Problem n.1

The Quality Control Manager (QCM) of a factory production sector is responsible for determining if screws produced by a machine are of acceptable quality (designated as "OK" or 1) or not (designated as "KO" or 2), based on their diameter in millimeters. The machine is old and sometimes goes into Fault (F), causing a higher proportion of "KO" screws to be produced. The manager believes that:

- when the machine works at Capacity (C), there is a 0.7 probability of producing an "OK" screw and there is a probability of 0.3 that the machine switches to the Fault state (F);

- when the machine is in the Fault state (F), there is a 0.2 probability of producing an "OK" screw and there is a probability of 0.55 that the machine returns to working at Capacity (C).

By modelling the described situation through a Hidden Markov Model, answer to the following questions:

a) Which are, in your opinion, the hidden states of the model and why? Report the Transition and Emission matrices related to the tentative guess of the QCM.

On the 7th of February 2023, from 8:30 am to 8:35 am, 50 screws were produced and the QCM attributed to them the following labels ("OK"=1, "KO"=2), in sequence: 1, 1, 2, 2, 2, 2, 1, 1, 2, 1, 2, 1, 1, 2, 1, 2, 2, 2, 1, 2, 2, 2, 2, 2, 1, 2, 2, 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 1, 2, 1, 2, 1, 1, 1, 2, 1, 1, 2, 1, 2.

b) What is the most probable path of hidden states of the machine that generated the observed sequence of labels? How do you estimate it? *[In the estimation, use defaults parameters and assume equal start probabilities]*

c) Estimate the Emission matrix and Transition matrix from the observed sequence of labels.

Between 4:40 pm and 4:45 pm, 50 additional screws were produced by the machine. The quality control employee, lacking the experience of the QCM, is unaware of the transition probabilities. However, he is knowledgeable in Statistics and he assumes that:

- when the machine operates at Capacity (C), the screw diameters are distributed according to a Gaussian $N(\mu_C, \sigma_C^2)$;

- when the machine is in a Faulty state (F), the screw diameters are distributed according to a Gaussian $N(\mu_F, \sigma_F^2)$.

The employee has recorded the sequence of diameter measurements in the file `diameters.txt`.

d) What is the most probable path of hidden states of the machine that generated the sequence of observed screw diameters? How do you estimate it? *[Before the estimation, write `set.seed(1)`]*

Remark: be careful with packages conflicts!

Upload your solution here