Problem 4: Modelling Insurance Claims

Deepmalya Dutta

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.

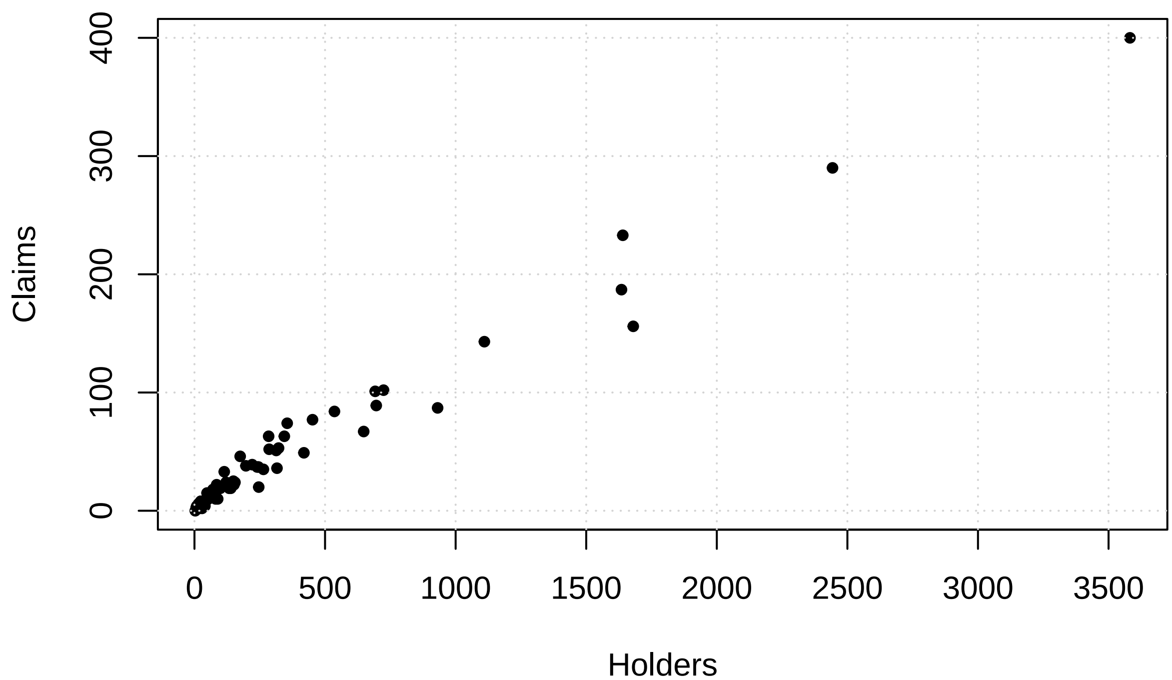
`Group` (an ordered factor): group of car with levels <1 litre, 1–1.5 litre, 1.5–2 litre, >2 litre.

`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,

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

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

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

```
claims <- Insurance$Claims
holders <- Insurance$Holders

gaussian_neglog <- function(theta, data1, data2){
  b0 <- theta[1]
  b1 <- theta[2]
  sigma <- theta[3]
  data2 <- b0 + b1*data2
  return(-sum(dnorm(data1, mean = data2, sd = sigma, log=TRUE)))
}
```

```r
gaussian_estimation <- optim(c(0,0,1), gaussian_neglog, data1 = claims, data2 = holders)
print(gaussian_estimation$par)
```

```
## [1]  8.132948  0.112690 11.849609
```

- So, The Maximum Likelihood Estimate of $\theta = (\beta_0, \beta_1, \sigma)$ is (8.132948, 0.112690, 11.849609)

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

```r
print(paste("The Bayesian Information Criterion (BIC) for this model is",
↪  3*log(length(claims))+2*gaussian_neglog(gaussian_estimation$par, claims, holders)))
```

```
## [1] "The Bayesian Information Criterion (BIC) for this model is 510.759815302333"
```

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

```r
dlaplace <- function(data, location, scale){
  return(log(exp(-abs(data-location)/scale)/(2*scale)))
}

laplacian_neglog <- function(theta, data1, data2){
  b0 <- theta[1]
  b1 <- theta[2]
  sigma <- theta[3]
  data2 <- b0 + b1*data2
  return(-sum(dlaplace(data1, location = data2, scale = sigma)))
}

claims <- Insurance$Claims
holders <- Insurance$Holders
laplacian_estimation <- optim(c(1,0,1), laplacian_neglog, data1 = claims, data2 =
↪  holders)
laplacian_estimation$par
```

```
## [1] 5.0841404 0.1166254 8.2217936
```

- So, The Maximum Likelihood Estimate of $\theta = (\beta_0, \beta_1, \sigma)$ is (5.0841404, 0.1166254, 8.2217936)

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

```r
print(paste("The Bayesian Information Criterion (BIC) for this model is",
↪  3*log(length(claims))+2*laplacian_neglog(laplacian_estimation$par, claims, holders)))
```

```
## [1] "The Bayesian Information Criterion (BIC) for this model is 498.687095702136"
```

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

```
lognormal_neglog <- function(theta, data1, data2){
  n <- length(data1)
  b0 <- theta[1]
  b1 <- theta[2]
  sigma <- theta[3]
  l <-  0
  for (i in 1:n){
    if (data1[i] > 0){
      m <- b0 + b1*data2[i]
      l <- l + dlnorm(data1[i], meanlog = m, sdlog=sigma, log=TRUE)
    }
  }
  return(-l)
}
claims <- Insurance$Claims
holders <- Insurance$Holders
lognormal_estimation <- optim(c(1, 0, 1), lognormal_neglog, data1 = claims, data2 =
↪  holders)
lognormal_estimation$par
```

```
## [1] 2.638435585 0.001474652 0.822601225
```

- So, The Maximum Likelihood Estimate of $\theta = (\beta_0, \beta_1, \sigma)$ is (2.638435585, 0.001474652, 0.822601225)

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

```
print(paste("The Bayesian Information Criterion (BIC) for this model is",
↪  3*log(length(claims))+2*lognormal_neglog(lognormal_estimation$par, claims, holders)))
```

```
## [1] "The Bayesian Information Criterion (BIC) for this model is 568.019648133591"
```

Part D: We want to fit the following models:

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

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

```
gamma_neglog <- function(theta, data1, data2){
  n <- length(data1)
  b0 <- theta[1]
  b1 <- theta[2]
  sigma <- theta[3]
  l <-  0
  for (i in 1:n){
    if (data1[i] > 0){
      m <- b0 + b1*data2[i]
      l <- l + dgamma(data1[i], shape=m, scale=sigma, log=TRUE)
    }
  }
  return(-l)
}
claims <- Insurance$Claims
holders <- Insurance$Holders
```

```r
gamma_estimation <- optim(c(1,0,1), gamma_neglog, data1 = claims, data2 = holders)
gamma_estimation$par
```

```
## [1] 1.89295081 0.04701952 2.58775933
```

- So, The Maximum Likelihood Estimate of $\theta = (\beta_0, \beta_1, \sigma)$ is (1.89295081, 0.04701952, 2.58775933)

(iii) Compare the BIC of all three models

- Ans: The BIC is lowest for the Laplace Distribution. So we will prefer that model for our estimation.