

parameters from the actual parameter. The model with the highest R

squared is selected or the model with the lowest Standard Error.

Since Negative binomial model has the smallest s. e = -5. 665e+79

compared to 0. 0748 of the INRA 1 model hence the negative binomial model

is selected for this study.

Conclusion

Models for analyzing data are progressively becoming sophisticated and the new and effective ways of research, data collection storage, analysis and presentation. New technology advances each day meaning that these are bound to more enhancements and complexities soon. The functions and derivative formulas have been explained in detail in the essay. The applications of both model is somewhat similar, but one is deemed as more effective than the other. As mentioned, the model with continuous cumulative functions, and keeps continuing other than having new data each and every time is quite efficient. The calculations and the deriving of formulas are quite tedious but technology that is up to date and enhanced has been ensured through computer systems.

In recap, industries such as insurance, accounting, epidemiology, research, earthquake counts which enable predictions via the study of seismic moves all depend on the same data acquisition and recording only with different outcomes, purposes, and fulfillments. Points to keep in mind should be patience and deficiency. All in all, as much as the processes will be quite costly and with major budget allocations to consider, it is quite worth the penny.

Bibliography

HILBE, J. M. (2011). Negative binomial regression. Cambridge, UK, Cambridge University

Press.

SIMONOFF, J. S. (2003). Analyzing categorical data. New York, NY [u. a.], Springer.

STUART, A. M., & HUMPHRIES, A. R. (1998). Dynamical systems and

numerical analysis.

Cambridge, Cambridge University Press.

APPENDIX

The graph is highly skewed to the right which implies the parameters cannot

be modeled with by non skewed distributions like a normal distribution.

The graph is highly skewed to the left which implies the parameters cannot

be modeled with by non skewed distributions like a normal distribution.

R Code

```
#Import data into R

library(fGarch)

setwd(" C:/Documents and Settings/Administrator/Desktop")

data= read. table(" zeger. txt", heade= F); data

#Extract the 1st column and call it x

x= data[, 1]; x

# Fitting the negative binomial model and simulating possible values of p

r= 1: length(x); r

j= r+1; j

p. r= choose(j-1, r-1)*(1-p)^(j-r)*p^r; p. r

hist(p. r)

#Estimating p

a= sum(p. r); a

mle. p. r= mean (p. r); mle. p. r

#Fitting the integer value AR and simulating possible values of p
```

model2= arima (x, c(1, 0, 0)); model2

hist (model2, col=" green", main=" Histogram of INRA 1")

model2= arima (x, c(1, 0, 0)); model2