Bayesian Model For Motor Claim Insurance

jimmynguyen

2022-05-31

```
##Initial Set-up Part A (Question 1 to Question 3)
##read simulated data for number of claims for each policy, 183,999 policies
in total
Pclaim<- read.csv("norauto.csv")

#list given variables of data sets
names(Pclaim)

## [1] "Male"        "Young"       "DistLimit"   "GeoRegion"   "Expo"
## [6] "ClaimAmount" "NbClaim"

#check the structure of data sets
str(Pclaim)

## 'data.frame':    183999 obs. of  7 variables:
##  $ Male       : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ Young      : int  0 1 1 0 0 0 0 1 0 1 ...
##  $ DistLimit  : Factor w/ 6 levels "12000 km","16000 km",..: 1 1 4 1 1 1 1
1 1 1 ...
##  $ GeoRegion  : Factor w/ 6 levels "High-","High+",..: 4 6 6 5 1 6 4 1 1 4
...
##  $ Expo       : num  0.789 0.2 0.285 0.247 0.995 0.222 0.794 0.146 0.674
0.663 ...
##  $ ClaimAmount: int  0 0 0 0 18158 0 11790 0 0 0 ...
##  $ NbClaim    : int  0 0 0 0 1 0 1 0 0 0 ...

#create a new data set of first 100 policies
Pclaim100 <- Pclaim[1:100,]
Pclaim100

##     Male Young      DistLimit GeoRegion  Expo ClaimAmount NbClaim
## 1      0     0       12000 km      Low+ 0.789           0       0
## 2      0     1       12000 km   Medium+ 0.200           0       0
## 3      0     1 25000-30000 km   Medium+ 0.285           0       0
## 4      0     0       12000 km   Medium- 0.247           0       0
## 5      0     0       12000 km     High- 0.995       18158       1
## 6      0     0       12000 km   Medium+ 0.222           0       0
## 7      0     0       12000 km      Low+ 0.794       11790       1
## 8      0     1       12000 km     High- 0.146           0       0
## 9      0     0       12000 km     High- 0.674           0       0
## 10     0     1       12000 km      Low+ 0.663           0       0
## 11     0     0       20000 km     High- 0.184           0       0
## 12     0     1       12000 km     High- 0.504           0       0
```

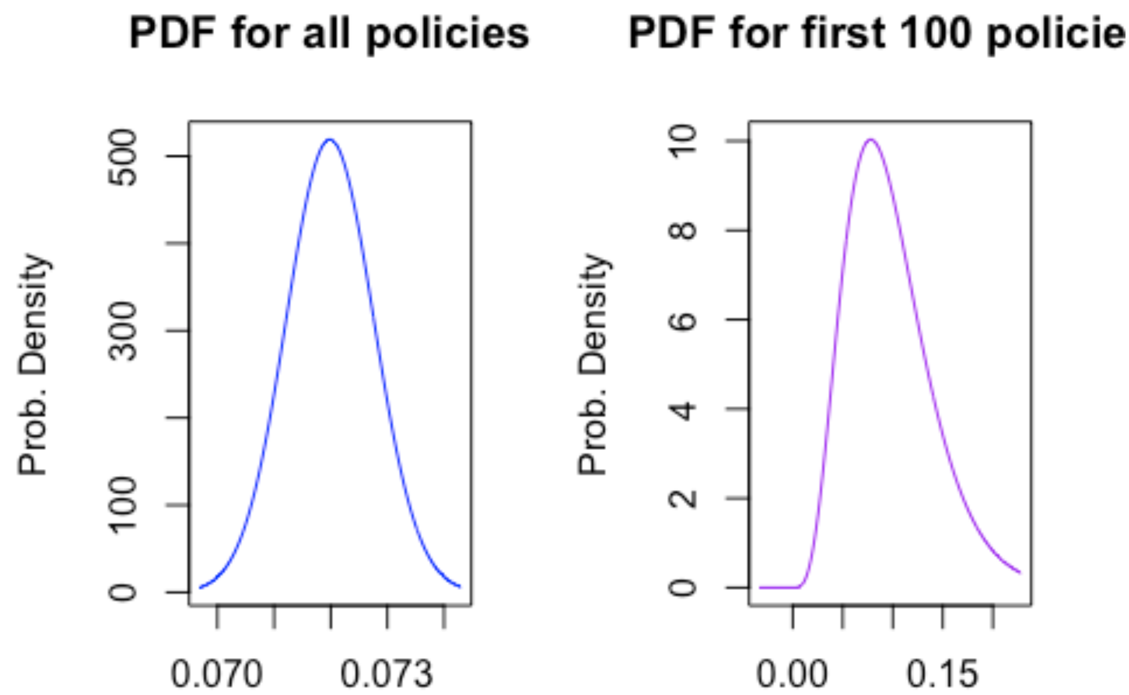
```
## 13      0      0       12000 km       Low+ 0.945           0        0
## 14      0      1       12000 km      High- 0.397           0        0
## 15      0      0       12000 km    Medium+ 1.000           0        0
## 16      0      0       20000 km    Medium- 1.000           0        0
## 17      0      1       12000 km      High- 0.526           0        0
## 18      0      0       12000 km      High- 0.318           0        0
## 19      0      1       20000 km    Medium+ 0.452           0        0
## 20      0      0       12000 km    Medium+ 0.293           0        0
## 21      0      1 25000-30000 km    Medium- 0.378           0        0
## 22      0      0       12000 km      High- 0.181           0        0
## 23      0      0 25000-30000 km    Medium- 1.003           0        0
## 24      0      0       20000 km       Low+ 0.271           0        0
## 25      1      1       12000 km      High+ 0.378           0        0
## 26      1      1       12000 km      High+ 0.134           0        0
## 27      1      0       12000 km      High+ 0.663           0        0
## 28      1      0       12000 km      High+ 0.255           0        0
## 29      1      0       20000 km      High+ 0.323           0        0
## 30      1      1       12000 km      High- 0.370           0        0
## 31      1      1       12000 km      High- 1.000           0        0
## 32      1      1 25000-30000 km      High- 1.000           0        0
## 33      1      0       12000 km      High- 0.778       39913        1
## 34      1      0       12000 km      High- 0.211           0        0
## 35      1      0       12000 km      High- 0.948           0        0
## 36      1      0       no limit      High- 0.271           0        0
## 37      1      1       20000 km    Medium+ 0.359           0        0
## 38      1      0       12000 km    Medium+ 0.997        7051        1
## 39      1      1       20000 km    Medium- 1.000           0        0
## 40      1      0       no limit    Medium- 0.340           0        0
## 41      1      1       no limit       Low- 0.575           0        0
## 42      1      1       12000 km      High+ 0.293           0        0
## 43      1      1       12000 km      High+ 0.518           0        0
## 44      1      0       12000 km      High+ 0.759           0        0
## 45      1      0       12000 km      High+ 0.222           0        0
## 46      1      1       12000 km      High- 0.496           0        0
## 47      1      1 25000-30000 km      High- 1.003           0        0
## 48      1      0       12000 km      High- 0.219           0        0
## 49      1      0       12000 km      High- 0.529           0        0
## 50      1      0       20000 km      High- 1.003           0        0
## 51      1      1       12000 km    Medium+ 1.003           0        0
## 52      1      1       12000 km    Medium+ 1.000           0        0
## 53      1      0       20000 km    Medium+ 0.641           0        0
## 54      1      1       20000 km    Medium- 0.937           0        0
## 55      1      0       12000 km       Low+ 0.285           0        0
## 56      1      0       20000 km       Low+ 0.482           0        0
## 57      1      1       12000 km      High+ 0.148           0        0
## 58      1      0       12000 km      High+ 0.723           0        0
## 59      1      0       12000 km      High+ 0.597           0        0
## 60      1      0       12000 km      High+ 0.400           0        0
## 61      1      0       20000 km      High+ 0.299           0        0
## 62      1      0 25000-30000 km      High+ 0.573           0        0
```

```
## 63      1     1       12000 km     High- 0.134          0        0
## 64      1     1       12000 km     High- 0.107          0        0
## 65      1     0       12000 km     High- 0.712          0        0
## 66      1     1       12000 km   Medium+ 0.441          0        0
## 67      1     1       12000 km   Medium+ 0.140          0        0
## 68      1     1       20000 km   Medium+ 0.677          0        0
## 69      1     0       12000 km   Medium+ 0.205          0        0
## 70      1     1       12000 km   Medium- 0.537          0        0
## 71      1     0       20000 km   Medium- 0.329          0        0
## 72      1     0       no limit   Medium- 0.249          0        0
## 73      1     1       no limit   Medium+ 0.340          0        0
## 74      1     0       20000 km   Medium+ 1.000          0        0
## 75      1     0       20000 km      Low+ 0.192          0        0
## 76      1     0       no limit     High- 0.214          0        0
## 77      1     1 25000-30000 km   Medium+ 0.337          0        0
## 78      1     0       12000 km   Medium+ 0.337          0        0
## 79      1     0       20000 km     High- 0.444          0        0
## 80      1     0       20000 km   Medium- 0.167          0        0
## 81      1     0       12000 km     High+ 0.482          0        0
## 82      1     0       12000 km   Medium+ 1.003          0        0
## 83      1     0       12000 km   Medium+ 0.197          0        0
## 84      1     0       no limit   Medium+ 0.444          0        0
## 85      1     0       12000 km     High+ 0.112          0        0
## 86      1     0       20000 km   Medium+ 0.416          0        0
## 87      1     0       12000 km   Medium- 0.515          0        0
## 88      1     0       12000 km   Medium- 0.416          0        0
## 89      1     0       no limit   Medium- 1.000          0        0
## 90      1     1       12000 km      Low+ 0.416          0        0
## 91      1     0       12000 km      Low+ 0.545          0        0
## 92      1     1       no limit      Low- 0.540          0        0
## 93      1     0       12000 km      Low- 0.860          0        0
## 94      1     1       12000 km   Medium+ 0.630          0        0
## 95      1     1       no limit      Low+ 0.499       6599        1
## 96      1     1       20000 km   Medium+ 0.734          0        0
## 97      1     0 25000-30000 km   Medium+ 0.375          0        0
## 98      1     1       12000 km     High+ 0.263          0        0
## 99      1     0       12000 km     High+ 0.208          0        0
## 100     1     1       20000 km     High- 0.444          0        0
```

```r
#valuate alpha and beta which are calculated in the report
alpha <- 0.018
beta <- 5/3

#Question 1: evaluate mean & sd for 183,999 policies and first 100 policies
#valuate alpha_n and beta_n
alpha_n <- sum(Pclaim$NbClaim)+alpha
alpha_n
```

```
## [1] 8772.018
```

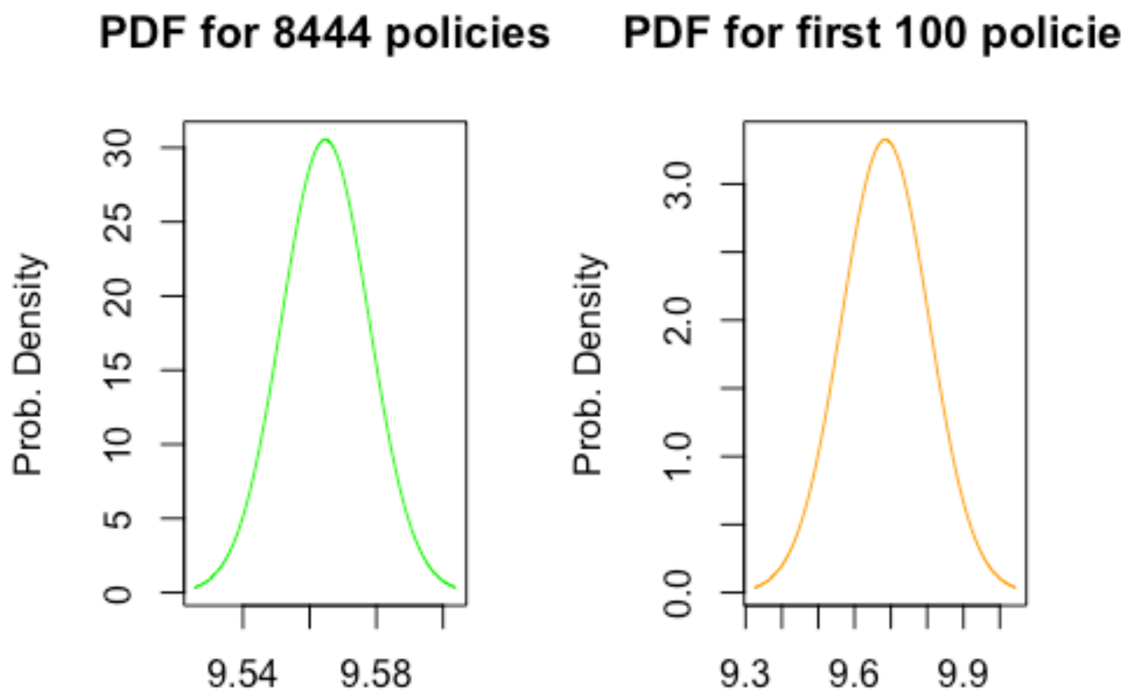
```r
beta_n <-  beta/(1+beta*sum(Pclaim$Expo))
beta_n

## [1] 8.207036e-06

#in the case of 183,999 policies
mu1 <- alpha_n*beta_n
mu1

## [1] 0.07199227

sd1 <- beta_n * sqrt(alpha_n)
sd1

## [1] 0.0007686632

#valuate alpha_100 and beta_100
alpha_100 <- sum(Pclaim100$NbClaim)+alpha
alpha_100

## [1] 5.018

beta_100 <-  beta/(1+beta*sum(Pclaim100$Expo))
beta_100

## [1] 0.01943257

#in the case of first 100 policies
mu2 <- alpha_100*beta_100
mu2

## [1] 0.09751263

sd2 <- beta_100 * sqrt(alpha_100)
sd2

## [1] 0.04353069

#Question 2: Plot the Bayesian posterior density for 183,999 policies and 100
policies
c1 <- seq(mu1-3*sd1, mu1+3*sd1, 0.01*sd1)
pdf1<- dgamma(c1, shape = alpha_n, scale = beta_n, log = FALSE)

c2 <- seq(mu2-3*sd2, mu2+3*sd2, 0.01*sd2)
pdf2<- dgamma(c2, shape = alpha_100, scale = beta_100, log = FALSE)


par(mfrow=c(1,2))

plot(c1, pdf1, col="blue",xlab="", ylab="Prob. Density", type="l",main="PDF
for all policies")
plot(c2, pdf2, col="purple",xlab="", ylab="Prob. Density", type="l",main="PDF
for first 100 policies")
```




```
#Compute 90% Bayesian confidence interval for λ
#in the case of 183,999 policies
BCILower1<- mu1 - qnorm(0.95)*sd1
BCIUpper1<- mu1 + qnorm(0.95)*sd1
BCI1<-c(BCILower1, BCIUpper1)
BCI1

## [1] 0.07072793 0.07325660

#in the case of first 100 policies
BCILower2<- mu2 - qnorm(0.95)*sd2
BCIUpper2<- mu2 + qnorm(0.95)*sd2
BCI2<-c(BCILower2, BCIUpper2)
BCI2

## [1] 0.02591102 0.16911424

#Question 3: evaluate the credibility estimator of λ and credibility factor
for 183,999 policies and 100 policies

pmu1 <- 0.03
psd1 <- 0.05

##in the case of 183,999 policies
```

```
z1 <- beta_n*sum(Pclaim$Expo)
z1

## [1] 0.9999951

lamda_mle1 <- sum(Pclaim$NbClaim)/sum(Pclaim$Expo)
lamda_mle1

## [1] 0.07199247

CRlamda1 <- pmu1*(1-z1)+lamda_mle1*z1
CRlamda1

## [1] 0.07199227

##in the case of 100 policies
z2 <- beta_100*sum(Pclaim100$Expo)
z2

## [1] 0.9883405

lamda_mle2 <- sum(Pclaim100$NbClaim)/sum(Pclaim100$Expo)
lamda_mle2

## [1] 0.09830908

CRlamda2 <- pmu1*(1-z2)+lamda_mle2*z2
CRlamda2

## [1] 0.09751263

##Initial Set-up Part B (Question 4 to Question 6)
#Obmit null data based on Claim Amount
Pclaim1 <- Pclaim[Pclaim$ClaimAmount != 0,]

#create new dataframe for first 100 policies
Pclaim1_100 <- Pclaim1[1:100,]

#calculate the log of claim amount
log_ClaimAmount1 <- log(Pclaim1$ClaimAmount)
log_ClaimAmount2 <- log(Pclaim1_100$ClaimAmount)

sd3 <- 1.2
pmu2 <- 6.0
tau <- 4.0^2

#Question 4: Compute the posterior mean and standard deviation of the unknown
mean parameter θ
#in case of 8,444 policies
n1 <- length(Pclaim1$ClaimAmount)
taupost1<-1/(n1/(sd3^2) + 1/tau)
```

```r
sd_taupost1 <- sqrt(taupost1)
sd_taupost1

## [1] 0.01305885

mupost1 <- taupost1*(pmu2/tau + n1*mean(log_ClaimAmount1)/sd3^2)
mupost1

## [1] 9.564685

#in case of 100 policies
n2 <- length(Pclaim1_100$ClaimAmount)
taupost2 <-1/(n2/(sd3^2) + 1/tau)
sd_taupost2 <- sqrt(taupost2)
sd_taupost2

## [1] 0.119946

mupost2 <- taupost2*(pmu2/tau + n2*mean(log_ClaimAmount2)/sd3^2)
mupost2

## [1] 9.684676

#Question 5: Plot the posterior and to compute a 90% Bayesian confidence
interval for θ

#in case of 8,444 policies
c3<-seq(mupost1-3*sqrt(taupost1),mupost1+3*sqrt(taupost1),
0.01*sqrt(taupost1))
pdf3<-dnorm(c3, mupost1, sqrt(taupost1))

#in case of 100 policies
c4<-seq(mupost2-3*sqrt(taupost2),mupost2+3*sqrt(taupost2),
0.01*sqrt(taupost2))
pdf4<-dnorm(c4, mupost2, sqrt(taupost2))

#plot the posterior for 2 cases
par(mfrow=c(1,2))

plot(c3, pdf3, col="green",xlab="", ylab="Prob. Density", type="l", main="PDF
for 8444 policies")
plot(c4, pdf4, col="orange",xlab="", ylab="Prob. Density", type="l",main="PDF
for first 100 policies")
```




```
#Compute a 90% Bayesian confidence interval for 8444 pocilies and 100 first
policies
#in case of 8,444 policies
BCILower3<- mupost1 - qnorm(0.95)*sqrt(taupost1)
BCIUpper3<- mupost1 + qnorm(0.95)*sqrt(taupost1)
BCI3<-c(BCILower3, BCIUpper3)
BCI3

## [1] 9.543205 9.586165

#in case of 100 policies
BCILower4<- mupost2 - qnorm(0.95)*sqrt(taupost2)
BCIUpper4<- mupost2 + qnorm(0.95)*sqrt(taupost2)
BCI4<-c(BCILower4, BCIUpper4)
BCI4

## [1] 9.487383 9.881970

#Question 6: evaluate the credibility estimator of λ and credibility factor
for 183,999 policies and 100 policies

sd3 <- 1.2
pmu2 <- 6.0
tau <- 4.0^2
```

```r
##in the case of 8,444 policies
z3 <- (n1*tau)/(n1*tau+sd3^2)
z3

## [1] 0.9999893

CRlamda3 <- z3*mean(log_ClaimAmount1) + (1-z3)*pmu2
CRlamda3

## [1] 9.564685

##in the case of 100 policies
z4 <- (n2*tau)/(n2*tau+sd3^2)
z4

## [1] 0.9991008

CRlamda4 <- z4*mean(log_ClaimAmount2) + (1-z4)*pmu2
CRlamda4

## [1] 9.684676
```