ECS 132 Fall 2020: Assignment 5 - 10 pts

November 7, 2020

Instructions

1. You may in no circumstances upload your homework to private tutoring websites such as CourseHero or Chegg. Remember all material related to this course is a property of the University of California and posting them is a violation of the copyright laws.
2. If you refer to a source (either a book or the internet), you must cite it.
3. You are highly urged to work on these problems on your own. See the homework grading policy. Not getting the answer correct has very low penalty. However, trying it and then figuring out where you went wrong will really help you understand the material better and you will be much better prepared for the exams. If you do discuss with others, you must list their names.
4. Write your answers in R Markdown and submit the knitted pdf on Gradescope; for due date and other details see the Homework Policy and Schedule.

1 Problem

In the lecture we discussed preliminary data analysis and used the inter-eruption time of the Old_Faithful data. Answer the following questions for the eruption duration (the last column of the data set which is the third column.) For this question, go ahead and read the lectures as well as R documentation regarding boxplot and quantiles.

1. Plot the frequency histogram of the eruption duration?
2. Draw the boxplot of the eruption duration?
3. What are the values for the 95, 97, 99 quantiles of the eruption duration?
4. Suppose we classify the eruption duration using the following simple rule: if the duration is less than or equal to 3 mins then we classify it as a short eruption otherwise (i.e., of if the duration is greater than 3 mins) it is a long eruption. Draw a scatter plot that compares each eruption duration with the following one. Suppose there are n data points $e[1] \dots e[n]$. (You can find the number of data points using len(data[,3]) see lecture notes). Plot $(x = e[i], y = e[i + 1])$ for $i = 1 \dots n - 1$. Note that both the X and Y axes are time. On the scatter plot now draw horizontal and vertical lines at 3 mins. These two line divide the area into 4 parts corresponding to long eruption followed by a long eruption, a long by a short, a short by a long, and short by a short. Summarize your observation.

5. Find the probabilities that a long eruption is followed by a long eruption, a long by a short, a short by a long, and short by a short.

1.1 Answer

```
data = read.table(file="/Users/dghosal/Documents/ECS132/Fall2020/Data/
    Old_Faithful.txt", header=TRUE)

par(mfrow=c(2,2))


hist(data[,3], xlab = "Bin", ylab = "Frequency Count", main = "Frequency
    Histogram of Eruption Duration")


boxplot(data[,3], main = "Boxplot of Eruption Duration")


q = c(.95, .97, .99)
quantile(data[,3], q)



n = length(data[,3])

currentd = numeric(n-1)
nextd = numeric(n-1)
pss = 0
psl = 0
pll = 0
pls = 0

for (i in seq(1,n-1,1)) {
  currentd[i] = data[i,3]
  nextd[i] = data[i+1,3]
  if ((currentd[i] >= 3.0) & (nextd[i] >= 3.0))
      {pll = pll + 1}
    else if ((currentd[i] >= 3.0) && (nextd[i] < 3.0))
       {pls = pls + 1}
    else if ((currentd[i] < 3.0) && (nextd[i] < 3.0))
       {pss = pss + 1}
    else
       {psl = psl + 1}
}

plot(currentd,nextd, xlab = "Current Duration", ylab = "Next Duration")

mtext(c("Short","Long"),side=3,line=2,at=c(2.5,4))
mtext(c("Short","Long"),side=2,line=2,at=c(2.5,4))
abline(v=3)
abline(h=3)
```

```
sprintf("Probability of Long Long is %f", pll/(n-1))
sprintf("Probability of Long Short is %f",pls/(n-1))
sprintf("Probability of Short Long is %f", psl/(n-1))
sprintf("Probability of Short Short is %f", pss/(n-1))
```

1.2 Points - 7 pts

- 1 pt - Properly import data and setup in R
- 1 pt (part 1) for properly generating a histogram
- 1 pt (part 2) for properly generating the boxplot
- 1 pt (part 3) for computing the 95th, 97th, and 99th quantiles
- 2 pt (part 4): 1 pt for setting up the problem as described, 1 pt for properly generating the figure and summarizing
- 1 pt (part 5) for identifying probabilities of long-short, long-long, short-short, short-long (0.25pt each)

2 Problem

This question deals with the Poisson process and is formulated in terms of a subway station. It could as well be formulated in terms of a networking problem. A subway station where different train lines intersect is like switch/router in the communication network with the different train lines corresponding to what are called Labelled Switched Paths (LSPs). As for the trains, you can think of them as packets or burst of packets. One can draw this analogy also in optical networks which forms the core of the Internet backbone.

Two one-way subway lines, the A train line and the B train line, intersect at a transfer station, A trains and B trains arrive at the station according to independently operating Poisson processes with rates $\lambda_A = 3\ trains/hr$ and $\lambda_B = 6\ trains/hr$. We assume that passenger boarding and un-boarding occurs almost instantaneously. At a random time, Bart, a prospective passenger, arrives at the station, awaiting an A train. [Note that the superposition of two independent Poisson processes with rates λ_1 and λ_2 is also a Poisson process with rate $\lambda_1 + \lambda_2$.]

- What is the probability that the station handles exactly 9 trains during any given hour?
- If an observer counts the number of trains that the station handles each hour, starting at 8:00 A.M. on Tuesday, what is the expected number of hours until he or she will first count exactly 9 trains during an hour that commences ”on the hour”? (e.g., 9: 00 A.M., 10: 00 A.M., 2: 00 P.M.)

2.1 Answers

- The combined process $N(t)$ of the A trains and B trains is a Poisson process with rate $\lambda_A + \lambda_B = 9$. The probability that there are exactly 9 trains in 1 hour is given by

$$\begin{aligned} P\{N(1) = 9\} &= e^{-\lambda}\frac{\lambda^9}{9!} \\ &= e^{-9}\frac{9^9}{9!} \\ &= 0.132 \end{aligned}$$

- Let success correspond to the event if there are exactly 9 trains in any given hour starting on the hour. The probability of success denoted by p is 0.132 as computed in the previous problem. The probability of failure which correspond to the event that the number of trains in a hour is any number other than 9 and this is $1 - p = 0.868$. Now each hour can be considered as a Bernoulli random variable. Furthermore, if we consider a sequence of hours we have a sequence of Bernoulli random variable which are independent. The independence comes from the fact that the underlying arrival process if Poisson. In a Poisson process the number of events in non-overlapping intervals are independent.

  The number of hours, (which is the number of trials) denoted by N until first success (observe exactly 9 trains) is geometrically distributed, i.e.,

  $$P\{N = n\} = p^{n-1}p \quad n = 1, 2, \ldots, \infty$$

  and we can show that the $E[N] = 1/p$. Thus, the expected number of hours is 7.6 hours.

2.2 Points - 3 pts

- 1 pt for properly setting up system
- 1 pt (part 1) for properly computing probability of 9 trains at one station in an hour
- 1 pt (part 2) for computing expected number of hours until 9 trains arrive in an hour (starting on the hour). Give yourself full credit if you correctly computed the expectation for a different interpretation of the waiting time - but please make sure you understand the correct interpretation from the solution.