

Tarea 5:

1) Reflexión 2.3.4.

Año		Maximo Muerto - 3 ^{er} Quantil	
2005	3 ^{er} Quantil	Valor Máximo	Mediana
2005	152,776	6287,67	35,997
2006	141,500	4761,277	33,895
2007	145,546	3533,754	37,552

	Promedio
2005	173,669
2006	181,778
2007	157,904

En el box plot al final de la sección 2.3.4, las cajas de los tres años tienen la misma forma, no se pueden apreciar correctamente, además puedo observar que en todos los años tengo valores extremos << outliers >> que seagan a la derecha, por lo cual yo quitaría alguna de ellas,

Aún no manejo muy bien los criterios pero en este caso delimitaria, en los tres años los siniestros hasta 2000 (\leq), se modificarían ~~to 2000~~ ~~2000~~ ~~2000~~ ~~permanecer~~. Esto para poder hacer una nueva comparación.

También puedo que hay una gran acumulación en los valores de "Claim-Amount" $\in (0, 100]$ (aprox), estos los analizaría aparte ya que considero se pueden segmentar entre los valores $\in (0, 100] \cup [100, 2000]$

Reflexión 3.1

No puedo hacer observaciones o reflexiones propias en esta sección, ya que no logro comprender la modelación.

- 2) Hay algunas cosas en el código de usted que sí se entiende, de hecho, al intentar correr su código me marcó un error en "loglik-grid" = \mathbb{Z} , supongo que ha de ser porque no coincide el rango de la matriz con las de thetalambda-grid

3) Lo que permite que se pueda calcular por separado es:

- La homogeneidad en t
- La simetría estocástica en t e $i = 1, \dots, N_t \rightarrow$ indep. estocástica entre periodos \times entre pólizas

Las condiciones que harían que no se pudieran calcular por separado serían:

- que ~~algunas~~ ~~algunas~~ al menos una póliza se siniestrara más de una vez \rightarrow necesitaría Bin Neg
- que se rompa la indep. estocástica

```
### 1
```

```
#install.packages("dplyr", verbose=T)
```

```
library("dplyr")
```

```
library("repmis")
```

```
data <- source_data("https://github.com/JCMO-  
ITAM/Data4Analysis/blob/master/d4a_allstateclaim_data.csv?raw=true")
```

```
head(data)
```

```
data <- as.data.frame(data)
```

```
data[, "iota"] <- 0
```

```
data[which(data$Claim_Amount>0), "iota"] <- 1
```

```
table(data[,c("Calendar_Year", "iota")])
```

```
#2005
```

```
data_x1 <- data[which(data$iota==1 & data$Calendar_Year==2005 & data$Claim_Amount <=  
2000), "Claim_Amount"]
```

```
summary(data_x1)
```

```
hist(data_x1, 100)
```

```
#2006
```

```
data_x2 <- data[which(data$iota==1 & data$Calendar_Year==2006 & data$Claim_Amount <=  
2000), "Claim_Amount"]
```

```
summary(data_x2)
```

```
hist(data_x2,100)
```

```
#2007
```

```
data_x3 <- data[which(data$iota==1 & data$Calendar_Year==2007 & data$Claim_Amount <= 2000),"Claim_Amount"]
```

```
summary(data_x3)
```

```
hist(data_x3,100)
```

```
data_x <- data[which(data$iota==1 & data$Claim_Amount <=2000),c("Calendar_Year", "Claim_Amount")]
```

```
data_x$Calendar_Year <- as.factor(data_x$Calendar_Year)
```

```
colnames(data_x)
```

```
library("ggplot2")
```

```
ggplot(data = data_x, aes(x = Calendar_Year, y = Claim_Amount)) +
```

```
  geom_boxplot(aes(fill = Calendar_Year), width = 0.8) + theme_bw()
```

```
#llegue a reducir el data$Claim_Amount hasta 100 y los boxplots de cada uno de los años
```

```
#se mantenía casi igual, por lo que afirmo que los valores entre 0 y 100 se debería graficar
```

```
# y analizar aparte de los demás datos
```

```
### 2) -----
```

```
theta_grid <- seq(0,1,0.01)
```

```
lambda_grid <- seq(0,1000,1)
```

```
if(!require('lattice')){install.packages("lattice")}
```

```
library("lattice")
```

```

data_n <- sum(data$iota)
data_J <- nrow(data)
data_xsum <- sum(data[which(data$iota==1),"Claim_Amount"])
theta <- data_n/data_J; lambda <- (data_xsum / data_n)
loglikelihood <- function(theta,lambda,data_n,data_J,data_xsum){
  loglik.theta <- data_n*log(theta) + (data_J-data_n)*log(1-theta)
  loglik.lambda <- data_n*log(lambda) -lambda*data_xsum
  loglik <- loglik.theta + loglik.lambda
  return(loglik)
}
loglikelihood(theta,lambda,data_n,data_J,data_xsum)

thetalambda_grid <- expand.grid( x = theta_grid, y = lambda_grid)
dim(thetalambda_grid)

loglik_grid <- matrix(NaN,ncol=1,nrow=nrow(thetalambda_grid))
G <- nrow(thetalambda_grid)
g <- 1
for(g in 1:G){loglik_grid[g] <-
loglikelihood(thetalambda_grid[g,1],thetalambda_grid[g,2],data_n,data_J,data_xsum)}

if(!require('plot3D')){install.packages("plot3D")}
library("plot3D")

scatter3D(thetalambda_grid[,1], thetalambda_grid[,2], loglik_grid,
  pch = 18, cex = 2,
  theta = 20, phi = 20, ticktype = "detailed",
  xlab = "theta", ylab = "lambda", zlab = "log-lik",

```

```
surf = list(x = thetalambda_grid[,1], y = thetalambda_grid[,2], z = loglik_grid,  
            facets = NA, fit = fitpoints), main = "")
```

#Le pido me tenga paciencia, aun estoy aprendiendo a manejar R, estoy poniendo empeño en los
#cursos de DataCamp, de hecho, con el 1er curso pude realizar mi analisis de la pregunta 1
#no se como corregir su codigo, lo que pienso es que son de diferentes longitudes las matrices
#por lo cual no las puedo graficar.

```
###3) -----
```