IMS Inleveropdracht 2

2023-10-19

Introduction

The goal of this statistical experiment is to determine whether the gas extraction from the Groningen gas field has some influence on the number of earthquakes in the Netherlands. This is accomplished by finding two model for a data set containing all earthquakes with a magnitude of 3.0 or larger on the Richter scale, where earthquakes that happen within 3 days of each other count as one. One model will contain all the data from the years before the gas extraction (1900-1962), the other model will contain all the data from the years after (1963-2022). The goal is to find a ML-estimator for both models, where the same family of distribution is used for both models. After that a mean squared error will be computed, and finally a confidence interval will be given.

Organizing the data

The available data is first organised so it can be used to calculate with.

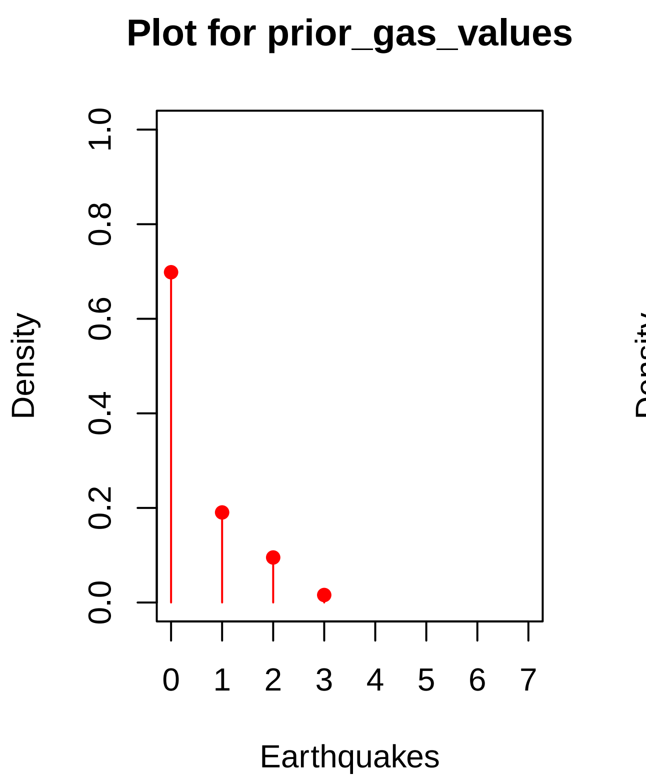
```
#setwd("C:\\Users\\Sam\\Desktop\\Coding\\Rstudio")
load("Earthquakes2.Rdata") #loading the data
sep_data <- stack(Data) #seperates the data from our data file
all_values <- sep_data$values #seperates just the values into a vector
prior_gas_values <- all_values[1:63] #data from the years prior to gas extraction
post_gas_values <- all_values[64:123] #data from the years after gas extraction began
```

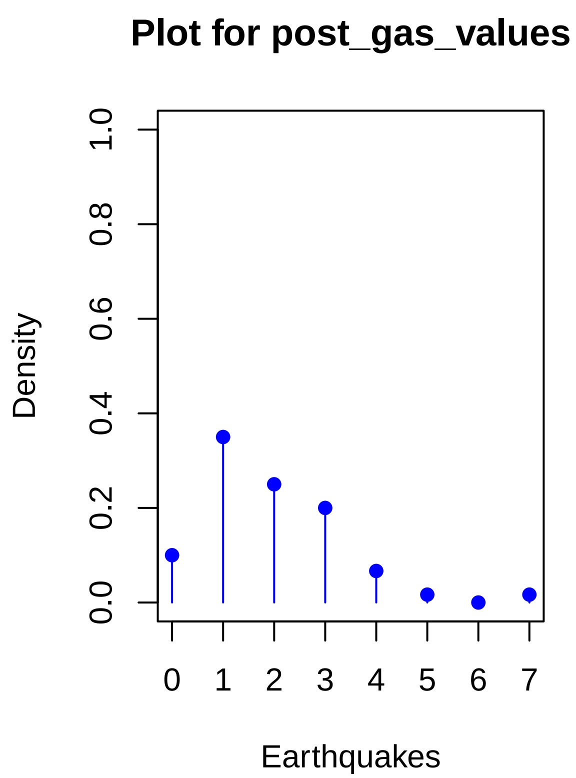
The point graphs

A point graph for both data sets with the x-axis being all the possible values (in our case 0-7) and the y-axis being the density of every point is created to visually establish what family of distributions could be used.

```
par(mfrow=c(1, 2))
#First we need to know some information, like the maximum amount of earthquakes to plot
max_amount = max(prior_gas_values, post_gas_values)

#we generate a table from the prior_gas_values vector
prior_freqs_table <-
  table(factor(prior_gas_values,
               levels = 0:max(prior_gas_values)))/length(prior_gas_values)
prior_freqs <- stack(prior_freqs_table)$values #this will be our y-axis vector
prior_x <- c(0:(length(prior_freqs)-1)) #this will be our x-axis vector

#we generate a table from the post_gas_values vector
post_freqs_table <-
  table(factor(post_gas_values,
               levels = 0:max(post_gas_values)))/length(post_gas_values)
post_freqs <- stack(post_freqs_table)$values #this will be our y-axis vector
post_x <- c(0:(length(post_freqs)-1)) #this will be our x-axis vector

# We create a blank plot
```

```r
plot(x = 1,
     type = "n",
     xlim = c(0, max_amount),
     ylim = c(0, 1),
     pch = 16,
     xlab = "Earthquakes",
     ylab = "Density",
     main = "Plot for prior_gas_values")
#Now we plot the points over it
points(prior_x, prior_freqs, pch = 16, lwd = 1, col = "red")
segments(x0 = prior_x, y0 = 0, x1 = prior_x, y1 = prior_freqs, col = "red")

#We create another blank plot
plot(x = 1,
     type = "n",
     xlim = c(0, max_amount),
     ylim = c(0, 1),
     pch = 16,
     xlab = "Earthquakes",
     ylab = "Density",
     main = "Plot for post_gas_values")
#And we plot the points over it
points(post_x, post_freqs, pch = 16, lwd = 1, col = "blue")
segments(x0 = post_x, y0 = 0, x1 = post_x, y1 = post_freqs, col = "blue")
```





The chosen distribution

reason why we choose the poisson distribution

Determining the ML-estimators

latex stuff

Calculating the mean:

```
#(could also be done with the mean function)
mean_prior <- sum(prior_gas_values)/63
mean_post <- sum(post_gas_values)/63
#printing the means to use in our calculations
print(mean_prior)
```

```
## [1] 0.4285714
```

```
print(mean_post)
```

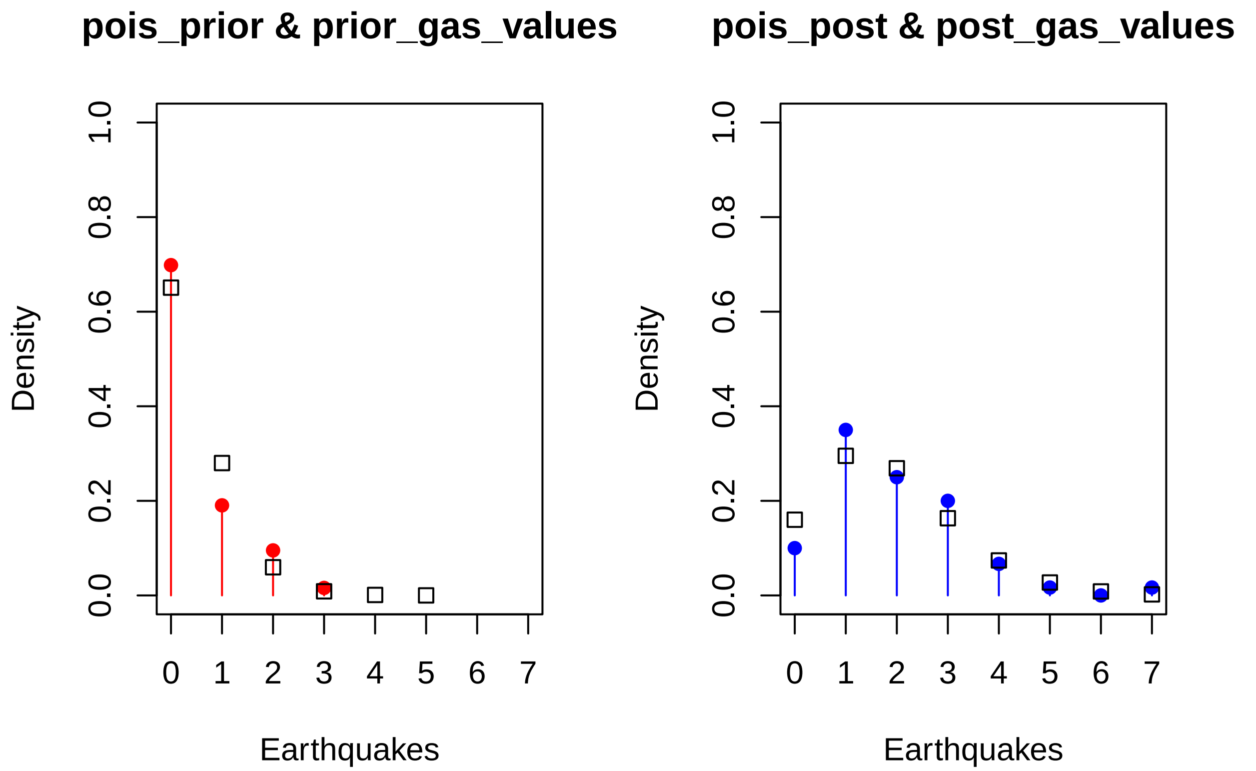
```
## [1] 1.825397
```

latex stuff

Updating the point graphs

```
par(mfrow=c(1, 2))
#we plot the Poisson distribution over the previous plots.
#First we establish the points for both Poisson distributions
pois_prior_values <- rpois(100000, 0.4285714)
pois_prior_freqs_table <-
  table(factor(pois_prior_values,
               levels = 0:max(pois_prior_values)))/length(pois_prior_values)
pois_prior_freqs <- stack(pois_prior_freqs_table)$values
pois_prior_x <- c(0:(length(pois_prior_freqs)-1))

pois_post_values <- rpois(100000, 1.825397)
pois_post_freqs_table <-
  table(factor(pois_post_values,
               levels = 0:max(pois_post_values)))/length(pois_post_values)
pois_post_freqs <- stack(pois_post_freqs_table)$values
pois_post_x <- c(0:(length(pois_post_freqs)-1))

#Create another blank plot for our prior values and the Poisson values
plot(x = 1,
     type = "n",
     xlim = c(0, max_amount),
     ylim = c(0, 1),
     pch = 16,
     xlab = "Earthquakes",
     ylab = "Density",
     main = "pois_prior & prior_gas_values")
points(prior_x, prior_freqs, pch = 16, lwd = 1, col = "red")
segments(x0 = prior_x, y0 = 0, x1 = prior_x, y1 = prior_freqs, col = "red")
points(pois_prior_x, pois_prior_freqs, pch = 0, lwd = 1, col = "black")

#Create another blank plot for our post values and the Poisson values
plot(x = 1,
     type = "n",
     xlim = c(0, max_amount),
```

```r
     ylim = c(0, 1),
     pch = 16,
     xlab = "Earthquakes",
     ylab = "Density",
     main = "pois_post & post_gas_values")
points(post_x, post_freqs, pch = 16, lwd = 1, col = "blue")
segments(x0 = post_x, y0 = 0, x1 = post_x, y1 = post_freqs, col = "blue")
points(pois_post_x, pois_post_freqs, pch = 0, lwd = 1, col = "black")
```



Computing the mean squared errors

The difference between the prior frequencies and the poisson prior frequencies is squared for all available points. Then the mean is taken which results in the mean squared error for the model. The same is done for the post gast extraction data.

```r
difference_prior <- ( prior_freqs - pois_prior_freqs[1:length(prior_freqs)] )^2
mse_prior <- mean(difference_prior)
print(mse_prior)
```

```
## [1] 0.002897977
```

```r
difference_post <- ( post_freqs - pois_post_freqs[1:length(post_freqs)] )^2
mse_post <- mean(difference_post)
print(mse_post)
```

```
## [1] 0.001096017
```

calculating the confidence intervals

"Construct also two-sided asymptotic 0.95 confidence intervals for your parameters" In this section the two-sided asymptotic 0.95 confidence interval for both of our parameters is calculated. ***latex stuff***

```
#final calculation?
```

Final conclusion

"What is your conclusion about the influence of the gas extraction?"