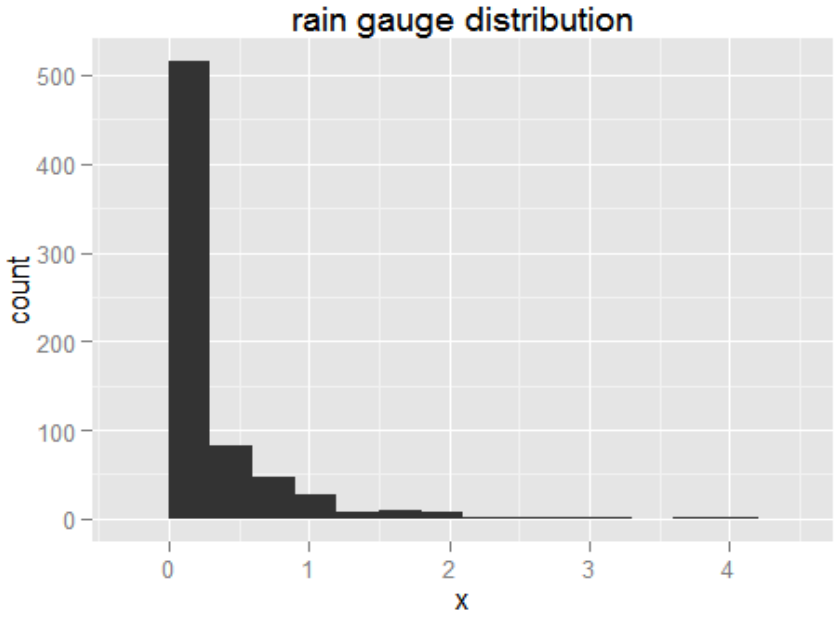
midtermrain-lina.R

Lina Zhou

Mon Oct 19 22:15:56 2015

```
## read data in loops
library(dplyr)

library(ggplot2)
raw<-NULL
Y<-c("00","01","02","03","04")
M<-c("01","02","03","04","05","06","07","08","09","10","11","12")
for(i in 1:5)
{
  for(j in 1:12)
  {
    name<-paste("L-",Y[i],"-",M[j],".txt",sep="")
    raw<-rbind(raw,read.csv(name,skip=2,stringsAsFactors = F,header=T))
  }
}

## data cleaning
raw<-raw[,2:25]
dim(raw)

## [1] 1827   24

## let T=0,other strange data=-1
raw[raw=="T"]<-0
raw[raw=="----"]<-(-1)
raw[raw=="M"]<-(-1)
raw[raw=="T    "]<-0
raw[raw=="M    "]<-(-1)

class(raw[1,1])

## [1] "character"

test<-as.matrix(raw)
data<-as.numeric(t(test))## trans the matrix and make it into a vector
length(data)

## [1] 43848

## count the rain.
rain<-NULL
temp<-0
k=1
```

```
while(k<length(data))
{
  if(data[k]>=0)
  {
    while(data[k]>=0)
    {
      temp=temp+data[k]
      k=k+1
    }
    rain<-c(rain,temp)
    temp=0
  }
  else{k=k+1}
}
rain

## MEM MLE
data1<-data.frame(x=rain)
data1<-subset(data1,data1$x!=0.00)
head(data1)

##      x
## 1 0.03
## 3 0.03
## 4 0.01
## 5 0.01
## 6 0.97
## 7 0.06

ggplot(data=data1)+geom_histogram(aes(x),binwidth=.3)+labs(title="rain gauge distribution")
```



```r
mean(data1$x)
```

```
## [1] 0.2831108
```

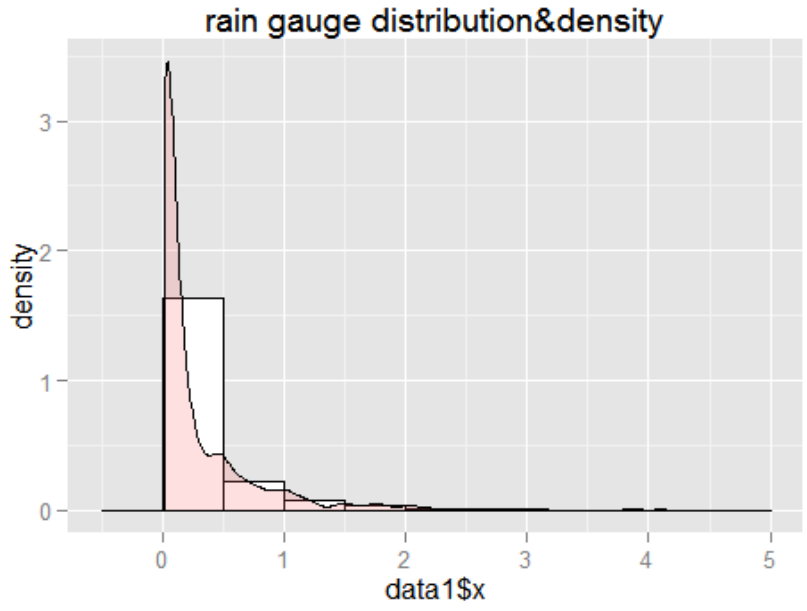
```r
var(data1$x)
```

```
## [1] 0.2218382
```

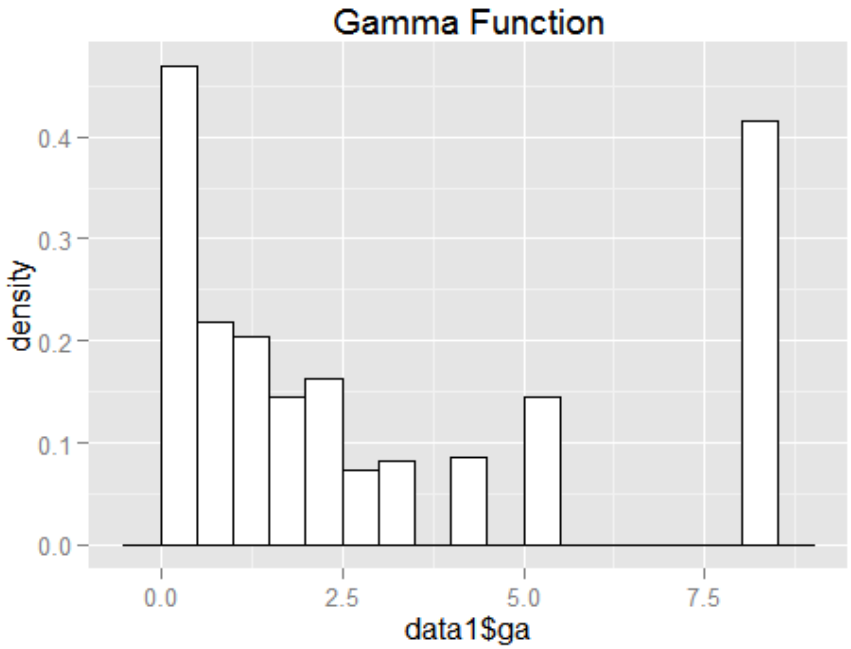
```r
## 2 parameters
alpha <- mean(data1$x)^2/var(data1$x)
lambda <- mean(data1$x)/var(data1$x)
## plot HW1
gam<-function(x)
{
  gamm<-(lambda^alpha)*x^(alpha-1)*exp(-lambda*x)/gamma(alpha)
  return(gamm)
}
data1$ga<-gam(data1$x)
head(data1)
```

```
##      x        ga
## 1 0.03 4.0063483
## 3 0.03 4.0063483
## 4 0.01 8.2902504
## 5 0.01 8.2902504
## 6 0.97 0.1310848
## 7 0.06 2.4765980
```

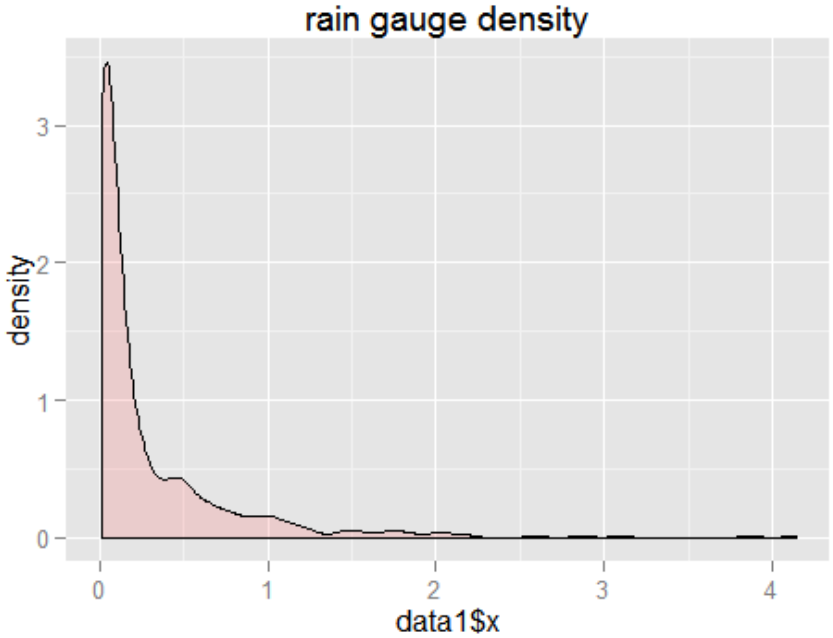
```r
## plot of density & count
ggplot(data1,aes(x=data1$x))+geom_histogram(aes(y=..density..),binwidth=.5,colour="black",fill="white")+labs(title="rain gauge distribution&density")+geom_density(alpha=.2,fill="#FF6666")
```



```
## plot of density function
ggplot(data1,aes(x=data1$ga))+geom_histogram(aes(y=..density..),binwidth=.5,colour="black",fill="white")+labs(title="Gamma Function")
```



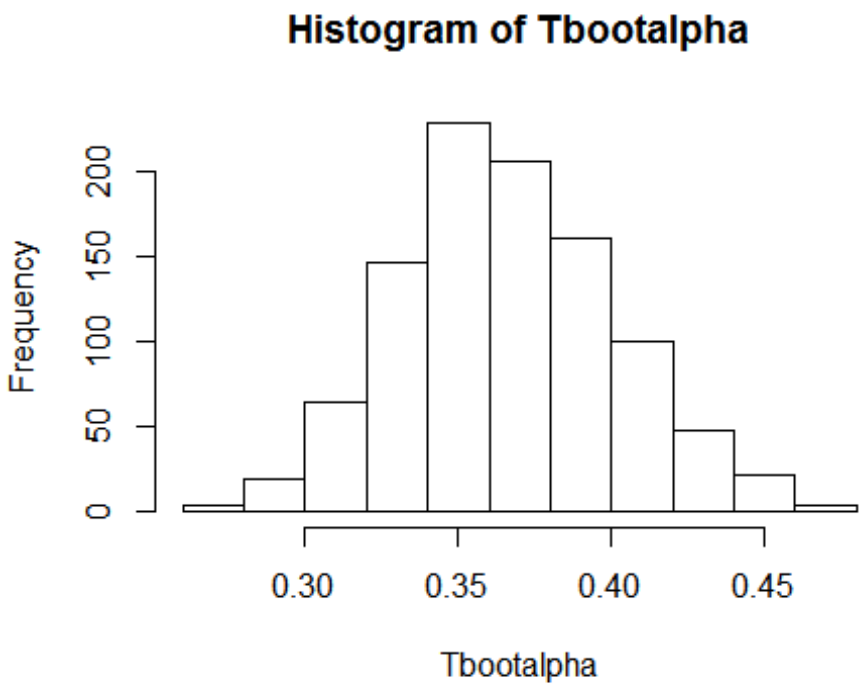

```
## plot of density
ggplot(data1,aes(x=data1$x))+geom_density(alpha=0.2,fill="#FF6666")+labs(title="rain gauge density")
```



```r
## bootstrap HW2
n<-length(data1$x)
B<-1000
Tbootalpha<-rep(0,B)
Tbootlambda<-rep(0,B)
for(i in 1:B){
  al.s<-sample(data1$x,n,replace=TRUE)
  lam.s<-sample(data1$x,n,replace=TRUE)
  Tbootalpha[i]<-mean(al.s)^2/var(al.s)
  Tbootlambda[i]<-mean(lam.s)/var(lam.s)
}
## standrad error for alpha
seal<-sqrt(var(Tbootalpha))
seal
```

```
## [1] 0.03557355
```

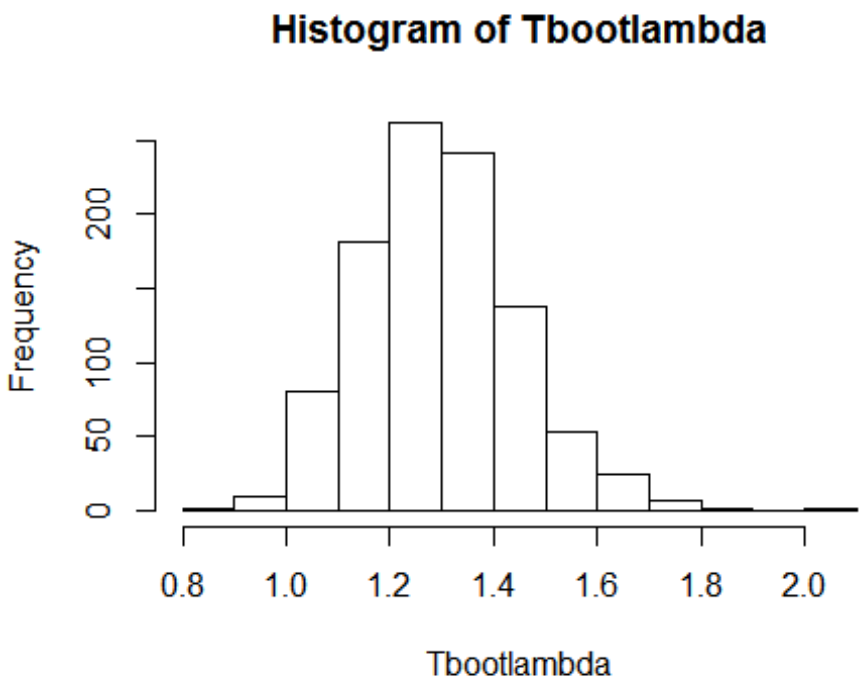
```r
hist(Tbootalpha)
```




```r
## standrad error for lambda
selam<-sqrt(var(Tbootlambda))
selam
```

```
## [1] 0.1504287
```

```r
hist(Tbootlambda)
```




```
## MLE
theta <- c(alpha,lambda)
minus.likelihood <- function(theta) {-(n*theta[1]*log(theta[2])-n*lgamm
a(theta[1])+(theta[1]-1)*sum(log(data1$x))-theta[2]*sum(data1$x))}
max.likelihood <- nlminb(start=c(alpha,lambda), obj = minus.likelihood)
theta
```

```
## [1] 0.3613071 1.2762037
```

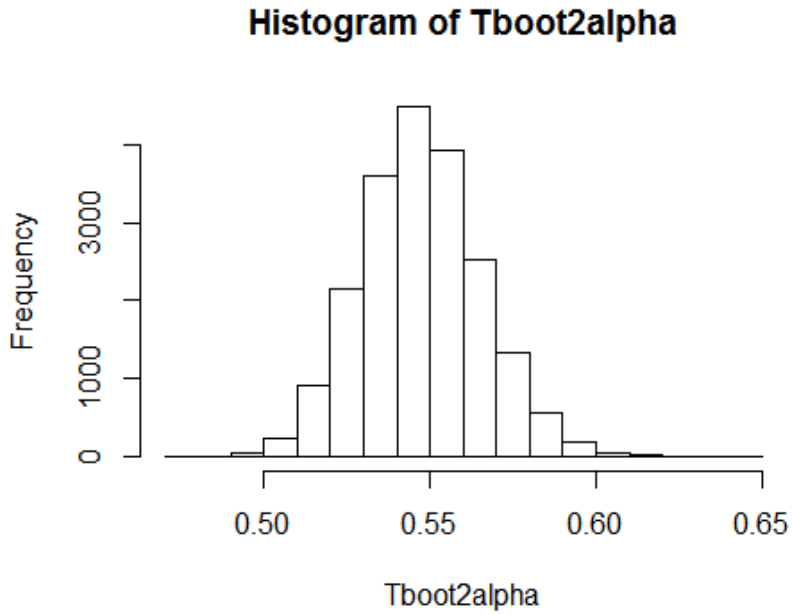
```
max.likelihood$par
```

```
## [1] 0.5461541 1.9291179
```

```
## bootstrap MLE
B2<-20000
Tboot2alpha<-rep(0,B2)
Tboot2lambda<-rep(0,B)
minus.likelihood2 <- function(theta) {-(n*theta[1]*log(theta[2])-n*lgam
ma(theta[1])+(theta[1]-1)*sum(log(boot))-theta[2]*sum(boot))}
for(i in 1:B2)
{  boot <- sample(data1$x, n, replace=TRUE)
  mle <- nlminb(start=c(.1794, 1.1062), obj = minus.likelihood2)
  Tboot2alpha[i]<-mle$par[1]
  Tboot2lambda[i]<-mle$par[2]
}
## standard error for alpha in mle
se2al<-sd(Tboot2alpha)
se2al
```

```
## [1] 0.017941
```

```
hist(Tboot2alpha)
```




```
## standard error for lambda in mem
se2la<-sd(Tboot2lambda)
se2la
```

```
## [1] 0.1584769
```

```
hist(Tboot2lambda)
```