homework\_three

Yi (Chris) Chen

October 23, 2017

## Homework Three

#### UNI:yc3356

#### Name: Yi Chen

#### Email:[yc3356@columbia.edu](mailto:yc3356@columbia.edu)

**Goals**: writing functions to automate repetitive tasks and using them as larger parts of code, some practice with ggplot, working with data frames and manipulating data from one form to another.

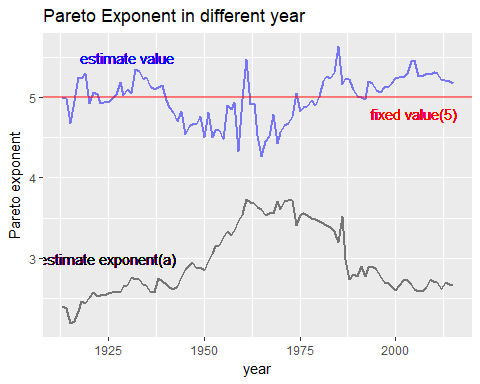
## Part 1: Estimating a on US data

### i.Write a function which takes P99.5, P99.9, and a, and calculates the lefthand side of that equation. Plot the values for each year using ggplot, using the data and your estimates of the exponent from lab (using the exponent.est ratio()). Add a horizontal line with vertical coordinate 5. How good is the fit?

#read the data  
setwd("C:/Users/cheny/Desktop/study/statistical computing and intro to data science/homework/homework three")  
Data <- read.csv('wtid-report.csv',header = TRUE)  
Data <- Data[,-1] # only take the col that needed  
colnames(Data) <- c('year','P99','P99.5','P99.9') # rename the col name  
  
#calculate the value of a  
exponent.est\_ratio <- function(p1=Data$P99,p2=Data$P99.9){  
 a <- 1- (log(10)/(log(p1/p2)))  
 return(a)  
}  
  
a <- exponent.est\_ratio()  
  
#function  
problem\_one <- function(p1=Data$P99.5,p2=Data$P99.9,a){  
 a <- exponent.est\_ratio()  
 return((p1/p2)^(-a+1))  
}  
   
b <- problem\_one()   
   
   
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.4.2

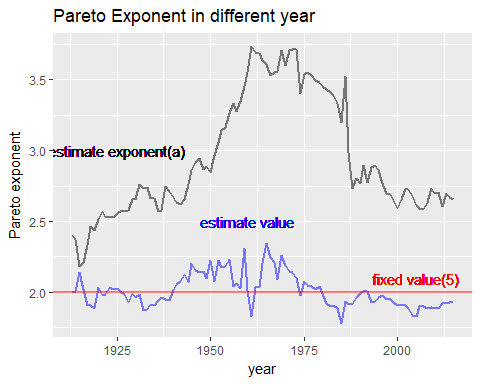
ggplot(data=Data) +   
 geom\_line(mapping = aes(x=year,y=a),size=1,col='black',alpha=0.5) +  
 labs(title='Pareto Exponent in different year',y='Pareto exponent',x='year') +  
 geom\_line(mapping = aes(x=year,y=b),size=1,col='blue',alpha=0.5)+  
 geom\_hline(yintercept = 5 , size=1, col='red',alpha=0.5)+  
 geom\_text(mapping = aes(x=1925, y=3, label = 'estimate exponent(a)'), size=4,col='black') +  
 geom\_text(mapping = aes(x=1930, y=5.5, label = 'estimate value'), size=4,col='blue') +  
 geom\_text(mapping = aes(x=2005, y=4.8, label = 'fixed value(5)'), size=4,col='red')



* In general, estimate value is roughly round the fixed value 5.

### ii. Repeatthe previous step with this formula. How would you describe this fit compared to the previous one?

problem\_two <- function(p1=Data$P99,p2=Data$P99.5,a){  
 a <- exponent.est\_ratio()  
 return((p1/p2)^(-a+1))  
}  
   
b <- problem\_two()   
   
   
library(ggplot2)  
ggplot(data=Data) +   
 geom\_line(mapping = aes(x=year,y=a),size=1,col='black',alpha=0.5) +  
 labs(title='Pareto Exponent in different year',y='Pareto exponent',x='year') +  
 geom\_line(mapping = aes(x=year,y=b),size=1,col='blue',alpha=0.5)+  
 geom\_hline(yintercept = 2 , size=1, col='red',alpha=0.5)+  
 geom\_text(mapping = aes(x=1925, y=3, label = 'estimate exponent(a)'), size=4,col='black') +  
 geom\_text(mapping = aes(x=1960, y=2.5, label = 'estimate value'), size=4,col='blue') +  
 geom\_text(mapping = aes(x=2005, y=2.1, label = 'fixed value(5)'), size=4,col='red')



* In general, estimate value is roughly round the fixed value 5.

### 