ST2137 AY 24/25 Sem 2

Assignment 1 (R portion)

Introduction

This assignment covers topics 1 to 4. The questions on this pdf correspond to the *R portion*. The dataset can be found on Canvas. Remember to also solve and submit the Python portion for this assignment!

For this R portion, you may want to look up functions such as `readLines` and `strsplit`. They may help you read the data in. For this assignment, you are not allowed to use any additional packages other than `lattice`.

Aircraft breakdowns

The file `aircraft_failure.txt` contains information on the time between breakdowns of 10 aircrafts. Each line in the file corresponds to a particular aircraft. For instance, aircraft number 9 first broke down 418 hours after it was commissioned. Following that repair, it broke down 18 hours later again, and so on.

1. Read the data into R and create a dataframe named `ftimes_df` with two columns. Here are the first few rows of the dataframe:

```
  aircraft failure_times
1        1           413
2        1            14
3        1            58
4        1            37
5        1           100
6        1            65
```

2. Create a lattice plot of histograms for each aircraft. Ensure that the plot has a title and proper axis labels.

3. Consider fitting a method of moments (MoM) gamma estimator to each aircraft's failure time. If X has a $\Gamma(\alpha, \beta)$ distribution, then its pdf is

$$f_X(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, \quad x > 0, \quad \alpha, \beta > 0$$

   In terms of the parameters, the mean and variance of a gamma distribution are:

$$E(X) = \alpha\beta, \quad Var(X) = \alpha\beta^2$$

   The MoM estimator is one method for obtaining an initial, decent estimate of parameters. These estimates are found by equating the first k sample moments to the corresponding k population moments. In the above case, for an i.i.d sample $x_1, \dots, x_n$, we set:

$$\sum_{i=1}^{n} x_i = \bar{x} \quad = \quad E(X) = \alpha\beta \tag{1}$$

$$\sum_{i=1}^{n} x_i^2 = \bar{x} \quad = \quad E(X^2) = \alpha\beta^2(1+\alpha) \tag{2}$$

Solving for α and β in the above equations returns the MoM estimators. Write a function in R, named `mom_gamma` that will return the MoM estimator. Here is how it should work:

```
ac1 <- ftimes_df$failure_times[ftimes_df$aircraft == 1]
mom_gamma(ac1)
```

```
    alpha        beta
0.6727961 142.2357426
```

4. One of the common assumptions when deriving estimators is that the sample is i.i.d. One check we can make for this assumption is to compute correlation of each vector *on itself!* Suppose that we have a sample of values $x_1, \ldots, x_n$. We can then compute Pearson correlation on the pairs

$$(x_1, x_2),\ (x_2, x_3),\ (x_3, x_4),\ \ldots (x_{n-1}, x_n)$$

This is known as the autocorrelation with lag 1. Compute this autocorrelation for all 10 aircrafts and store them in a vector `ac_vec`.