Modelling Insurance Claims

Shashi Bhushan Singh

Submission Nov 16-2022 (Wednesday)

Problem 4: Modelling Insurance Claims

Consider the `Insurance` datasets in the `MASS` package. The data given in data frame `Insurance` consist of the numbers of policyholders of an insurance company who were exposed to risk, and the numbers of car insurance claims made by those policyholders in the third quarter of 1973.

This data frame contains the following columns:

`District` (factor): district of residence of policyholder (1 to 4): 4 is major cities.

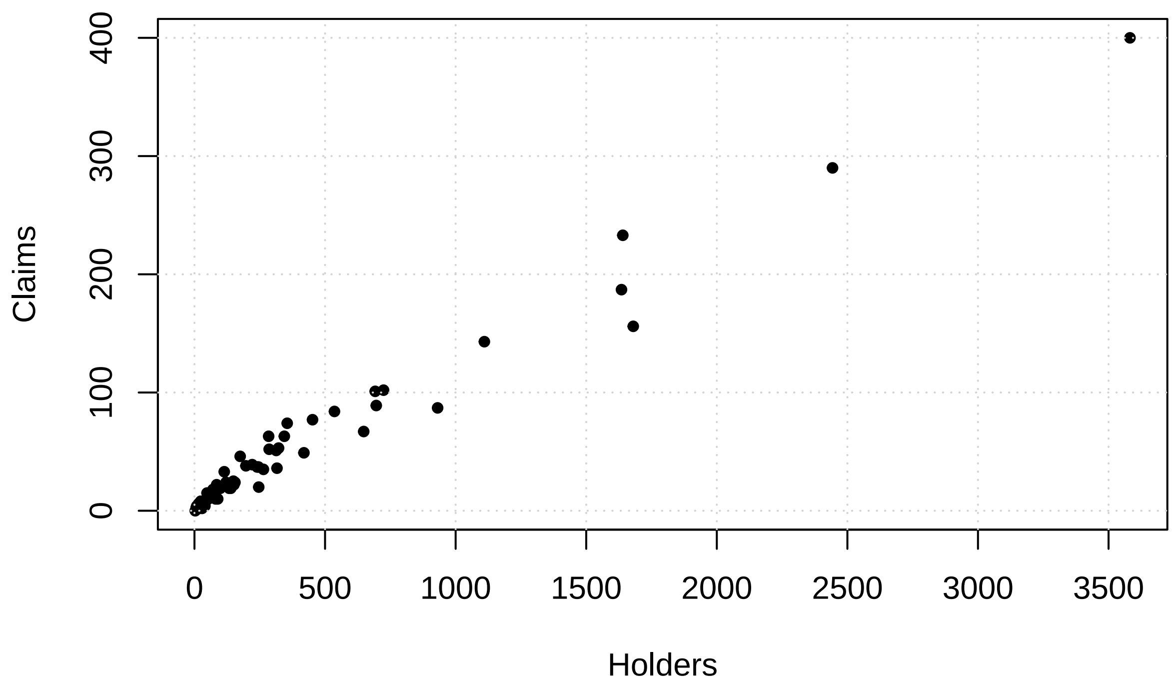
`Group` (an ordered factor): group of car with levels <1 litre, 1–1.5 litre, 1.5–2 litre, >2 litre.

`Age` (an ordered factor): the age of the insured in 4 groups labelled <25, 25–29, 30–35, >35.

`Holders` : numbers of policyholders.

`Claims` : numbers of claims

```
library(MASS)
plot(Insurance$Holders,Insurance$Claims
     ,xlab = 'Holders',ylab='Claims',pch=20)
grid()
```



Note: If you use built-in function like `lm` or any packages then no points will be awarded.

Part A: We want to predict the `Claims` as function of `Holders`. So we want to fit the following models:

$$\texttt{Claims}_i = \beta_0 + \beta_1 \ \texttt{Holders}_i + \varepsilon_i, \quad i = 1, 2, \cdots, n$$

Assume : $\varepsilon_i \sim N(0, \sigma^2)$. Note that $\beta_0, \beta_1 \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$.

The above model can alse be re-expressed as,

$$\texttt{Claims}_i \sim N(\mu_i, \sigma^2), \quad where$$

$$\mu_i = \beta_0 + \beta_1 \ \texttt{Holders}_i + \varepsilon_i, \quad i = 1, 2, \cdots, n$$

(i) Clearly write down the negative-log-likelihood function in `R`. Then use `optim` function to estimate MLE of $\theta = (\beta_0, \beta_1, \sigma)$

(ii) Calculate **Bayesian Information Criterion** (BIC) for the model.

```
modelA =function(parm){

  l=0
  for (i in 1:nrow(Insurance)) {
    m=parm[1]+parm[2]*Insurance$Holders[i]
    l=l+dnorm(Insurance$Claims[i], mean=m, sd=exp(parm[3]), log=T) }

  return (-l)
}
```

```
p_A=optim(c(0,0,1),modelA)
cat("MLE : b0:", p_A$par[1], "b1:", p_A$par[2], "b3:", exp(p_A$par[3]), " ")
```

```
## MLE : b0: 8.092468 b1: 0.1126364 b3: 11.85255
```

```
# Bayesian Information criterion= k*ln(n) -2 ln(L`)
BIC_A = 3 * log(nrow(Insurance)) + 2 *p_A$value

cat("BIC :", BIC_A)
```

```
## BIC : 510.7594
```

Part B: Now we want to fit the same model with change in distribution:

$$\texttt{Claims}_i = \beta_0 + \beta_1 \ \texttt{Holders}_i + \varepsilon_i, \quad i = 1, 2, \cdots, n$$

Assume : $\varepsilon_i \sim Laplace(0, \sigma^2)$. Note that $\beta_0, \beta_1 \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$.

(i) Clearly write down the negative-log-likelihood function in `R`. Then use `optim` function to estimate MLE of $\theta = (\beta_0, \beta_1, \sigma)$

```
dlap=function(x,m,b) {
  return( log( 1/(2*b) * exp(-abs(x-m)/b)

  ))
}


modelB =function(parm){

  l=0
  for (i in 1:nrow(Insurance)) {
    m=parm[1]+parm[2]*Insurance$Holders[i]
   l=l+dlap(Insurance$Claims[i], m, exp(parm[3]))

     }

  return (-l)
}



p_B=optim(c(0,0,1),modelB)
cat("MLE : b0:", p_B$par[1], "b1:", p_B$par[2], "b3:", exp(p_B$par[3]), " ")
```

```
## MLE : b0: 5.084757 b1: 0.1166252 b3: 8.203108
```

```
BIC_B = 3 * log(nrow(Insurance)) + 2 *p_B$value

cat("BIC :", BIC_B)
```

```
## BIC : 498.687
```

(i) Calculate **Bayesian Information Criterion** (BIC) for the model.

Part C: We want to fit the following models:

$$\texttt{Claims}_i \sim LogNormal(\mu_i, \sigma^2), where$$

$$\mu_i = \beta_0 + \beta_1 \log(\texttt{Holders}_i), \quad i = 1, 2, ..., n$$

Note that $\beta_0, \beta_1 \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$.

(i) Clearly write down the negative-log-likelihood function in R. Then use `optim` function to estimate MLE of $\theta = (\alpha, \beta, \sigma)$

(ii) Calculate **Bayesian Information Criterion** (BIC) for the model.

```
Insurance1=Insurance[-61,]

modelc =function(parm){

  l=0
  for (i in 1:nrow(Insurance1)) {
    mu=parm[1]+parm[2]*log(Insurance1$Holders[i])


   l=l+dlnorm(Insurance1$Claims[i], meanlog=mu, sdlog= exp(parm[3]),log=T)

     }

  return (-l)
}


p_C=optim(c(2,2,1),modelc)
cat("MLE : b0:", p_C$par[1], "b1:", p_C$par[2], "b3:", exp(p_C$par[3]), " ")
```

```
## MLE : b0: -1.024094 b1: 0.8478622 b3: 0.3292471
```

```
BIC_C = 3 * log(nrow(Insurance1)) + 2 *p_C$value

cat("BIC :", BIC_C)
```

```
## BIC : 452.6034
```

Part D: We want to fit the following models:

$$\texttt{Claims}_i \sim Gamma(\alpha_i, \sigma), where$$

$$log(\alpha_i) = \beta_0 + \beta_1 \log(\texttt{Holders}_i), \quad i = 1, 2, ..., n$$

(iii) Compare the BIC of all three models

```
modeld =function(parm){

  l=0
  for (i in 1:nrow(Insurance1)) {
    m=exp(parm[1]+parm[2]*log(Insurance1$Holders[i]))
   l=l+dgamma(Insurance1$Claims[i], m, exp(parm[3]),log=T)

     }

  return (-l)
}

Insurance1=Insurance[-61,]

p_D=optim(c(1,1,2),modeld)
cat("MLE : b0:", p_D$par[1], "b1:", p_D$par[2], "b3:", exp(p_D$par[3]), " ")
```

```
## MLE : b0: -1.642424 b1: 0.8370277 b3: 0.4859753
```

```
BIC_D = 3 * log(nrow(Insurance1)) + 2 *p_D$value

cat("BIC :", BIC_D)
```

```
## BIC : 437.3382
```

```
BIC_model= data.frame("Model" = c("A","B","C","D"), "Bic_val"= c(BIC_A,BIC_B,BIC_C,BIC_D))

bst=BIC_model[BIC_model$Bic_val==min(BIC_model$Bic_val),]
print(paste("The best model is ",bst[,1], " with BIC value ", bst[,2]))
```

```
## [1] "The best model is  D  with BIC value  437.338200746668"
```