CE93—Engineering Data Analysis
Midterm 1 (10/13)

Answer any four questions. Show your work. Partial credit generously given.

1. Telegraph Avenue merchants are saying that they are losing business because it takes too long for people to find parking; thus, they want the city to build a new parking structure. You, the city parking expert, don't think there is currently any problem finding parking, but you need more than a gut feeling to convince the city council that a new structure is a bad idea. You interview 48 drivers as they step out of their cars and ask them to state how long it took them to find parking. Luckily, they all had stopwatches with them, so they knew to the minute and second how long they were searching for a parking spot. The following table lists the search time (in **minutes:seconds**). Note that the values are listed in the ascending order.

0:30	2:14	3:05	4:35
0:35	2:16	3:07	4:37
0:37	2:30	3:14	5:08
1:01	2:38	3:28	5:11
1:03	2:46	3:33	5:24
1:15	2:52	3:34	5:43
1:48	2:59	4:05	6:03
1:52	2:59	4:17	6:22
2:01	3:01	4:23	6:58
2:02	3:03	4:28	8:15
2:08	3:03	4:29	15:02
2:13	3:03	4:30	23:10

a) *Sketch* a cumulative frequency plot of these data. Make sure to identify both axes. Include at least four points on the plot, including the point where it reaches the value of 1.

b) Find the median time to find parking. What is the 0.125-quantile value of search time? What is the value of search time at the third quartile? What is the interquartile range of the data?

c) Now assume that you need to divide the data into categories: excellent (0-2 minutes), good (2-4 minutes), fair (4-8 minutes), unacceptable (8-25 minutes). Draw a histogram of the data using these categories (bins).

d) Suppose that a new structure needs be built if more than 25% of drivers experience unacceptable delay (as defined in part c.) in searching for parking. Given this, does the structure need to be built?

2. Professor Fungus, chair of the Civil Engineering Department at Mold Miss University, wants to make sure that the undergraduate majors in his department are getting enough exercise—he's heard that few have time left over for this because of the demands of course work and drinking. He commissions a survey of CE majors, and finds

- 40% exercise zero days a week
- 0% exercise one day a week
- 20% exercise two days a week
- 20% exercise three days a week
- 10% exercise four days a week
- 0% exercise five days a week
- 0% exercise six days a week
- 10% exercise seven days a week

The number of exercise days a week can be thought of as a discrete random variable, *C*.

a. Determine the expected value, mode, and median of *C*. Describe in words what each of these values mean.
b. Sketch the PMF and CDF for *C*.
c. Calculate the variance and coefficient of variation of *C*.
d. Is *C* positively or negatively skewed?
e. If all of the current non-exercisers (40% of the total) suddenly started exercising 1 day a week, would the mode change? Would the mean remain the same, increase, or decrease? What about the median? Please briefly explain your reasoning.

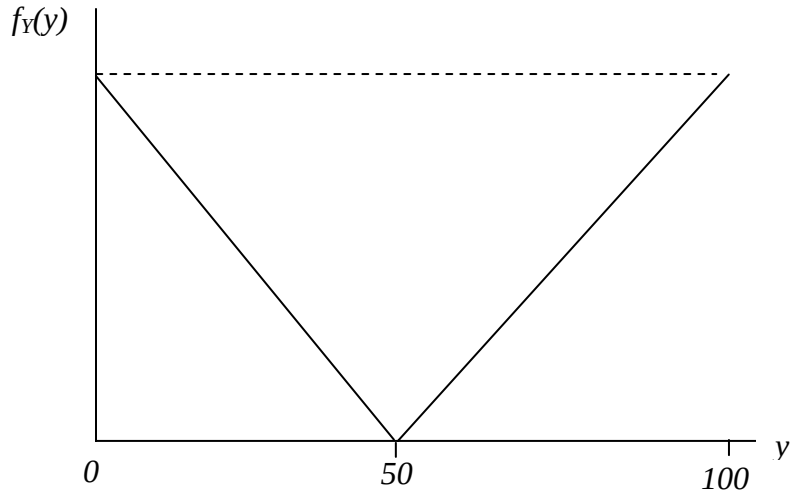
3. The probability of a person passing the Air Traffic Controllers exam is .6 if she has an Associates' degree in air traffic control, and .2 if she does not have such a degree. Of the people taking the exam, 20% have the degree.

a) What is the probability of passing the exam?
b) What is the probability that a person who has passed the exam does <u>not</u> have the degree?

4. An ambitious but slightly incompetent entry level engineer is asked to find information about hourly maximum wind speed in an area where a skyscraper is going to be built. He obtains many years of hourly data, and decides to summarize it by developing a function, $N(s)$, which gives the number of hours where the maximum hourly speed (that is the maximum instantaneous wind speed measure for that hour) was *greater* than s. He finds that this function is $N(s) = 70000(1-\frac{\sqrt{s}}{7})$. He e-mails this result to you, his supervisor, and then dies in an accident on a space trip he won by correctly guessing the number of commas in Jane Austen's *Pride and Prejudice*. You are thus left with the task of trying to make sense of his results. As a first step, you find from other sources that the proportion of hours with no wind $(s = 0)$ is 0.2.

a) Use the information provided to develop the CDF for S, $F_S(s)$.
b) What is the PDF for S, $f_S(s)$, for s>0?
c) Explain why $f_S(0)$ is undefined.
d) What is the median of *S*?

5. The PDF for final exam scores (ranging from 0 to 100) for a large class has the shape shown by the solid line below.



a) What is the formula for this PDF?
b) What is the formula for the associated CDF?
c) What is the sign of the skewness coefficient? Explain your reasoning.
d) What is the probability that four students, chosen at random, each receive a score between 25 and 75 on the exam?
e) Explain, in words, what seems to have happened in the final exam.

6. A new transportation system has two kinds of vehicles with seating capacities 2 and 4. They become available to the dispatcher at a terminal in mixed trains having one to two cars. If both of the possible train lengths are equally likely and the two vehicles appear independently and in equal relative frequencies, what is the probability that exactly four seats will be available for dispatch in a particular train?