HW2

Group T_8

March 21, 2018

Contents

1 Task 27:

1.1 The method of moments

The theoretical moments of the Lomax distribution are:

$$\mu_1 = \frac{\lambda}{\alpha - 1}$$

$$\mu_2 = \frac{\lambda^2 \alpha}{(\alpha - 1)^2 (\alpha - 2)}$$

The sample moments are:

$$\bar{x} = \frac{1}{n} \sum x_i$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum (x_i - \bar{x})^2$$

Now we solve the system for the parameters:

$$\mu_1 = \bar{x}$$

$$\mu_2 = \hat{\sigma}^2$$

And get the next solution:

$$\hat{\alpha} = \frac{2\hat{\sigma}^2}{\hat{\sigma}^2 - \bar{x}^2}$$

$$\hat{\lambda} = \bar{x} \frac{\hat{\sigma}^2 + \bar{x}^2}{\hat{\sigma}^2 - \bar{x}^2}$$

1.2 The method of maximum likelihood

Since RVs are i.i.d and Lomax distributed, their joint frequency function is the product of the marginal frequency functions. So the log-likelihood:

$$l(\alpha, \lambda) = \sum_{i=1}^{n} \log \frac{\alpha \lambda^{\alpha}}{(\lambda + x_i)^{\alpha+1}} = \sum_{i=1}^{n} (\log \alpha + \alpha \log \lambda - (\alpha + 1) \log (\lambda + x_i))$$

Finally, by setting the partial derivatives equal to zero we will find the equations that the MLE must satisfy:

$$\frac{\partial l(\alpha, \lambda)}{\partial \alpha} = \sum_{i=1}^{n} \left(\frac{1}{\alpha} + \log \lambda - \log (\lambda + x_i) \right) = 0$$

$$\frac{\partial l(\alpha, \lambda)}{\partial \lambda} = \sum_{i=1}^{n} \left(\frac{\alpha}{\lambda} - \frac{\alpha + 1}{\lambda + x_i} \right) = 0$$

2 Task 28:

The CDF of Weibull distribution, $X \sim Weibull(c, \gamma)$ is the following:

$$F_X(x) = 1 - e^{cx^\gamma}$$

Let $Y = X^\gamma$

$$F_Y(y) = P(Y \leq y) = P(X^\gamma \leq y) = P(X \leq y^{\frac{1}{\gamma}}) = F_X(y^{\frac{1}{\gamma}}) = 1 - e^{-cy}$$

Calculated CDF of Y corresponds to the CDF of exponential distribution with the rate parameter c, so, we conclude that $Y \sim E(c)$.

Method of moments

The moments of Weibull distribution:

$$E(X^n) = \frac{1}{c}^n \Gamma\left(1 + \frac{n}{\gamma}\right).$$

Let Y1, Y2, . . . , Yn be a sample from Weibull(α, β) distribution. The first two moments about origin are given by

$$\mu_1' = E(Y) = \beta^{\frac{1}{\alpha}}\Gamma(1+\frac{1}{\alpha}),$$

and

$$\mu_2' = E(Y^2) = \beta^{\frac{2}{\alpha}}\Gamma(1+\frac{2}{\alpha}).$$

The sample moments are given by,

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

and

$$m_1' = \frac{1}{n}\sum_{i=1}^{n} Y_i^2.$$

So,using method of moments, we have

$$\begin{aligned} m_1' &= \mu_1' \\ \Rightarrow \frac{1}{n}\sum_{i=1}^{n} Y_i &= \beta^{\frac{1}{\alpha}}\Gamma(1+\frac{1}{\alpha}) \end{aligned}$$

and

$$\begin{aligned} m_2' &= \mu_2' \\ \Rightarrow \frac{1}{n}\sum_{i=1}^{n} Y_i^2 &= \beta^{\frac{2}{\alpha}}\Gamma(1+\frac{2}{\alpha}). \end{aligned}$$

Solving for α and β we get the estimators

Method of percentiles: The first and third theoretical quartiles of Weibull distribution are

$$B_{25} = c^{-\frac{1}{\gamma}}(\log\frac{4}{3})^{\frac{1}{\gamma}}$$

$$B_{75} = c^{-\frac{1}{\gamma}}(\log 4)^{\frac{1}{\gamma}}$$

Let y_{25} and y_{75} be the first and third empirical sample quartiles. Now estimate the parameters by equating the

1st and 3rd empirical and theoretical quartiles.

$$B_{25} = y_{25}$$

$$B_{75} = y_{75}$$

Estimated parameters are:

$$\gamma = \log_{\frac{y_{75}}{y_{25}}}\left(\frac{\ln(4)}{\ln(4/3)}\right)$$

$$c = \frac{\ln(4/3)}{y_{25}^{\gamma}}$$

3 Task 29:

The given density is the density of a Gumbel distribution. We find the MLE with the usual procedure.

$$f(x) = \frac{1}{\sigma} e^{-\left(\frac{x-\mu}{\sigma}\right)}\ e^{-e^{-\left(\frac{x-\mu}{\sigma}\right)}}$$

$$L(\mu, \sigma|x) = \prod_{i=1}^{n} \frac{1}{\sigma}\ e^{-\left(\frac{x_i-\mu}{\sigma}\right)}\ e^{-e^{-\left(\frac{x_i-\mu}{\sigma}\right)}} = \left(\frac{1}{\sigma}\right)^{n}\ e^{-\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)}\ e^{-\sum_{i=1}^{n}\left(e^{-\frac{x_i-\mu}{\sigma}}\right)}$$

$$l(\mu, \sigma|x) = nlog\left(\frac{1}{\sigma}\right)\ - \sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)\ - \sum_{i=1}^{n}\left(e^{-\frac{x_i-\mu}{\sigma}}\right)$$

Partial derivatives being set to 0 give us:

$$\frac{dl}{d\mu} = 0 \implies \frac{n}{\sigma} - \frac{1}{\sigma}\sum_{i=1}^{n} e^{-\frac{x_i-\mu}{\sigma}} = 0$$

$$\frac{dl}{d\sigma} = 0 \implies -\frac{n}{\sigma} + \sum_{i=1}^{n}\frac{x_i-\mu}{\sigma^2} - \sum_{i=1}^{n} e^{-\frac{x_i-\mu}{\sigma}}\frac{x_i-\mu}{\sigma^2} = 0$$

Solving this equations give us estimations $\hat{\mu}$ and $\hat{\sigma}$

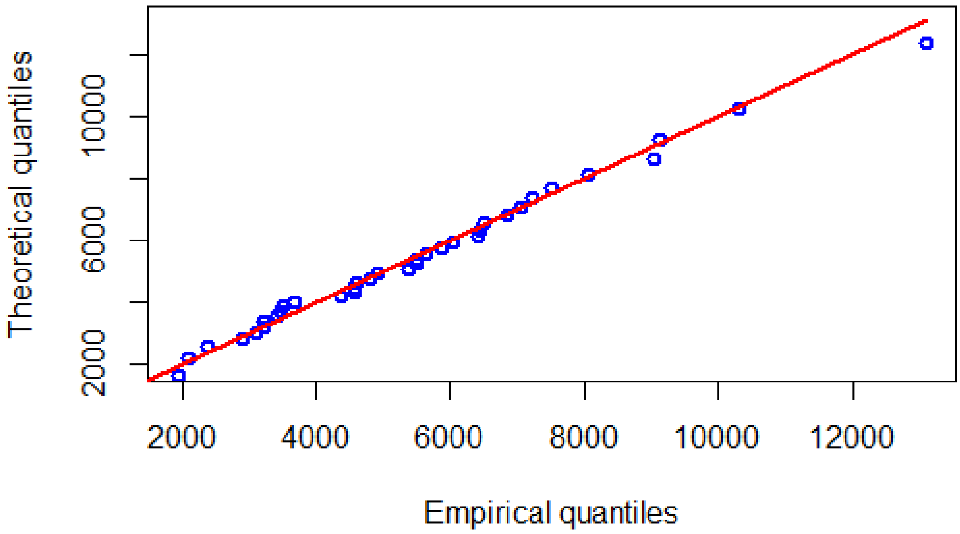
```
require(fitdistrplus)
require(reliaR)

  x <- c(5500, 4380 ,2370, 3220, 8050, 4560, 2100, 6840, 5640, 3500, 1940,
         7060, 7500, 5370, 13100, 4920, 6500, 4790, 6050, 4560, 3210, 6450,
         5870, 2900, 5490, 3490, 9030, 3100, 4600, 3410, 3690, 6420, 10300,7240, 9130)

  dgumbel <- function(x, a, b) 1/b*exp((a-x)/b)*exp(-exp((a-x)/b))
  pgumbel <- function(q, a, b) exp(-exp((a-q)/b))
  qgumbel <- function(p, a, b) a-b*log(-log(p))
  g <- fitdist(x,'gumbel',method = 'mle', start = list(a = 4000, b = 3000))
  summary(g)$estimate
  mu = summary(g)$estimate[1]
  sigma = summary(g)$estimate[2]


  qq.gumbel(x,mu,sigma)
```

$\hat{\mu} = 4393.654$ and $\hat{\sigma} = 1881.137$



We see the QQ-Plot looks actually quite good.

4 Task 31:

Let's find the probability that a claim exceeds 10 000:

$$P(x > 10000) = 1 - P(x \leq 10000) = 1 - F(10000) = 1 - (1 - e^{-10000\lambda}) = e^{10000\lambda}$$

where F(x) - CDF of exponential distribution.
Now let's calculate likelihood:

$$L(\lambda|x) = \prod_{1}^{68} \lambda e^{-\lambda x_i} * \prod_{69}^{80} e^{-10000\lambda} = \lambda^{68} e^{-\lambda \sum_{i=1}^{68} x_i} e^{-120000\lambda}$$

The log-likelihood is:

$$l(\lambda|x) = 68log(\lambda) - \lambda \sum_{i=1}^{68} x_i - 120000\lambda = 68log(\lambda) - 340000\lambda$$

Find MLE for $\hat{\lambda}$:

$$l'(\lambda|x) = \frac{68}{\lambda} - 340000 = 0 \to \hat{\lambda} = 0.0002$$

Calculate an approximate 95 % confidence interval for λ:

$$\hat{\lambda} \pm \frac{1.96}{\sqrt{nI_1(\hat{\lambda})}}$$

where

$$I_1(\hat{\lambda}) = \frac{d^2 l(\hat{\lambda})}{(d\hat{\lambda})^2} = \frac{68}{\hat{\lambda}^2}$$

Substituting n = 68, $\hat{\lambda} = 0.0002$ we have 95% confidence interval: (0.00019424;0.00020576)

5 Task 32:

5.1 a

$\hat{\mu}$ by MLE method is $\bar{x}$

Let $\bar{X}_i$ be a i-th bootstrap sample mean. Since in bootstrap procedure number of samples is rather big by the central limit theorem bootstrap samples mean $\bar{X}$ is normally distributed.

In particular:

$$\mathbb{E}\left[\bar{X}_i\right] = \mathbb{E}\left[\frac{1}{n}\sum_{j=1}^{n} X_{i,j}\right] = \frac{n\hat{\mu}}{n} = \hat{\mu}$$

$$\mathrm{Var}\left(\bar{X}_i\right) = \mathrm{Var}\left(\frac{1}{n}\sum_{j=1}^{n} X_{i,j}\right) = \frac{n\hat{\sigma}^2}{n^2} = \frac{\hat{\sigma}^2}{n}$$

5.2 b

Subtraction of the constant μ_0 from the RV $\hat{\mu}$ does not change the normality. Moreover, it does not influence the variance of the RV. So, we have

$$\mathrm{Var}\left(\hat{\mu} - \mu_0\right) = \mathrm{Var}\left(\hat{\mu}\right) = \frac{\hat{\sigma}^2}{n}$$

$$\mathbb{E}\left[\hat{\mu} - \mu_0\right] = \mathbb{E}\left[\hat{\mu}\right] - \mu_0 = \mu_0 - \mu_0 = 0$$

5.3 c

We know that $\hat{\mu} - \mu_0 \sim N(0, \frac{\hat{\sigma}^2}{n})$

Denote the $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of this distribution by δ_{low} and δ_{high} i.e.:

$$P(\hat{\mu} - \mu_0 \leq \delta_{low}) = \frac{\alpha}{2}$$

$$P(\hat{\mu} - \mu_0 \leq \delta_{high}) = 1 - \frac{\alpha}{2}$$

Then

$$P(\delta_{low} \leq \hat{\mu} - \mu_0 \leq \delta_{high}) = 1 - a$$

and from manipulation of the inequalities,

$$P(\hat{\mu} - \delta_{high} \leq \mu_0 \leq \hat{\mu} - \delta_{low}) = 1 - a$$

The confidence interval based on the t-distribution is:

$$\hat{\mu} \pm t_{n-1, \frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

where n - sample size, s - sample standard deviation, $t_{n-1, \frac{\alpha}{2}}$ - t - distribution value with n-1 degrees of freedom for the designed confidence level α

6 Task 33:

From the previous exercise we have bootstrap confidence interval:

$(\hat{\theta} - \delta_{high}, \hat{\theta} - \delta_{low})$ where , $\delta_{low}, \delta_{high}$ are $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$ quantiles of the distribution of $\hat{\theta} - \theta$

In our case $\delta_{high} = \theta_U - \hat{\theta}$ and $\delta_{high} = \theta_L - \hat{\theta}$

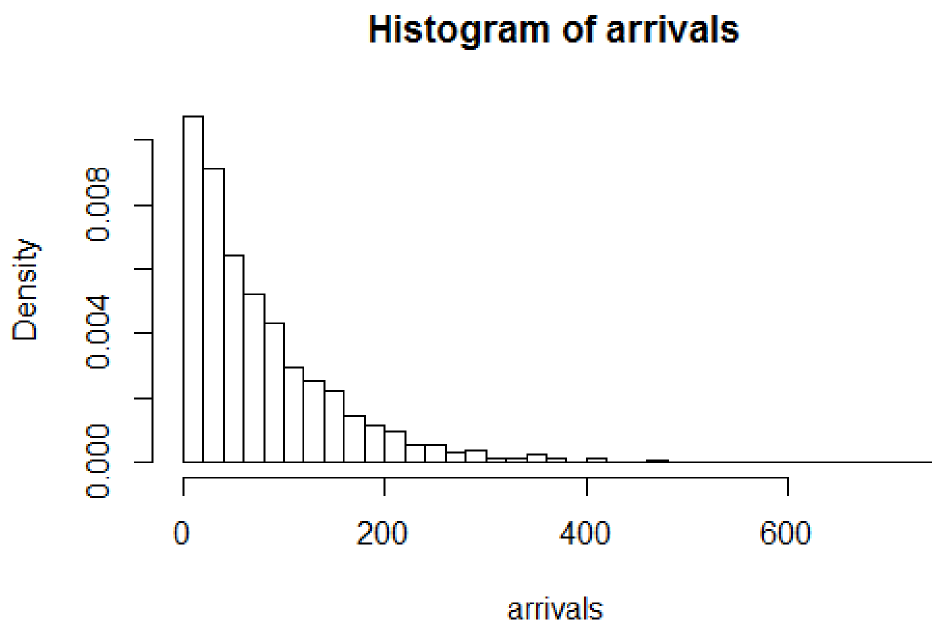
Then confidence interval is $(2\hat{\theta} - \theta_U, 2\hat{\theta} - \theta_L)$

In case of symmetric distribution we have $\hat{\theta} = \frac{\theta_U + \theta_L}{2}$

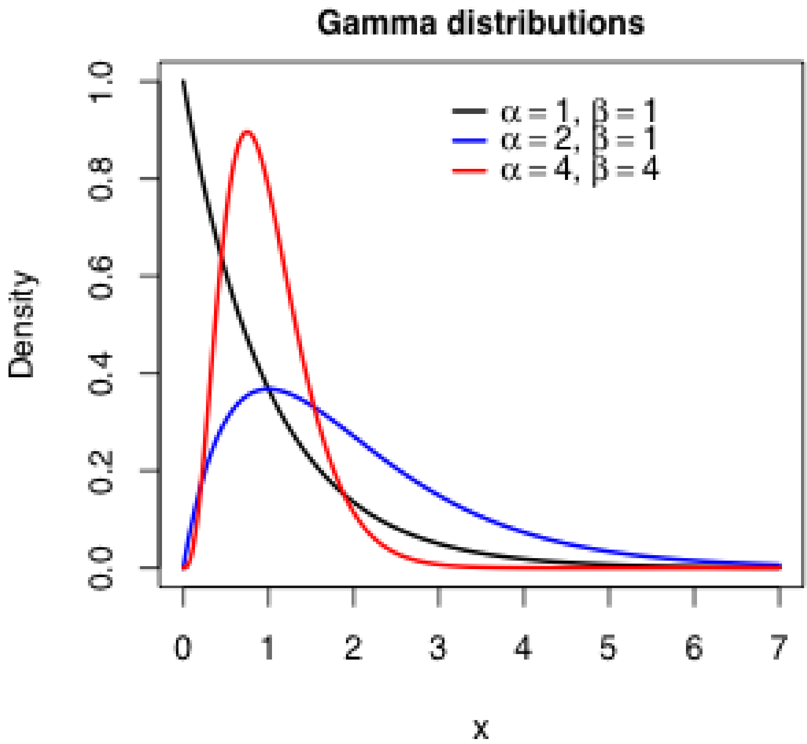
Substitution gives us exactly what we need: (θ_L, θ_U)

7 Task 34:

7.1 a)



Histogram of arrivals
Density
0.000
0.004
0.008
0
200
400
600
arrivals



Assumption seems to be reasonable. Gamma distributions can reflect a strong decreasing behavior of the data. Here we can see the same picture. The parameters gives necessary flexibility to fit the data rather good.

7.2 b)

Fit the parameters by the method of moments and by maximum likelihood. How do the estimates compare?

```
arrivals <- read.table("data.txt")
arrivals <- unname(unlist(arrivals))
hist(arrivals, breaks=30, probability = T)

gammaparameters <- fitdist(arrivals, "gamma", method = "mme")
GammaShapeMME <- summary(gammaparameters)$estimate[1]
GammaRateMME <- summary(gammaparameters)$estimate[2]
MME <- c(GammaShapeMME, GammaRateMME)
print(MME)


gammaparameters <- fitdist(arrivals, "gamma", method = "mle")
GammaShapeMLE <- summary(gammaparameters)$estimate[1]
GammaRateMLE <- summary(gammaparameters)$estimate[2]
MLE <- c(GammaShapeMLE, GammaRateMLE)
print(MLE)
```

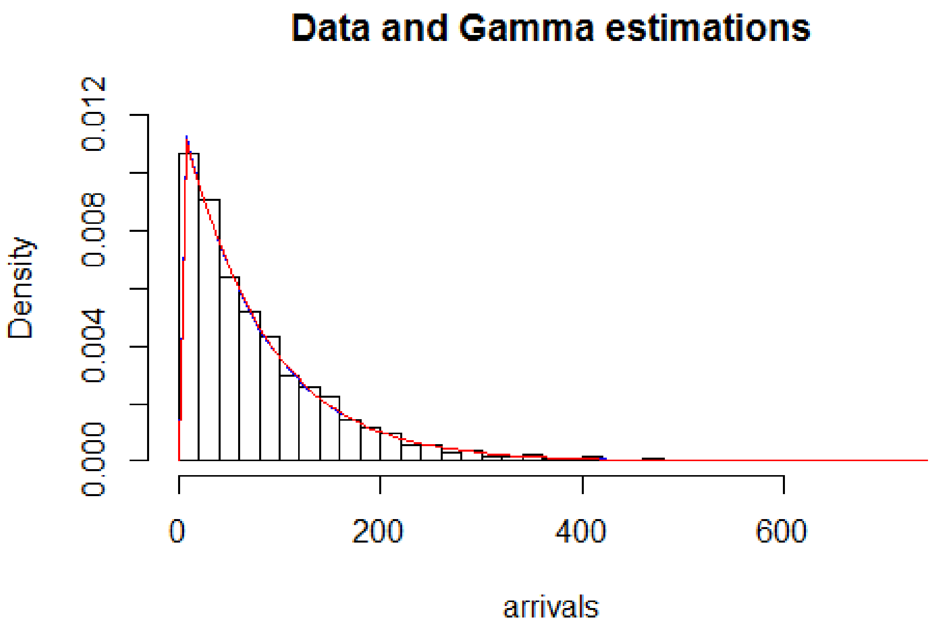
```
> print(MME)
     shape       rate
1.01235221 0.01266466
> print(MLE)
     shape       rate
1.02658382 0.01284249
```

As we can see the results of both methods are very close in values

7.3 c

Plot the two
fitted gamma densities on top of the histogram. Do the
ts look reasonable?

```
hist(arrivals, breaks=30,probability = T, ylim = c(0,0.012),
     main="Data and Gamma estimations")
x <- seq(0,600,by=0.1)
curve(dgamma(x, shape=MME[1], rate=MME[2]),
      add=TRUE, col='blue')
curve(dgamma(x, shape=MLE[1], rate=MLE[2]),
      add=TRUE, col='red')
```



Here is 2 lines (for moments and MLE) but they are basically the same.

7.4 d

For both maximum likelihood estimate and the method of moments, use the bootstrap to form approximate con

dence intervals for the parameters. How do the con

dence intervals for the two methods compare?

```r
require(matrixStats)
require(RBesT)
require(resample)
MOMsample <- matrix(rgamma(length(arrivals)*1000, shape=MME[1],
                           rate = MME[2]), nrow=length(arrivals), ncol=1000)
MOMshapes <- colMeans(MOMsample)^2/colVars(MOMsample)
MOMrates <- colMeans(MOMsample)/colVars(MOMsample)

MLEsample <- matrix(rgamma(length(arrivals)*1000,
                           shape=MLE[1], rate = MLE[2]), nrow=length(arrivals),
                    ncol=1000)
sampleMLEparameters <- numeric(1000)
for (i in 1:1000)
  sampleMLEparameters[i] <- fitdist(MLEsample[i,], "gamma", method = "mle")
MLEshapes <- numeric(1000)
MLErates <- numeric(1000)
for (i in 1:1000){
  MLEshapes[i] <- sampleMLEparameters[[i]][1]
  MLErates[i] <- sampleMLEparameters[[i]][2]}
head(MLEshapes)
head(MLErates)

  # shape MOM
MOM950s <- MME[1] - (sort(MOMshapes, decreasing = T)[950] - MME[1])
# rate MOM
MOM950r <- MME[2] - (sort(MOMrates, decreasing = T)[950] - MME[2])
# shape MLE
MLE950s <- MLE[1] - (sort(MLEshapes,decreasing = T)[950] - MLE[1])
# rate MLE
MLE950r <- MLE[2] - (sort(MLErates, decreasing = T)[950] - MLE[2])
c(MOM950s,MLE950s)
c(MOM950r,MLE950r)
```

```
> c(MOM950s,MLE950s)
   shape    shape
1.065345 1.088821
> c(MOM950r,MLE950r)
      rate       rate
0.01336404 0.01378168
```

We see confidence intervals which are very close.

7.5 e

It is consistent with a Poisson process model for the arrival times. The Gamma distribution is deeply intertwined with Poisson processes.

It follows that the time of the nth arrival has a Gamma(n, λ) distribution, since the sum of n i.i.d. Expo(λ) r.v.s is Gamma(n, λ).

$$P(T_n \leq t) = P(N_t \geq n) = P(G_n \leq m(t)),$$

where $G_n \sim \mathrm{Gamma}(n, 1)$ and $m(t)$ is as above.

Reference to the detailed Prove