ETC 2420/5242 Lab 4 2016

Di Cook

Week 4

Purpose

This lab is to examine different statistical distributions, fit distributions to samples by estimating the parameters by maximum likelihood and checking the fit with QQ-plots.

Reading

Read the material on maximum likelihood estimation at https://onlinecourses.science.psu.edu/stat414/node/191.

Read the code in the lecture notes from Week 3. Particularly look at the functions for making QQ-plots, computing and plotting the likelihood functions.

Warmup exercises

- Compute these probabilities for $X \sim N(2, 0.5)$
  - $P(X < 1.3)$
  - $P(X > 1.9)$
  - $P(1.8 < X < 2.2)$
- Compute the quantile value (X) for $X \sim N(-3, 4)$ which matches these probabilities
  - $P(X < x) = 0.53$
  - $P(X < x) = 0.12$
  - $P(X < x) = 0.84$
  - $P(X < x) = 1.2$
- Compute the value of the density function for a $N(-1.2, 0.8)$ corresponding to $X =$
  - -1.0
  - -0.2
  - -2.0
- Plot the density curves a
  - $Weibull(2.5, 2)$
  - $Weibull(2.5, 1)$
  - $Weibull(1, 1)$ on the same plot.

Question 1

1. Simulate samples of size $n = 30$, 100, 500 from these distributions
   a. Lognormal(2, 0.5)
   b. Gamma(2, 4)
2. Make a QQ-plot of each these samples.

Question 2

Using this code, generate a sample of size $n = 267$ from a $Gamma(1.2, 0.25)$ distribution.

```
set.seed(123)
X2 <- data.frame(x=rgamma(n=267, 1.2, 0.25))
```

a. Plot the sample, using a histogram, describe the shape of the distribution.
b. What parameters of the gamma distribution were used to simulate the sample? (α, β)
c. If we are to use maximum likelihood distribution what values would we expect to get as the parameter estimates?
d. Write a function to compute the likelihood function.
e. Plot the likelihood function for a range of values of α, β that shows the maximum likelihood estimates for each parameter.
f. Look up the function `fitdistr` from the `MASS` library. Explain what this does. Use it to find the MLE estimates for α, β. How do these compare with the values you read off your plot?

Question 3

Take a look at the data set `danishuni` from the `CASdatasets` library.

```
library(CASdatasets)
data(danishuni)
```

a. Make a histogram of the `Loss`. Describe the shape.
b. Fit both a gamma and lognormal distribution to the sample, i.e. find the MLEs.

c. Produce a QQ-plot for each of the distributions.
d. Which is the better fit to the sample?
e. Re-do steps a-d after log-transforming the `Loss`.

TURN IN

- Your `.Rmd` file
- Your Word (or pdf) file that results from knitting the Rmd.
- Make sure your group members are listed as authors, one person per group will turn in the report
- DUE: Wednesday after the lab, by 7am, loaded into moodle

Resources

- PSU lecture notes on MLE
- CASdatasets