Theory of Statistics Likelihood Assigment

Sean Soutar STRSEA001[a], Fabio Fehr FHRFAB001[b]

[a] *UCT Statistics Honours, Cape Town, South Africa*

[b] *UCT Statistics Honours, Cape Town, South Africa*

Abstract

This project will explore the Accidents dataset and try fit a Poisson, Negative Binomial, Mixture of 2 Poissons and zero inflated Poisson models to the data. The model with the strongest support will be chosen and discussed. Profile likelihoods and confidence intervals for the parameters will be found and displayed of the chosen model.

Keywords: Likelihood, Overdispersion, Soek

JEL classification

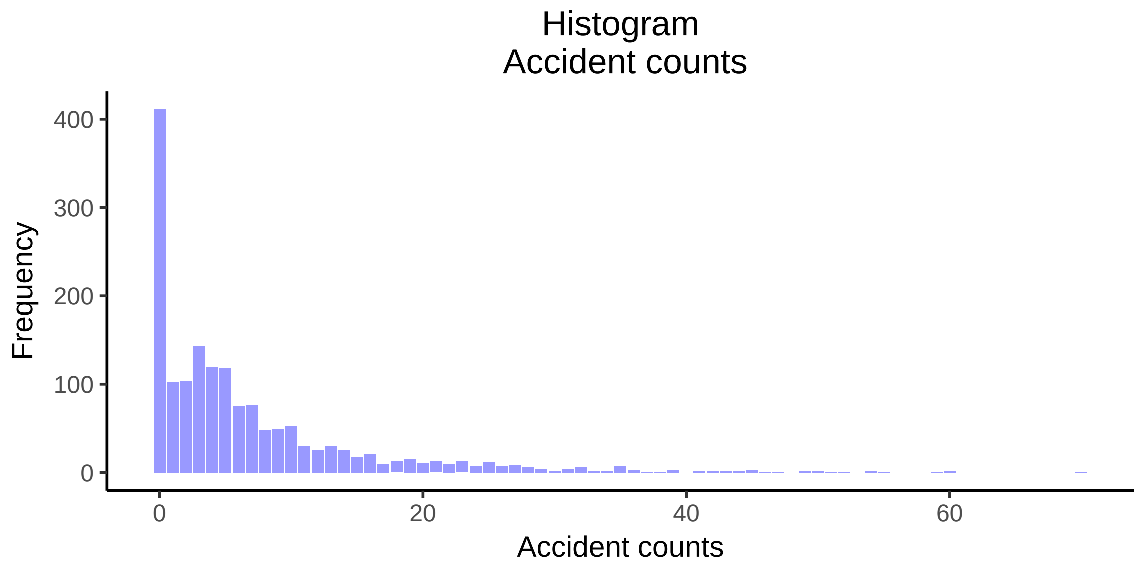
1. Introduction

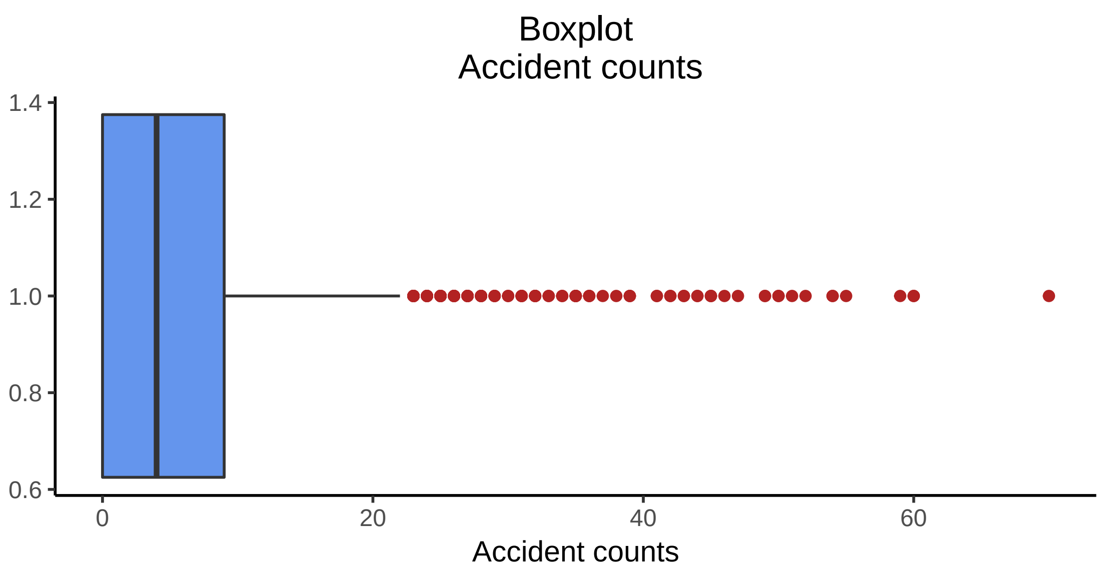
This assignment is an explorative report on a dataset containing the number of accidents on two-lane (same direction) road segments in Cape Town over a five-year period. The segments differ in length between 0.2 and 7.2 km. The aim of the report is to find and fit a model which accurately describes the accident dataset. This report will first explore the data then fit different adequate distributions and choose the most appropriate one. Once a model has been selected the profile likelihood and confidence intervals will be programmed and calculated from from first principles. The results will then be analysed critically and conclusions will be made and consider further considerations in the study.

1.1. Exploratory data analysis

To better understand our data this report shall explore the following properties; Firstly we examine the type of data within the accidents dataset and discuss whether our data is discrete ordinal or continuous. After the symmetry of the data and bounds will be discussed. This leads the exploration to outliers and extreme values.

Email addresses: `sean.soutar@gmail.com` (Sean Soutar STRSEA001), `FHRFAB001@myuct.ac.za` (Fabio Fehr FHRFAB001)





1.1.1. Data type

There are many instances where zero accidents were observed. This accounts for approximately 25.18% of the data. This suggests that the zero-inflated Poisson should be considered as this proportion is much higher than what would be expected of a regular Poisson distribution. The accident counts are discrete random variables. Specifically, they are discrete positive definite random variables on the interval $R \in \{0; +\infty\}$. Summary statistics of the data are shown below.

Table 1.1: Summary statisitics

Mean	Variance	Median
6.9179	85.0858	4

Mean	Variance	Median

In the Poisson distribution, the mean should equal the variance. The sample variance far exceeds the sample mean. This indicates overdispersion if the Poisson distribution were to be used. This is when the observations are more variable than what would be expected. This suggests that alternative count models and mixture distributions should be used.

1.1.2. Symmetry

This property is visually seen in the histogram and boxplot. All counts are greater than zero with a median value of 4 accidents. The largest accident observed is 70 accidents. THe histogram shows that the data are non-symetrical and positively skewed which is usually expected of count data.

1.1.3. Outliers

From the boxplot it clear that many outliers exist. One common method of classifying a point as an extreme value or outlier is if it falls more than 1.5 times the inner-quartile range above the upper quartile. The proportion of outliers within our data set amount to 15.26%.

2. Methods

2.1. Model Formulation

The data is discrete, asymmetric, positive definite, contains many positive outliers and many zeros. This would suggest distributions such as Poisson, Negative Binomial, mixture distribution of 2 Poissons and a zero inflated Poisson.

2.2. Akiake Information Coefficient (AIC)

The AIC metric can be used to compare models from different families of distributions. They can be used to compare relative goodness of fit between models. Overfitting the model with too many parameter is penalised by an increased AIC, thus a lower AIC value indicates a better fitting model.

$$\text{AIC} = -2l(\hat{\theta}) + 2\text{p}$$
$$p = \text{Number of estimated parameters}$$

2.3. *Bayesian Information Criterion (BIC)*

This metric, like AIC, also compares relative goodness of fit between models but penalises complex models when the sample size is large. BIC tends to produce simpler models than AIC.

$$\text{BIC} = -2l(\hat{\theta}) + log(n)$$
$$n = \text{Number of observations}$$

2.4. *Model Distributions*

2.4.1. *Poisson*

The Poisson is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant rate and independently of the time since the last event.This is typically used in count data where your mean and variance are equal.

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in \{0, 1, \ldots, \infty\}, \lambda > 0$$

$$L(\lambda|x) = \prod_{i=1}^{n} p(x_i)$$

$$L(\lambda|x) = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

$$l(\lambda|x) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right) \ln\lambda - \sum_{i=1}^{n} \ln(x_i!)$$

The Poisson is characterised by the λ parameter which denotes the population average rate of event occurence. In this context it would be the average number of accidents per unit time frame. Due to the variance being far higher than our mean in our data we would expect that a Poisson distribution would not fit very well.

Table 2.1: Poisson MLE's & information metrics

$\hat{\lambda}$	AIC	BIC
6.9179	20263.64	20269.04

2.4.2. Negative Binomial

The Negative Binomial distribution is a discrete probability function of the number of successes in a sequence of independant and identitically distributed Bernoulli trials. The parameters p & r measure the probability of success in an individual trial and the number of successes until r failures occure. The mean in a negative binomial is defined as $m = \frac{pr}{1-p}$ thus we can reparamatize the distribution in terms of the mean parameter m and shape parameter r such that $p = \frac{m}{m+r}$ and $1 - p = \frac{r}{m+r}$. We can also manipulate the constant term as follows.

$$\binom{x+r-1}{x} = \frac{(x+r-1)(x+r-2)\cdots(r)}{x!} = \frac{\Gamma(x+r)}{x!\,\Gamma(r)}$$

This gives us the negative binomial in the form

$$p(x) = \frac{\Gamma(r+x)}{x!\,\Gamma(r)} \left(\frac{m}{r+m}\right)^x \left(\frac{r}{r+m}\right)^r \quad \text{for } x = 0, 1, 2, \ldots$$

$$L(m, r|x) = \prod_{i=1}^{n} p(x_i)$$

$$L(m, r|x) = [\frac{1}{\Gamma(r)}]^n \prod_{i=1}^{n} \frac{\Gamma(r+x_i)}{x_i!} (\frac{m}{r+m})^{\Sigma_{i=1}^{n} x_i} (\frac{r}{r+m})^{nr}$$

$$l(m, r|x) = -n\ln[\Gamma(r)] + \sum_{i=1}^{n} \ln(\Gamma(r+x_i)) - \sum_{i=1}^{n} \ln x_i! + \sum_{i=1}^{n} x_i \ln(\frac{m}{r+m}) + nr\ln(\frac{r}{r+m})$$

It is important to note that the variance of a Negative Binomial under this parameterisation is $m + \frac{m^2}{r}$ and always larger than our mean m. This would suggest a better fit than our Poisson model. Shape parameters are often regarded as nuisance parameters and do not play a meaningful role in maximising likelihood. Therefore, since we desire no under or over dispersion, we can express the shape parameter as a function of the mean parameter to be esimated and the sample variance.

$$Var(x) = m + \frac{m^2}{r}$$

$$r = \frac{m^2}{Var(x) - m}$$

$$\hat{r} = \frac{\hat{m}^2}{S^2 - \hat{m}}$$

Table 2.2: Negative Binomial MLE's & information metrics

Mean $\hat{m}$	Shape $\hat{r}$	AIC	BIC
6.8108	0.5926	9626.92	9630.31

2.4.3. Mixture of 2 Poissons

A finite mixture distributon of two Poisson variables will now be explored. A possibly reason for the overdispersion is that the data are from two separate Poisson distributions. Since it is not known from which distribution that any given data point is from, presuming that the two distribution mixture is appropriate, an additional mixing proportion parameter p needs to be estimated.

$$p(x|\lambda_1, \lambda_2, p) = p\frac{e^{-\lambda_1}\lambda_1^x}{x!} + (1-p)\frac{e^{-\lambda_2}\lambda_2^x}{x!}, \quad x \in \{0, 1, \ldots, \infty\}, \lambda_1, \lambda_2, p > 0$$

$$L(\lambda_1, \lambda_2, p|x) = \prod_{i=1}^{n} p(x_i)$$

$$L(\lambda_1, \lambda_2, p|x) = \prod_{i=1}^{n} p\frac{e^{-\lambda_1}\lambda_1^{x_i}}{x_i!} + (1-p)\frac{e^{-\lambda_2}\lambda_2^{x_i}}{x_i!}$$

$$l(\lambda_1, \lambda_2, p|x) \sum_{i=1}^{n} \ln[p\frac{e^{-\lambda_1}\lambda_1^{x_i}}{x_i!} + (1-p)\frac{e^{-\lambda_2}\lambda_2^{x_i}}{x_i!}]$$

The parameters λ_1 and λ_2 refer to the average number of road accidents per road stretch for the first and second distribution respectively. The parameter p is the proportion parameter. This represents the probability that a given observation belongs to distribution 1. Therefore, the probability that an observation belongs to distribution 2 is the $1 - p$.

Table 2.3: Poisson-Poisson MLE's & information metrics

$\hat{\lambda_1}$	$\hat{\lambda_2}$	Proportion p	AIC	BIC
19.3809	2.8407	0.2465	12075.12	12076.52

2.4.4. Zero inflated Poisson

The Zero Inflated Poisson is also a finite mixture distribution. This model supposes that the data can come from two distributions. The one is a Zero Process and the other is a Poisson process that can only take on non-zero values. This model is useful if there are many zeroes in the data. This was seen to be the case as discussed in 1.1.1. The Zero Inflated Poisson is a piecewise defined distribution with different mass functions for predicting the probability that a given observation will be zero rather than non-zero.

$$p(x_i = 0) = \pi + (1 - \pi)e^{-\lambda}$$

$$p(x_i \neq 0) = (1 - \pi)\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}, \qquad x_i \geq 1$$

$$L(\lambda, \pi|x) = L(\lambda, \pi|x = 0)L(\lambda, \pi|x \neq 0)$$

An indicator variable I is defined.

$$I = \begin{cases} 0 & x = 0 \\ 1 & x \neq 0 \end{cases}$$

$$L(\lambda, \pi | x) = \prod_{i=1}^{n} p(x_i = 0)^{1-I} p(x_i \neq 0)^I$$

$$L(\lambda, \pi | x) = \prod_{i=1}^{n} [\pi + (1-\pi)e^{-\lambda}]^{1-I} [(1-\pi)\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}]^I$$

$$l(\lambda, \pi | x) = \sum_{i=1}^{n} \ln[(1-I)[\pi + (1-\pi)e^{-\lambda}] + I[(1-\pi)\frac{\lambda^{x_i} e^{-\lambda}}{x_i!}]]$$

The parameter λ is the average rate of accidents per road stretch. The parameter π is the probability of additional zeroes observed in our data. This mixture distribution has a mean of $(1-\pi)\lambda$ and variance $(1-\pi)\lambda(1+\pi\lambda)$. It is clear that the variation in this distribution is always greater than the average thus a good potential model to consider.

Table 2.4: Zero inflated Poisson MLE's & information metrics

$\hat{\lambda}$	$\hat{\pi}$	AIC	BIC
9.2456	0.2518	15556.16	15559.56

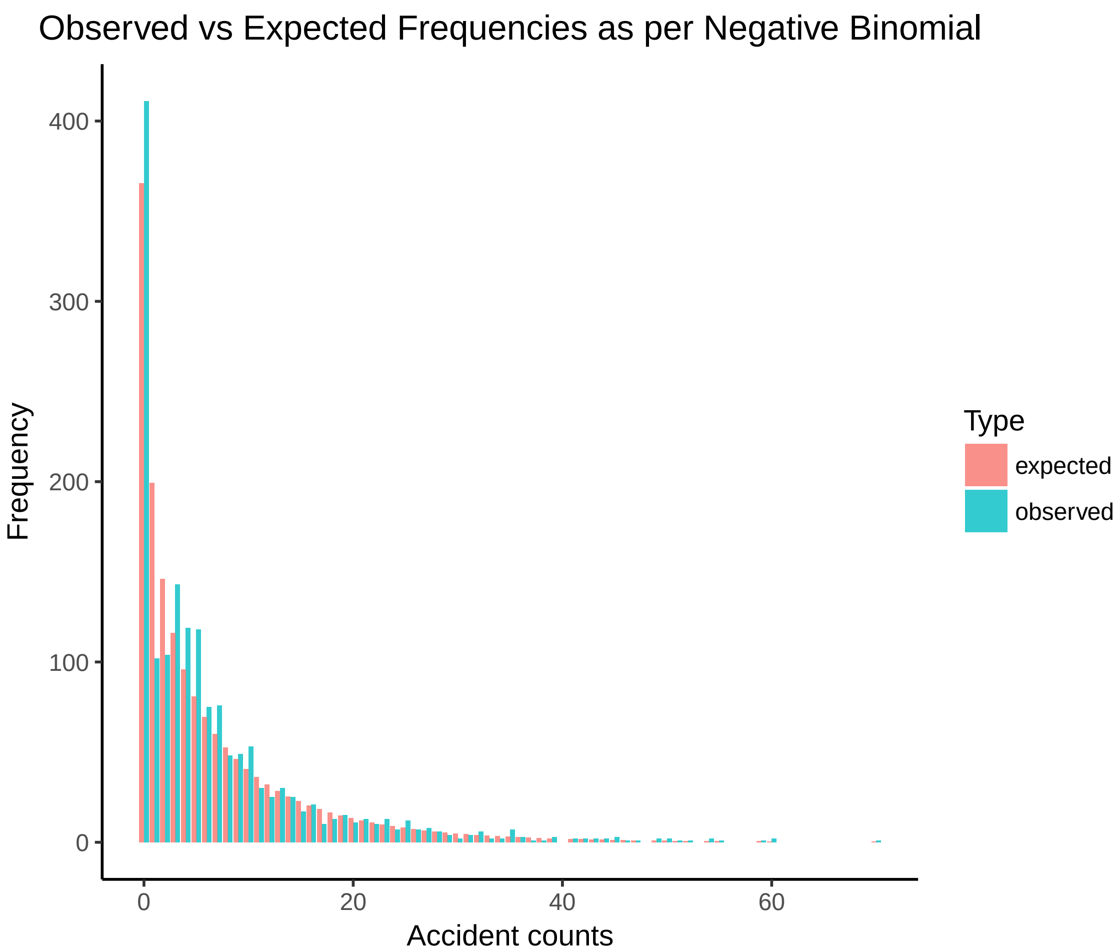
2.5. Model Selection

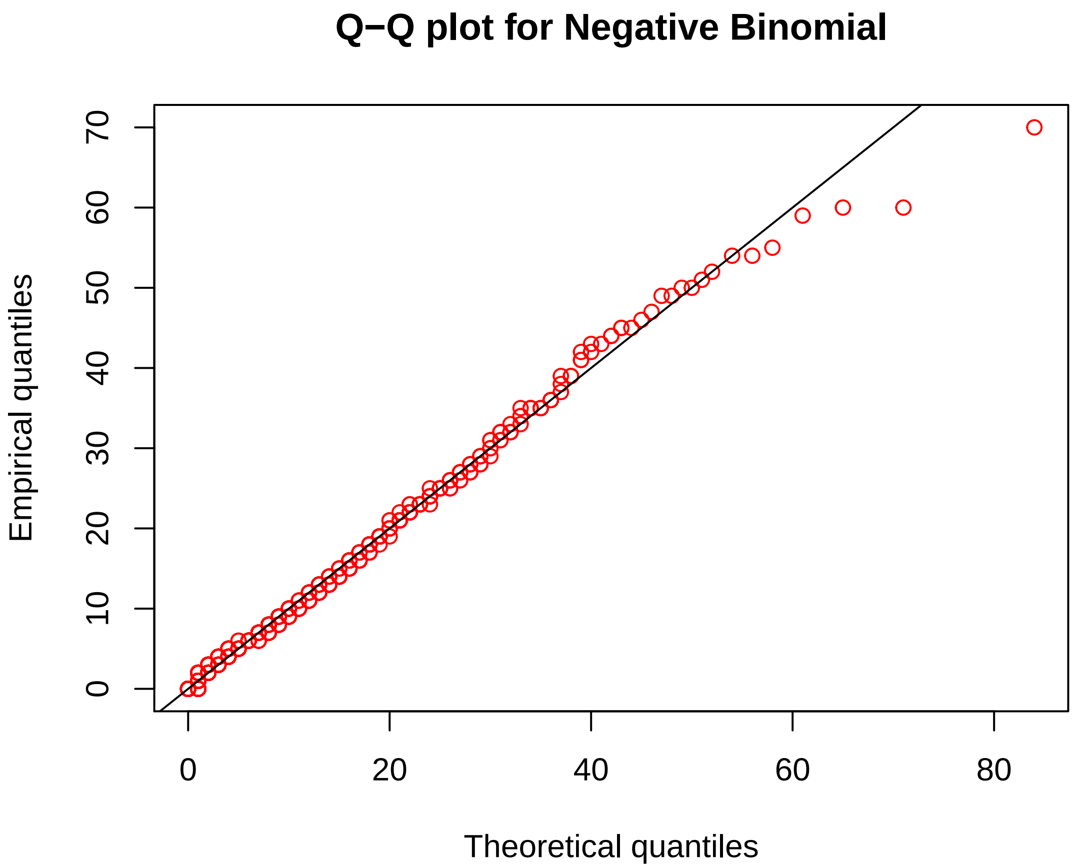
The AIC and BIC results for each model are summarised below.

Table 2.5: Models fitted to accident data

Model	AIC	BIC
Poisson	20263.64	20269.04
Negative Binomial	**9626.92**	**9630.31**
Poisson Mixture	12075.12	12076.52
Zero Inflated Poisson	15556.16	15559.56

The Negative Binomial has the lowest AIC and BIC value at 9626.92 and 9630.31 respectively. This implies that the Negative Binomial is the best fitting model when compared to the others. The goodness of fit will now be assessed further.

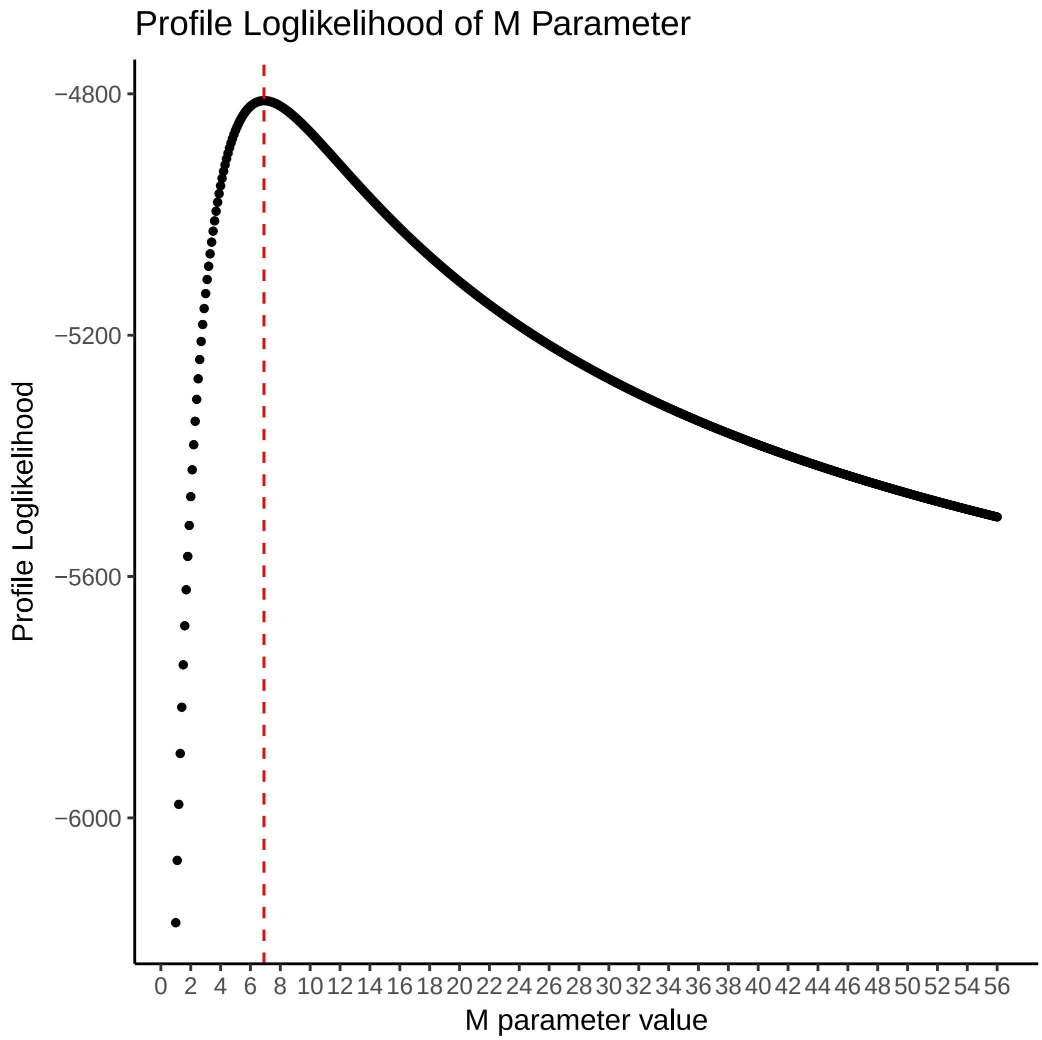

Observed vs Expected Frequencies as per Negative Binomial
Frequency
400
300
200
100
0
0
20
40
60
Accident counts
Type
expected
observed

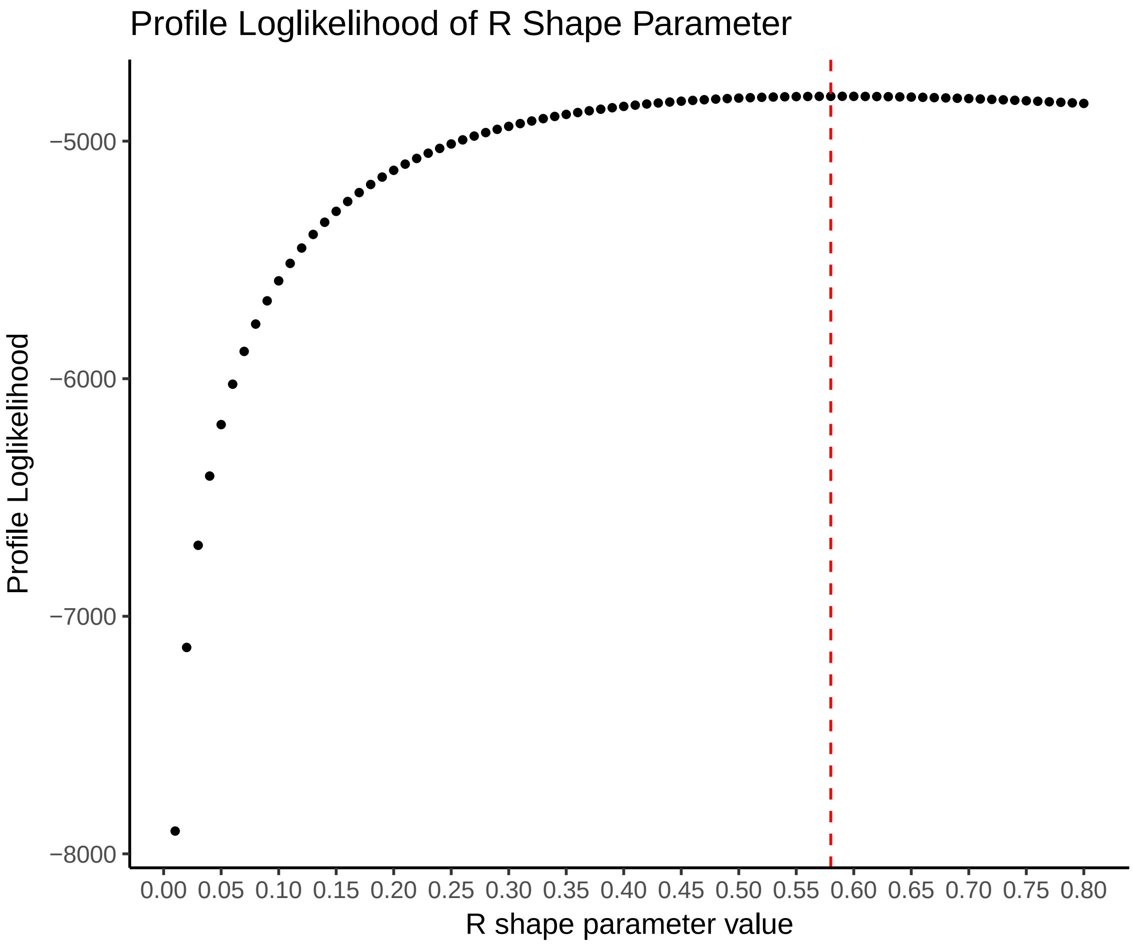



```
##
##  Pearson's Chi-squared test
##
## data:  filter(comparison_df, Type == "observed")$Frequency and filter(comparison_df, Type !=
## X-squared = 1456, df = 1430, p-value = 0.3101
```

2.6. Profile Likelihood & Confidence Intervals

The profile likelihood is a techinque used to estimate the likelihood function of a single parameter when multiple parameters are estimated simultaenously. The profile likelihoods for the m and r parameters are calculated as follows:


Profile Loglikelihood of M Parameter
Profile Loglikelihood
−4800
−5200
−5600
−6000
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56
M parameter value



It can be seen that the profile likelihood for the M parameter is somewhat quadratic , but this is clearly not the case for the R shape parameter. Therefore it would be inappropriate to use a Wilks Likelihood interval or Wald interval for the shape parameter. The direct likelihood interval should be used. However, the quadracity of the M parameter profile likelihood means that asymptotic intervals relying on quadracity could be used.

2.6.1. Wald Interval

We can attain an interval for out Maximum Likelihood Estimates if we assume they have asymptotic normal distribution. This means $n \to \infty$ the estimates will be approximately normal with the following parameters.

$$\hat{r} \dot{\sim} N(r, I(r)^{-1})\hat{m} \dot{\sim} N(m, I(m)^{-1})$$

This means that we can form asymptotic confidence intervals for m and r know as a Wald interval.

The advantages of this is it is simple to calculate provide you have a standard error estimate.

2.6.2. Wilks Likelihood ratio

The likelihood interval can be found as $R(\theta) > \gamma^p$, where γ can be based on a χ^2 approximation and p is the dimension of θ. Equivalently we can use the deviance:

$$-2[\ell(\theta_p) - \ell(\hat{\theta_p})] \sim \chi^2_p.$$

We solve for points of θ where the deviance equals the 95th percentile of χ^2_p.

If the assumption of multivariate normality (likelihood is quadratic in all dimensions) is not met, it is better to use the likelihood function. For a likelihood interval we would choose all θ where

$$W = 2\text{log}R(\theta) \sim \chi^2_p$$

where p is the dimension of θ. This translates to all values of θ where

$$\ell(\theta) > \ell(\hat{\theta}) + p \log(\gamma)$$

3. Results

<<<<<<< Updated upstream

======= -Are parameter estimates correlated? >>>>>>> Stashed changes

4. Conclusion

-What are the next steps and how can we improve the models

5. References