Indian Institute of Technology Dharwad, Karnataka, India

EE 201 / EE 227: Data Analysis / Data Analysis (Second Half)

Assignment (Autumn 2021)

Date: 09/10/2021

Time Limit: 9:00 a.m. to 12:00 noon, 3 Hours + 10 more minutes for submission

---

Instructions

1. Make suitable assumptions when necessary and state them clearly in your solutions.

2. Please write your own solutions. Any kind of copying will fetch zero marks for the entire assignment.

3. Submission is on Google Classroom. Late submissions will fetch zero marks regardless of how much late they are. So, plan your submissions properly.

Questions

1. (10 points) **Remember Hot Wheels?:** Let's go back to your childhood days, where you had some small cars to play with. At your school, your teacher draws a huge circle on the ground whose radius is R. Without loss of generality let us say that the centre of the circle is the origin. You start your toy car at any point on the y-axis chosen uniformly randomly such that the car definitely meets the circle. The car moves in parallel to the x-axis. You measure the distance covered by the toy car within the circle. Determine the average distance covered
   (a) (7 points) when the y-axis value is chosen to be a random variable.
   (b) (3 points) when the angle at which you begin is chosen to be a random variable.

2. (10 points) : **As promised:**
   (a) (5 points) Give a proof for the central limit theorem using moment generating functions (MGFs).
   (b) (5 points) Show that the correlation coefficient is between -1 and +1.

3. (10 points) **Special functions of random variables:**
   (a) (5 points) From the class, you have learned that for a random variable X and a function $g(.)$, $Y = g(X)$ is also a random variable. Now, let X is any continuous RV and I choose the function $g(.)$ to be the CDF of X, i.e., $Y = F_X(X)$. Determine the CDF and PDF of Y.
   (b) (5 points) Let $Z = X^2 + Y^2$, where X and Y are i.i.d Gaussian random variables with mean μ and variance σ^2. Find the CDF and PDF of Z.

4. (10 points) **Conditionally lost:** You take a walk in Dharwad for the first time and are lost at a point. There are three roads from that point. The first one brings you back to the same point after 10 minutes. The second road brings you back to the same point after 1 hour. The third road will take you to the city bus stand (CBT) in 20 minutes. Assume that your choice of picking a road is equally likely and you conveniently forget your past choices. Determine the average time taken for you to reach CBT

   (a) (5 points) using conditional expectations.

   (b) (5 points) by modelling the entire problem as a random variable and then finding the mean.

5. (10 points) **Bengaluru transport:** You have secured a job in the silicon valley of India, Bengaluru, and choose to reach office through public transport. You are at a bus-stop and buses to your destination arrive randomly. Now a large set of people are doing the same thing as you, thanks to the ever increasing number of people landing with a job in Bengaluru. Hence, you always end up taking the third bus. Let X_1 denote the time (in minutes) before the first bus arrives, X_2 be the interval between the arrivals of the first and second buses, and X_3 be the interval between the arrival of the second and third buses. Assuming X_i's to be i.i.d uniform random variables between 0 and 10 minutes, find the upper bounds to the probability that your wait time exceeds 25 minutes using the Markov, Chebyshev and Chernoff bounds (bonus 5 marks if you find the value of s minimizing the Chernoff bound by writing code or any other method)

6. (10 points) **Estimating a pass/fail course:** Suppose you are all taking a course where you are graded +1 and -1 for "pass" and "fail", respectively and the grades are equally likely. Let $X_1, X_2, \cdot X_n$ denote the grades of n people in the class. You have designed two estimators for "pass" - a) $\hat{\theta_1} = X_1 + 1$ (yes, just the first person's grade + 1!!) and b) $\hat{\theta_2} = \max\{X_1, X_2, \cdot X_n\}$. Determine whether these estimators are biased and consistent.