18.05 Problem Set 6, Spring 2022

Problem 1. (10 pts.) **Continuous MLE**
Suppose we have a distribution with the following pdf (called a gamma distribution)

$$f(x|a) = \frac{a^5}{(4)!} x^4 \mathrm{e}^{-ax}.$$

Suppose we have independent data x_1, x_2, ..., x_m drawn from this distribution. Find the maximum likelihood estimate (MLE) for a.

Suggestion: The likelihood can be compactly written in terms of the sum S and the product P of the data.

Problem 2. (25: 5,10,10 pts.) **Least squares**
In this problem we will use maximum likelihood estimates to develop Gauss' method of least squares for fitting lines to data.

Bivariate data means data of the form

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n).$$

For bivariate data the simple linear regression model assumes that, for some values of the parameters a and b, we have

$$y_i = ax_i + b + \text{ random measurement error.}$$

The model assumes the measurement errors are independent and identically distributed and follow a $\mathrm{N}(0, \sigma^2)$ distribution. (The values x_i may or may not be random.)

In general terms, we can say that the value of x 'explains' the value of y except for some random noise. Graphically, the model says to make a scatter plot and find the line that best fits the data. This is called a simple linear regression model.

It turns out that, under some assumptions about random variation of measurement error, one way to find a "best" line is by solving a maximum likelihood problem.

The goal is to find the values of the model parameters a and b that give the MLE for this model. To guide you, we note that the model says that

$$y_i \sim \mathrm{N}(ax_i + b, \sigma^2).$$

Also remember that you know the density function for this distribution.

(a) For a general datum (x_1, y_1) give the likelihood and log likelihood functions (these will be functions of y_1, x_1, a, b, and σ.)

(b) Consider the data $(1, 8)$, $(3, 2)$, $(5, 1)$. Assume that $\sigma = 3$ is a known constant and find the maximum likelihood estimate for a and b.

Note: since there are *two* variables a and b, in order to find a critical point you will have to take partial derivatives and set them equal to 0. This part of the problem takes a fair amount of tedius algebra –sorry.

Note: We gave you a specific value of σ, to avoid the distraction of one more symbol. If you look at your calculations, you should see that the value of σ plays no role in finding the MLE for a and b. We get the same answer no matter what the value.

(c) Use R to plot the data and the regression line you found in part (ii) The commands `plot(x,y, pch=19)` and `abline()` will come in handy. For `abline` be careful: the parameter `a` is the intercept and `b` is the slope – exactly the opposite of our usage. Print the plot and turn it in.

Problem 3. (15: 10,5 pts.) **Estimating uniform parameters**

(a) Suppose we have data `1.2, 2.1, 1.3, 10.5, 5` which we know is drawn indepenedently from a uniform(a, b) distribution. Give the maximum likelihood estimate for the parameters a and b.

Hint: in this case you should not try to find the MLE by differentiating the likelihood function.

(b) Suppose we have data $x_1, x_2, \ldots, x_n$ which we know is drawn indepenedently from a uniform(a, b) distribution. Give the maximum likelihood estimate for the parameters a and b.