PROJECT 2

Abstract

. The scope of this project is dealing with Chapters 9 of Bain, L.J. and Engelhardt M. Introduction to Probability and Mathematical Statistics 1992.

Question 1:

1. Problem Statement:

Summarize the three estimation methods (MME, MLE and Bayes) on one page.

Method 1 - MME:

ASSUME| $X_1, .., X_n$ is a random sample taken from a population with known distribution function $f(x; \theta_1, ..\theta_k)$

STEP 1: Determine the k-number of unknown parameters to estimate, in this case: $\theta_1, .., \theta_k$

STEP 2: Find the k^{th} population moment about the origin with: $\mu_k^{'} = E[X^k]$

STEP 3: Find the k^{th} observed sample moment with: $m_k^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i^j$

STEP 4: For 1 unknown parameter:

4.1: Find the MM_Estimate by setting the first population moment equal to the first observed sample moment with: $\widehat{\mu_1^{'}} = \overline{x}$ with $\overline{x}$ being the MM_Estimate for θ.

4.2: Give the MM_Estimator with: $\underset{\sim}{\widehat{\mu_1^{'}}} = \overline{X}$ with $\overline{X}$ being the MM_Estimator for θ.

STEP 5: For k unknown parameters:

5.1: Consider that for k-number of parameters we will need to evaluate k- number of population-and sample moments.

5.2: For each of the k-number of parameters find the MM_Estimates with $\widehat{\mu_k^{'}} = m_k^{'} = \frac{1}{n}\sum_{i=1}^{n} x_i^j$

5.3: Solve with substitution and or back-substitution, each of these k-number of equations to arrive at k-number of MM_Estimators of form: $\underset{\sim}{\widehat{\mu_k^{'}}} = m_k^{'} = \frac{1}{n}\sum_{i=1}^{n} X_i^j$

2) Method 2 - MLE:

ASSUME| $X_1, .., X_n$ is a random sample taken from a population with known distribution function $f(x; \theta_1, ..\theta_k)$

CASE 1| *The support of the distribution function is NOT bounded by the parameters:*

STEP 1: Find the Likelihood Function: $L(\theta_1, .., \theta_k) = \prod_{i=1}^{n} f(x; \theta_1, ..\theta_k)$

STEP 2: Find: $ln[L(\theta_1, .., \theta_k)]$

STEP 3.1: For 1 parameter, solve: $\frac{d}{d\theta} ln[L(\theta)] = 0$, for your ML_Estimate: $\widehat{\theta}(x)$ and ML_Estimator: $\underset{\sim}{\widehat{\theta}}(X)$

STEP 3.2: For 2 parameters, solve for $k = 1, 2$: $\frac{\partial^k L}{\partial \theta_k} ln[L(\theta_1, ..., \theta_k)] = 0$, for your ML_Estimates: $\widehat{\theta_1}(x)$, $\widehat{\theta_2}(x)$ and ML_Estimators: $\underset{\sim}{\widehat{\theta_1}}(X)$, $\underset{\sim}{\widehat{\theta_2}}(X)$

STEP 4.1: For 1 parameter, to show that our estimate provides a maximum value for $L(\theta)$ solve and show: $\frac{d^2}{d\widehat{\theta}^2} ln[L(\widehat{\theta})] < 0$

CASE 2| *The support of the distribution function is bounded by the parameter:*

STEP 1: ...repeat Step 1 and Step 2 of Case 1.

STEP 2.1: For 1 parameter, we need to determine which values of the parameter, given the observed values from the random sample, will maximize $L(\theta)$ or $ln[L(\theta)]$ while taking into account the support of the distribution function. This will yield the ML_Estimate and ML_Estimator.

STEP 2.2: For 2 parameters, with the support of the distribution function only bounded by 1 parameter for example θ_1. Firstly we keep θ_2 constant and determine for which values of θ_1 we will get a maximum for $L(\theta_1, \theta_2)$, this will yield the first ML_Estimate $\widehat{\theta}_1$. Secondly we keep θ_1 constant and solve for the second ML_Estimate, $\widehat{\theta}_2$, with $\frac{d}{d\theta_2} ln[L(\theta_1, \theta_2)] = 0$.

3) Method 3 - Bayes:

ASSUME| $X_1, .., X_n$ is a random sample taken from a population with known distribution function

STEP 1: Set $f_{X|\theta}(x|\theta) :=$ distribution of the sample $X_1, .., X_n$

STEP 2: Set $p(\theta) =$ density function of the parameter which will be our Prior Density

STEP 3: Since $\widehat{\theta_B} = E_{\theta|\underline{X}}(\theta)$ we first need to find the Posterior Distribution of θ with:

$f_{\theta|\underline{X}}(\theta|\underline{x}) = \frac{f_{\underline{X}|\theta}(\underline{x}|\theta)p(\theta)}{\int_a^b f_{\underline{X}|\theta}(\underline{x}|\theta)p(\theta)d\theta}$ with the support of θ being between a and b.

STEP 4: Since $f_{\theta|\underline{X}}(\theta|\underline{x})$ depends only on θ we know we only have to solve $f_{\theta|\underline{X}}(\theta|\underline{x}) \propto f_{\underline{X}|\theta}(\underline{x}|\theta)p(\theta)$

STEP 5: Since we need the likelihood of n number of observations we set $f_{\underline{X}|\theta}(\underline{x}|\theta) = \prod_{i=1}^{n} f_{X|\theta}(x|\theta)$,

then solving and reducing to see which kernel of a known distribution it matches up to.

STEP 6: Now we can find $E_{\theta|\underline{X}}(\theta)$ since we have all the information we need.

Question 2:

1. Problem Statement:

Suppose we have a sample x from the model $f(x;\lambda,\theta) = \frac{\lambda}{\theta}(\frac{x}{\theta})^{\lambda-1}e^{-(\frac{x}{\theta})^{\lambda}}$ $\forall x > 0$. Assume $\theta = 1$ is fixed and use λ_0 as the true value of λ (Choose it yourself). Use a simulation study to calculate the point estimate of λ. You must use MME, MLE and Bayes with an Exponential prior for λ such that *a priori*, $E(\lambda) = \lambda_0$. Use the mode of the posterior density as the Bayes point estimate.Show graphically the maximum of the likelihood on the likelihood function as well as the mode of the posterior.

2. Methodology and Design of Simulation Study:

Now the supplied model from which the sample is taken is ~$WEI(\theta = 1, \lambda)$ distributed with λ the parameter under estimation. Each of the 3 methods will be considered separately:

2.1 MME Approach:

SINCE| $f(x;\lambda,\theta) \sim WEI(1,\lambda) \implies \mu_1^{'} = E(X) = \Gamma(1+\frac{1}{\lambda}) \ldots$ the first population moment

ALSO| $m_1^{'} = \overline{x}$

THEN| $\overline{x} = \Gamma(1+\frac{1}{\lambda})$

```
theta=1
n=100
lambda=rs_1=x=trueMeanVal=trueMean=simMean=simMean1=0

trueMeanFun = function (r){
  for(i in 1:r){
    trueMeanVal=1+1/i
    trueMean[i]=gamma(trueMeanVal)
  }
  return(trueMean)
}

set.seed(1492)
simMeanFun = function(q){
  for(i in 1:q){
    simMean[i] = 0.01*sum(rweibull(100, shape=i, scale=1))
  }
  return(simMean)
```

```
}
set.seed(1492)
simMeanFun1 = function(q){
  for(i in 1:q){
    simMean1[i] = 0.001*sum(rweibull(1000, shape=i, scale=1))
  }
  return(simMean1)
}
colors <- c("Mean Estimate" = "red", "simSampleMean for n=100" = "blue",
            "simSampleMean for n=1000" = "black")
xf=1:100
df = data.frame(xf, trueMeanFun(100), simMeanFun(100), simMeanFun1(100))

ggplot(df, aes(xf)) +
  geom_line(aes(y=trueMeanFun(100), color = "Mean Estimate")) +
  geom_line(aes(y=simMeanFun(100), color = "simSampleMean for n=100")) +
  geom_line(aes(y=simMeanFun1(100), color = "simSampleMean for n=1000")) +
  labs(title = "Estimate of the Sample Mean vs Simulated Sample Mean",
       x = "Gamma Scale Parameter",
       y = "Sample Mean for Given Scale Parameter",
       caption = "Figure 1",
       color = "Legend") +
  scale_color_manual(values = colors)+
  theme(plot.title = element_text(hjust = 0.5, face = "bold"),
        plot.caption = element_text(hjust = 0.5, colour = "blue"),
        legend.justification = c("right", "bottom"),
        legend.position = c(.95, .15),
        legend.box.just = "right",
        legend.margin = margin(6, 6, 6, 6),
        axis.line = element_line(size = 0.5, colour = "black", linetype=1))+
  scale_x_continuous(expand = c(0, 0), limits = c(0,100), oob = scales::censor) +
  scale_y_continuous(expand = c(0, 0), limits = c(0.8,1), oob = scales::censor)
```

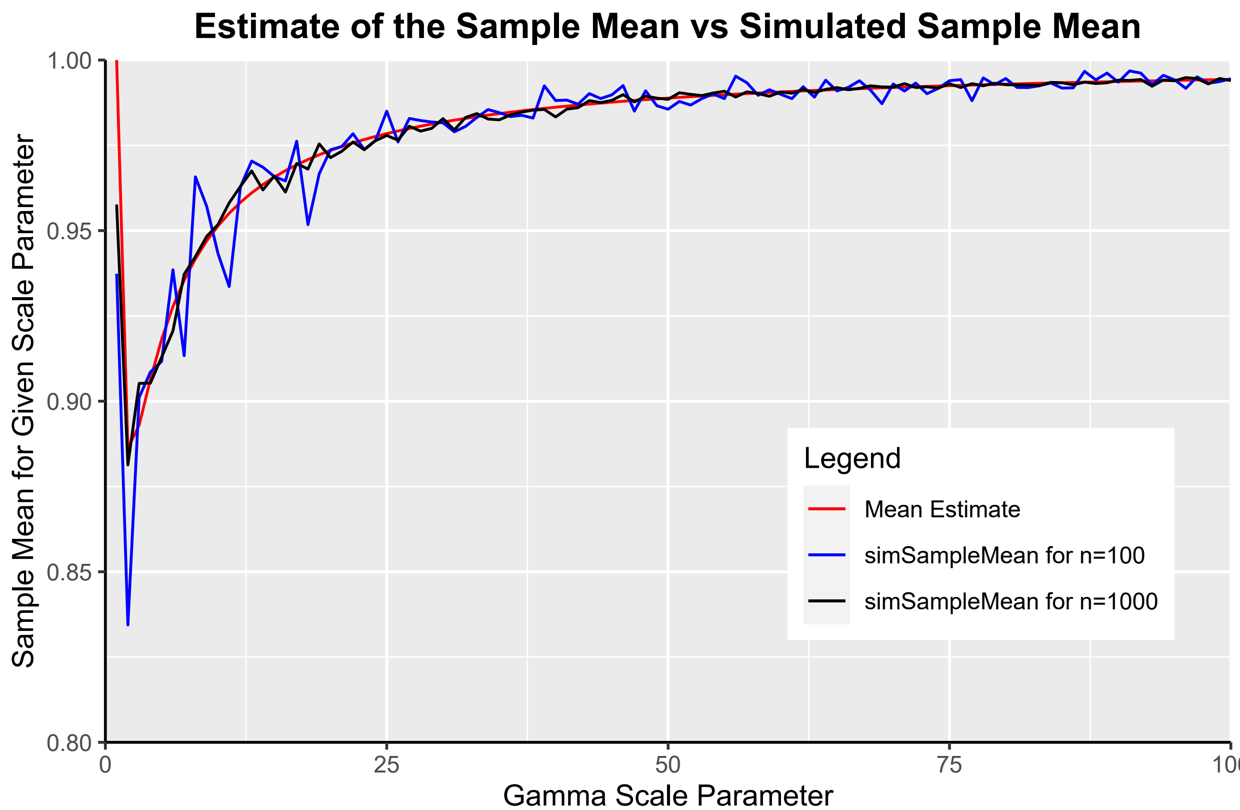

Figure 1

2.2 MLE Approach:

```
n_samples = 100000;

sampleData = rweibull(n = n_samples, shape=0.75, scale=1)

rate_fit_R = fitdist(data = sampleData, distr = 'weibull', method = 'mle')

print(rate_fit_R)
```

```
## Fitting of the distribution ' weibull ' by maximum likelihood
## Parameters:
##        estimate  Std. Error
## shape 0.747427 0.001841256
## scale 1.001723 0.004463314
```