

Project #4: Fitting probabilistic models

Milica Cudina

2021-10-02

```
#library(nimble)
```

Problem #1. (10 points)

Read in the data from the “mystery-data.csv” file. Your goal is to fit a probabilistic model to the data and visually substantiate your choices. More precisely, you need to do the following:

- **(2 points)** visually display your data;
- **(2 points)** propose a **named parametric distribution** to fit to your data and justify your choice;
- **(2 points)** using the data, propose a point estimate for any parameters in your model;
- **(4 points)** superimpose the appropriate graph for your model onto the appropriate graph of the data to convince your reader that your model is valid.

Problem #2 (14 points)

Here, the goal is to provide a probabilistic model for some of the data available here:

[Data set associated with the textbook](#)

First, download the data set and import it into R. Do not forget that the documentation for the data set is also available under the above link.

(2 points) Focus on the measurements of the male respondents’ calves’ maximum girth in centimeters. Does this set of measurements have any outliers?

(4 points) If there are outliers, discard them from your data set. Visualize the remaining data points appropriately.

(2 points) Propose a **named parametric distribution** to fit to your data and justify your choice.

(2 points) Using the data, propose a point estimate for any parameters in your model.

(4 points) Superimpose the appropriate graph for your model onto the appropriate graph of the data to convince your reader that your model is valid.

Problem #3 (22 points)

Here, the goal is to provide a probabilistic model for some of the data available here:

[Data set associated with the textbook](#)

First, download the data set and import it into R. Do not forget that the documentation for the data set is also available under the above link.

We want to construct a probabilistic model for the number of severe earthquakes in a year based on this data set.

(6 points) Create a `table` which contains the number of times that a particular yearly number of earthquakes was recorded in the above data set.

(2 points) We must not forget that in the years not mentioned in the data set, there were no earthquakes. Add this information to the table you created above. Do not forget to include the appropriate name to the added observation.

(2 points) Choose an appropriate plot and visualize the data in the table you obtained above.

(6 points) Now, we want to construct a probabilistic model for these data. When it comes to counts of rare events, the **Poisson distribution** is oftentimes used. Before you proceed, review the Poisson distribution (Section 4.5 in the textbook and/or your *M362K Probability* notes). Then, familiarize yourself with the built-in commands related to the Poisson distribution in R. You can easily access the `help` by typing `dpois` in the console in RStudio. Now, fit the Poisson distribution to the data by estimating its parameter.

(6 points) Superimpose the appropriate graph for your model onto the appropriate graph of the data to convince your reader that your model is valid.

Problem #4 (14 points)

Go to [Google historical data](#)

Then, download the data for the last 252 (or so) trading days, i.e., for the last year.

(2 points) Draw the time-plot of the evolution of the closing stock price (not the adjusted). You do **not** need to put the calendar days on the horizontal axis, but you **do** need to label your axes and give your time-plot a title indicating the dates.

The **simple daily return** of the stock over a day indexed by t is defined as

$$\frac{\text{price at end of day } t - \text{price at end of day } (t - 1)}{\text{price at end of day } (t - 1)}$$

(4 points) Construct the vector of simple daily returns over the last year. Provide a visualization of the returns. What can you say about the characteristics of the distribution based on the above plot?

(3 points) Fit the normal distribution to the above returns. Superimpose the appropriate graph for your model onto the appropriate graph of the data to convince your reader that your model is valid.

(5 points) You will have to install a package to solve this part of your project. First, run the following in your console:

```
install.packages('nimble')
```

When that is finished, you need to run this in your un-comment

```
library(nimble)
```

from the first chunk in this document.

Next, you should learn more about the **Laplace (double exponential)** distribution. This is easily done by visiting:

[Wikipedia: The Laplace distribution](#)

Now, you are equipped to fit the Laplace (double exponential) distribution to the above returns. To learn about the parametrization of the Laplace distribution in R, type `?ddexp` into the console in RStudio.

After you have completed the fit, superimpose the appropriate graph for your model onto the appropriate graph of the data.