

# ETC 2420/5242 Lab 4 2017

*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  $X \sim N(3.3, 1.1)$ 
  - $P(X < 1.3)$
  - $P(X > 1.9)$
  - $P(1.8 < X < 2.2)$
- Compute the quantile value ( $x$ ) for  $X \sim N(-10, 4)$  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  $N(12, 5)$  corresponding to  $X =$ 
  - 13.0
  - 4.0
  - 20.0
- Plot the density curves a
  - $Weibull(3, 1.5)$
  - $Weibull(2, 2)$
  - $Weibull(1, 1)$  on the same plot.

## Question 1

- Simulate samples of size  $n = 30, 100, 500$  from these distributions
  - Lognormal(4, 2)
  - Gamma(3, 3)
- Do an internet search to find an example of where a lognormal distribution, and a gamma distribution might be used.
- Make a QQ-plot of each these samples. Explain how closely the samples, of different sizes, appears to match the theoretical distribution.

## Question 2

Using this code, generate a sample of size  $n = 544$  from a  $Gamma(3.2, 1.7)$  distribution.

```
set.seed(123)
X2 <- data.frame(x=rgamma(n=544, 3.2, 1.7))
```

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

## Question 3

Take a look at the data set `usworkcomp` from the `CASdatasets` library. Read the documentation about this data on <http://cas.uqam.ca/pub/R/web/CASdatasets-manual.pdf>.

A copy of the data is provided with the lab, in case the `CASdatasets` are not all available.

- Make a histogram of the `LOSS`. Describe the shape.
- Fit both a gamma and lognormal distribution to the sample, i.e. find the MLEs.
- Produce a QQ-plot for each of the distributions.
- Which is the better fit to the sample?

## TURN IN

- Your `.Rmd` file
- Your html 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

## Resources

- PSU lecture notes on MLE
- `CASdatasets`