

Question 4 [30 points]

Many people in Montreal take the metro as part of their travel around the city. In this question, you will write code to simulate travel times to be used in an app to predict travel times for their routes.

The idea of the statistical part of the app is the following:

- The number of stops and the number of lines vary between individuals, but each user also knows how many metro stops they need to travel on each line.
- Each user will need to tell the app how many lines they are taking for their trip and for how many stops they will be on each line that they take.
- The app (which you're not writing) will collect this information and give it to you in a certain format (see the questions for details) to use in your simulations.
- Your goal is to simulate different possible trips to compute the 10th and 90th percentile values for their journey and return them to the app.

We will assume the time between metro arrivals can be generated in the following way: each time between metro arrivals (including doors opening and closing, passenger exchange, any delays) is either EXACTLY 2 minutes with probability 0.8 **or**, with probability 0.2, is equal to 2 minutes plus a delay time that is randomly distributed according to a Gamma distribution with shape parameter 5 and scale parameter 0.5. Your function should take a single argument, the number of stops that the passenger will be on the line (not including the station where they begin, but including the station where they end).

- (a) **[10 points]** Write a function below, `one_line`, which simulates the length of time spent on a single line.

Hint: You can use the `sample` or `rbern` function in R to determine whether the time between each arrival is exactly 2 minutes or 2 minutes plus a random delay. You can use the `rgamma` function in R to generate the possible delays at each stop.

CONTINUED ON NEXT PAGE

- (b) **[10 points]** Write a new function, 'one_trip' which uses your function in part (a) to simulate the amount of time taken for one entire **trip**, i.e. over all lines. Your function should take in one argument: a vector which is of length equal to the number of lines that someone will take on their trip where each element of the vector is equal to the number of stops they will take on that line. You do not need to account for the time to walk from one platform to the others.

Your function should return a single value, the total amount of time for their trip.

CONTINUED ON NEXT PAGE

- (c) **[10 points]** Write a new function, 'lower_and_upper' which uses your function in part (b) to simulate a certain number of trips and returns the 10th and 90th percentile values, to provide a realistic interval for how long the passenger should expect their trip to take. The function should take in three arguments, the random seed for the simulation, a vector which is of length equal to the number of lines that someone will take on their trip where each element of the vector is equal to the number of stops they will take on that line, and the number of the trips that should be simulated in order to estimate the 10th and 90th percentiles.