Assignment 2

Problem 4

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.

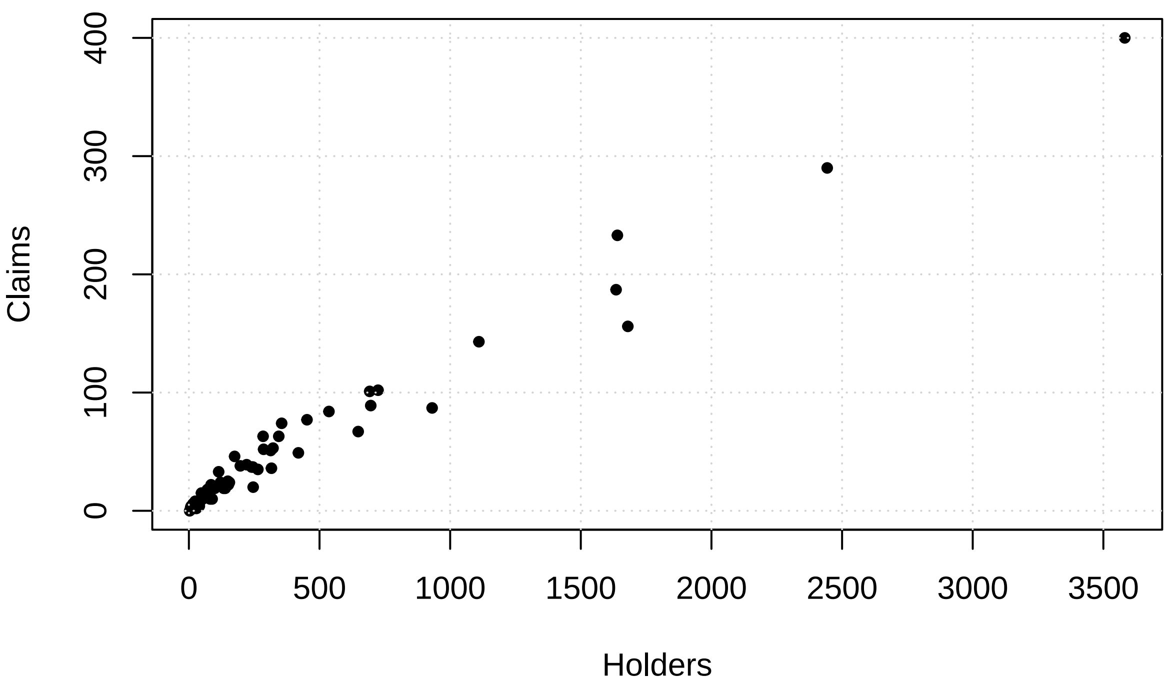
`Group` (an ordered factor): group of car with levels <1 litre, 1–1.5 litre, 1.5–2 litre, >2 litre.

`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.

We will write functions for calculation of MLE and BIC for each part here.

```r
Holders = Insurance$Holders
Claims = Insurance$Claims

#MLE and BIC functions

MyMLE <- function(func,x,initial){
  func_one<- function(p){
    func(p,x)
  }
  optim(par = initial, func_one) #We integrate the optim function here
}

MyBIC <- function(MLEv, parm, data){
  N <- length(data)
  (length(parm))*log(N)-2*(MLEv$value)
}
```

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 $\theta = (\beta_0, \beta_1, \sigma)$

```r
#Part A

ll_A <- function(parm, data=list(Claims,Holders)){
  beta0 <- parm[1]
  beta1 <- parm[2]
  sigma <- exp(parm[3])
  N <- length(data)
  -sum(log(dnorm((data$Claims - beta0 - beta1*data$Holders),0,sigma)))
}

MLEvA <- MyMLE(ll_A, Insurance, c(100,100,50))
cat("MLE for model A :", c(MLEvA$par[1], MLEvA$par[2], exp(MLEvA$par[3])))
```

```
## MLE for model A : 8.128316 0.1126432 11.87061
```

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

```r
b_A <- MyBIC(MLEvA, MLEvA$par, Insurance)
cat("BIC value for model A :", b_A)
```

```
## BIC value for model A : -493.4538
```

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 $\theta = (\beta_0, \beta_1, \sigma)$

```
#Part B

dLaplace <- function(x, mu=0, b=1, params=list(mu, b),...){
  if(!missing(params)){
    mu <- params$mu
    b <- params$b
  }
  d <- exp(-abs(x-mu)/b) / (2*b)
}

ll_B <- function(parm, data=list(Claims,Holders)){
  beta0 <- parm[1]
  beta1 <- parm[2]
  sigma <- exp(parm[3])
  N <- length(data)
  -sum(log(dLaplace((data$Claims - beta0 - beta1*data$Holders),0,sigma^2)))
}

MLEvB <- MyMLE(ll_B, Insurance, c(100,100,50))
cat("MLE for model B :", c(MLEvB$par[1], MLEvB$par[2], exp(MLEvB$par[3])))
```

```
## MLE for model B : 208.4413 -0.05895382 12.85039
```

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

```
b_B <- MyBIC(MLEvB, MLEvB$par, Insurance)
cat("BIC for model B :", b_B)
```

```
## BIC for model B : -865.3872
```

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 $\theta = (\alpha, \beta, \sigma)$

```
dLognormal <- function(x, mu=0, b=1, params=list(mu, b),...){
  if(!missing(params)){
    mu <- params$mu
    b <- params$b
  }
  d <- 1/(x*b*sqrt(2*pi)) * exp(-((log(x)-mu)^2)/(2*b^2))
}

ll_C <- function(parm, data=list(Claims,Holders)){
  beta0 <- parm[1]
  beta1 <- parm[2]
  sigma <- exp(parm[3])
  N <- length(data)
  x <- 0
```

```r
  for (i in 1:N){
    if(data$Claims[i] == 0){}
    else {
      x = x - log(dLognormal(data$Claims[i],(beta0 + beta1*log(data$Holders[i])),sigma))
    }
  }
  x
}

MLEvC <- MyMLE(ll_C, Insurance, c(100,100,50))
cat("MLE for model C :", c(MLEvC$par[1], MLEvC$par[2], exp(MLEvC$par[3])))
```

```
## MLE for model C : 235.1279 -38.50073 31.108
```

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

```r
b_C <- MyBIC(MLEvC, MLEvC$par, Insurance)
cat("BIC for model C :", b_C)
```

```
## BIC for model C : -82.44094
```

Part D: We want to fit the following models:

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

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

```r
#Part D

dGamma <- function(x, alpha=0, beta=1, params=list(alpha, beta),...){
  if(!missing(params)){
    alpha <- params$alpha
    beta <- params$beta
  }
  d <- ((exp(-x/beta)*x^(alpha-1))/((beta^alpha)*gamma(alpha)))
}

ll_D <- function(parm, data=list(Claims,Holders)){
  beta0 <- parm[1]
  beta1 <- parm[2]
  sigma <- parm[3]
  N <- length(data)
  x <- 0
  for (i in 1:N){
    if(data$Claims[i] == 0){}
    else {
      x = x - log(dGamma(data$Claims[i],(exp(beta0 + beta1*log(data$Holders[i]))),sigma))
    }
  }
  x
}

MLEvD <- MyMLE(ll_D, Insurance, c(3,0,10))
cat("MLE for model D :", c(MLEvD$par[1], MLEvD$par[2], MLEvD$par[3]))
```

```
## MLE for model D : -1.91316 0.7408866 4.381646
```

```r
b_D <- MyBIC(MLEvD, MLEvD$par, Insurance)
cat("BIC for model D :", b_D)
```

```
## BIC for model D : -36.0393
```

(iii) Compare the BIC of all three models

```r
bic_values <- c('A'=b_A, 'B'=b_B, 'C'=b_C, 'D'=b_D)
bic_values = sort(bic_values)
cat("Comparison of BIC values :")
```

```
## Comparison of BIC values :
```

```r
bic_values
```

```
##          B          A          C          D
## -865.38717 -493.45377  -82.44094  -36.03930
```