ST 501 Sample Midterm Exam

This is a list of sample problems for Midterm 1 for ST 501. The real exam is **closed book** and **closed notes**. However, you are allowed to bring one A4 page, double sided, formula/fact sheet. The real exam paper will also include a formula/fact sheet like the one at the end of this document (note that there are things on the fact sheet that will not be relevant to the current exam)

- The exam time is 90 minutes. There will be around 5 problems.
- A calculator is **not required** but you are welcome to bring one if you wish.
- These sample questions is meant to give you a rough idea of the format or type of questions you might see in the exam. The questions on the real exam are expected to be different, although we will try to make it so that the difficulty of the sample problems reflect, to some extent, the difficulty of the problems on the real exam (there will thus be a mix of easier problems and somewhat more complicated problems).
- Show all of your work, as detailed as you can. You do not need to provide exact numbers/answers, but the answers should be sufficiently detailed so that we can be sure that you know them. Note that if you do not show your work or explain your reasoning then it is much harder for us to grade your work; we might not be able to give you partial/full credits, even in cases where your answers are almost correct.
- The exam will cover Chapter 1, section 2.1, and parts of sections 4.1 and 4.2 of the textbook (the parts related to discrete distributions). You are expected to know the **binomial** and **geometric** distributions.
- Solution sketches are provided at the end of this document.

Question 1 (15pts)

A girl scout is selling cookies door-to-door. On each day, she can choose to visit either one house with probability $1/4$ or to visit two houses with probability $3/4$. For each house, she either makes a sale with probability 0.6 (or fails to make a sale with probability 0.4). If the sales are **independent** events and each sale nets her \$20 in revenue, what is the **expected** amount of money she will earn on a randomly selected day ?

Question 2 (10pts)

There is an urn with n balls in it. Each ball has a **distinct** number painted on it. The value on each ball is either $1, 2, \dots, n$. For example, if $n = 3$ then the urn contains 3 balls that are painted with the numbers 1, 2, and 3. You take one ball at random from the urn. Let X be the value painted on this ball. Find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$ (for the variance expression, simplify as much as you can).

Question 3 (10pts)

A fleet of twelve taxis is to be dispatched to four airports. Call these airports A, B, C, and D.

- (a) (5pts) Suppose three taxis go to airport A, four taxis go to airport B, three taxis go to airport C, and two taxis go to airport D. How many distinct ways can we dispatch the taxis ?
- (b) (5pts) Suppose there are **two** taxis that need repairs. If the taxis are dispatched to airports completely at random, what is the probability that airport A and airport B each receive **exactly** one of the taxis that need repair ? You should still assume that three taxis go to airport A, four to airport B, and so on like in part (a).

Question 4 (15pts)

Suppose a long-haul airplane has four engines and needs **at least** three **working** engines in order to fly. Another airplane has two engines and needs **at least** one working engine to fly. Suppose that the engines are **mutually independent** and each engine has a **constant** probability p of staying functional during a flight.

(a) (5pts) Find the probability that the the four engines plane completes the flight
(b) (5pts) Find the probability that the two engines plane completes the flight
(c) (5pts) Compare the probabilities in part (a) and part (b) for several values of p, for example, $p = 0.2, 0.5$ and 0.8. Which plane do you think is more reliable ?

Question 5 (10pts)

n friends are at a restaurant. They play a game of "odds man out" where, for each round, they each toss a **fair** coin. If at any given round there is **one** person whose coin have a different face (from his/her $n-1$ other friends) then the game stop and that person have to pay the restaurant bill. Otherwise the game continues. What is the probability that the game lasts for **exactly** four rounds ? For example, if $n = 5$ and the tosses for a round is either $HHHTT$ or $HHHHH$ then the game continues. Otherwise, if, for example, the tosses for a round is $HHHHT$ then the game ends **immediately** and the person whose toss is a T will pay the bill.

Question 6 (10pts)

A certain blood test for a disease gives a positive result 99% of the time among patients having the disease. But it also gives a positive result 1% of the time among people who do not have the disease. It is believed that 2% of the population has this disease. What is the probability that a person with a positive test result indeed has the disease ?

Question 7 (10pts)

Suppose n numbers are drawn uniformly at random from $\{1, 2, \ldots, N\}$. Let $k \in \{1, 2, \ldots, N\}$ be **given**

(a) (5pts) What is the probability that the maximum is k if the numbers are drawn **with replacement**?
(b) (5pts) What is the probability that the maximum is k if the numbers are drawn **without replacement**? (You can assume that $k \geq n$ for this part of the problem)

Hint: Let X denote the rv for the maximum among the n numbers and note that $P(X = k) = P(X \leq k) - P(X \leq k - 1)$. Calculations of $P(X \leq k)$ is much easier.

Question 8 (10pts)

An urn contains four red balls and four green balls. These balls are taken out of the urn at random, one at a time and **without replacement**. Let X be the random variable for the draw at which a green ball is **first** taken out. For example, if $X = 3$ then the first two balls drawn are red balls and the third ball drawn is a green ball. Write down the values of the **probability mass function** for X and compute $\mathbb{E}[X]$.

Question 9 (20pts)

At a large university, students are classified as being either conservative, moderate, liberal, or anarchist, with each student being pigeon-holed into **exactly** one category. It is estimated that

- 25% of the students are classified as conservative.
- 35% of the students are classified as moderate.
- 30% of the students are classified as liberal.
- 10% of the students are classifed as anarchists

There is an election going on between two candidates A and B.

- Given that a randomly selected student is a conservative, the probability that the student prefers candidate A is 0.4.
- Given that a randomly selected student is a moderate, the probability that the student prefers candidate A is 0.5
- Given that a randomly selected student is a liberal, the probability that the student prefers candidate A is 0.2
- Given that a randomly selected student is a anarchists, the probability that the student prefers candidate A is 0.9

From the above information, answer the following questions.

(a) (10pts) What is the probability that a **randomly selected** student prefers candidate A ?

(b) (10pts) Suppose we randomly choose 200 students who prefer candidate A. What is the **expected** number of conservatives we have (among this group of 200 students) ?

Question 10 (10pts)

There are 20 bears at the zoo, of which 12 are black bears and 8 are grizzly bears. A zookeeper randomly chooses 6 bears to train for a circus act.

(a) (5pts) What is the probability that all 6 bears that are chosen are from the **same** species ?
(b) (5pts) What is the probability that there are **at most** 2 grizzly bears among the 6 chosen bears ?

Question 11 (10pts)

Suppose that an electrical system consists of 10 components that are connected sequentially (one after another). Each component has two units, a **main** unit and a **backup** unit. A given component is non-operational if and only if **both** the main and backup units fail. The main and backup units are **independent** and each unit fails with probability $p = 0.1$. The system is operational if all 10 connected components are operational (see figure below for a notional depiction of the system). What is the probability that the system is operational ? (You can assume that the 10 components are mutually independent).

Basic results in probability

Let Ω be a sample space and P a probability measure on Ω.

- $P(A) = 1 - P(\bar{A})$ for any event $A \subset \Omega$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ for any events $A, B \subset \Omega$.
- $P(A) = P(A \cap B) + P(A \cap \bar{B})$ for any events $A, B \subset \Omega$.
- Law of total probability

  $$P(A) = \sum_i P(A \cap B_i) = \sum_i P(A \mid B_i) \times P(B_i)$$

  for any **partition** $\{B_i\}$ of the sample space Ω.
- Bayes Law:

  $$P(B_i \mid A) = \frac{P(A \mid B_i) \times P(B_i)}{\sum_j P(A \mid B_j) \times P(B_j)}$$

- Independence: If A and B are independent then $P(A \cap B) = P(A) \times P(B)$
- De Morgan's law.

  $$(A \cup B)' = A' \cap B', \quad (A \cap B)' = A' \cup B'$$

- Distributive law.

  $$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

Basic facts about discrete random variables

- If X is a **binomial** random variable with n trials and success probability p then

  $$P(X = x) = \tbinom{n}{x} p^x (1-p)^{n-x}, \quad \mathbb{E}[X] = np, \quad \mathrm{Var}[X] = np(1-p).$$

- If X is a **negative** binomial random variable for the number of trials at which the rth success **first** occurs and p is the success probability for each trial then

  $$P(X = x) = \tbinom{x-1}{r-1} p^r (1-p)^{x-r}, \quad \mathbb{E}[X] = \frac{r}{p}, \quad \mathrm{Var}[X] = \frac{r(1-p)}{p^2}.$$

  The case when $r = 1$ is known as the **geometric** distribution.

Solution Sketches

Q1. Let X be the number of sales made. Then $X \in \{0, 1, 2\}$ and furthermore

$$P(X = 0) = \frac{1}{4} \times 0.4 + \frac{3}{4} \times 0.4^2 = 0.22,$$
$$P(X = 1) = \frac{1}{4} \times 0.6 + \frac{3}{4} \times 2 \times 0.6 \times 0.4 = 0.51,$$
$$P(X = 2) = \frac{3}{4} \times 0.6^2 = 0.27.$$

For example, $X = 2$ if the girl scout visit two houses (probability 3/4) and **conditional on visiting two houses**, makes a sale at both houses (with probability $0.6^2 = 0.36$).

The expected amount of money the girl scout will make is then

$$\mathbb{E}[X] = \$20 \times P(X = 1) + \$40 \times P(X = 2) = \$21.$$

Q2. This problem is the discrete uniform distribution in disguise (Problem 4.2 from your textbook). Here X takes on the values $\{1, 2, \dots, n\}$ with $P[X = 1] = P[X = 2] = \cdots = P[X = n] = 1/n$. We thus have

$$\mathbb{E}[X] = \sum_{k=1}^{n} kP[X = k] = \sum_{k=1}^{n} k \times \frac{1}{n} = \frac{1}{n}\sum_{k=1}^{n} = \frac{1}{n} \times n(n+1)/2 = \frac{n+1}{2},$$
$$\mathbb{E}[X^2] = \sum_{k=1}^{n} k^2 P[X = k] = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6n} = \frac{(n+1)(2n+1)}{6},$$
$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \frac{(n-1)(n+1)}{12}.$$

Q3. For part (a), we use the multinomial coefficients. We can think of the airports as the four sets. The number of ways of assigning 12 elements to 4 sets with the sets having 3, 4, 3, and 2 elements is then

$$\binom{12}{3, 4, 3, 2} = \frac{12!}{3!4!3!2!} = 277200.$$

There are thus 277200 ways of assigning the 12 taxis to the 4 airports.

For part (b), the number of ways of assigning the taxis so that airport A and airport B each gets exactly one broken taxi is

$$\binom{2}{1} \times \binom{10}{2, 3, 3, 2} = \frac{2! \times 10!}{2!3!3!2!} = 50400.$$

In other words, first we choose, among the two broken taxis, which of them to sent to airport A (the other taxi will be send to airport B). There are now 10 taxis left, of which we need to send $3 - 1 = 2$ to airport A, $4 - 1 = 3$ to airport B, and 3 and 2 to airport C and D.

The **probability** of having exactly one taxi that need repair at each of the airport A and B is then

$$\frac{50400}{277200} = \frac{2}{11}.$$

Q4. For part (a), the four engine plane completes the flight if either all four engines work or, 3 out of the 4 engines work. As the engines are independent, this probability is

$$p^4 + 4p^3(1 - p) = 4p^3 - 3p^4.$$

For part (b), the two engine plane completes the flight if either both engines work or, one of the two engines work. This probability is then

$$p^2 + 2p(1 - p) = 2p - p^2.$$

Substituting $p = 0.2, 0.4, 0.6$ and 0.8 into the above expression for the four engines plane yield the probabilities

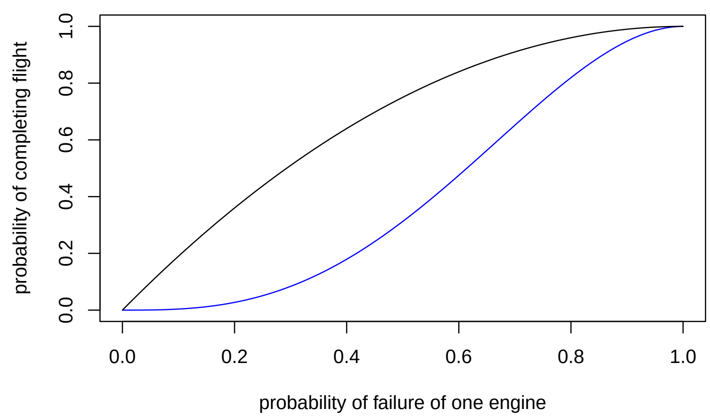
$$0.0272, \quad 0.1792, \quad 0.4752, \quad 0.8192$$

The corresponding probabilities for the two engines plane are

$$0.36, \quad 0.64, \quad 0.84, \quad 0.96$$

The two engines plane is thus more reliable. You can also plot the two probabilities for various values of p as follows. We see that the two engines plane (black line) is always more reliable than the four engines plane (blue line).

```
pseq <- seq(from = 0, to = 1, by = 0.01)
plot(pseq, 4*pseq^3 - 3*pseq^4, type = "l", col= "blue",
     xlab = "probability of failure of one engine",
     ylab = "probability of completing flight")
lines(pseq, 2*pseq - pseq^2, type = "l", col = "black")
```



Q5. In any given round, the probability that the game ends is $n2^{-n}$ as there are n different possibilities to select one person among the n people to get a different face then the remaining $n-1$ people (and the probability of having a H or a T are the same, i.e, $1/2$). The probability that the game continues is then $1 - n2^{-n}$. Therefore, the probability that the game lasts exactly four rounds is

$$\left(1 - n2^{-n}\right)^3 \times n2^{-n}$$

Q6. Let T be the event that the test result is positive and D be the event that the person has the disease. By Bayes' theorem, we have

$$\mathbb{P}[D \mid T] = \frac{\mathbb{P}[T \mid D] \times \mathbb{P}[D]}{\mathbb{P}[T \mid D] \times \mathbb{P}[D] + \mathbb{P}[T \mid D^c] \times \mathbb{P}[D^c]} = \frac{0.99 \times 0.01}{0.99 \times 0.01 + 0.01 \times 0.99} = \frac{1}{2}.$$

Q7. Let X be the random variable for the maximum value among the n selected numbers. For part (a), as the sampling is **with replacement** we have

$$P(X \leq k) = \frac{k^n}{N^n}.$$

We can interpret this expression as follows. $X \leq k$ if each of the selected numbers are less than or equal to k. The number of possible choices to select n numbers, each of which is less than or equal to k, is simply k^n.

The total number of choices to select n numbers from $\{1, 2, \dots, N\}$ is N^n. Given the above expression for $P(X \leq k)$, we have

$$P(X = k) = P(X \leq k) - P(X \leq k-1) = \frac{k^n - (k-1)^n}{N^n}.$$

For part (b), as the sampling is **without replacement** and we had assumed that $k \geq n$, we have

$$P(X \leq k) = \frac{\binom{k}{n}}{\binom{N}{n}}.$$

We therefore have

$$P(X = k) = P(X \leq k) - P(X \leq k-1) = \frac{\binom{k}{n} - \binom{k-1}{n}}{\binom{N}{n}}.$$

Q8. We note that as the balls are sampled **without replacement**, we can not use a **binomial** distribution or a **geometric** distribution (also the maximum possible value for X is 5 and so X surely cannot be a geometric random variable).

For this problem we just have to list all the possibilities for X. As the balls are sampled without replacement, X can take on any values in $\{1, 2, 3, 4, 5\}$. Now $X = 1$ if the first selected ball is green. This happens with probability $4/8 = 1/2$. Next $X = 2$ if the first selected ball is red and the second selected ball is green. This happens with probability $4/8 \times 4/7 = 2/7$. Continuing the same pattern, we have

$$\begin{aligned}
P(X = 3) &= P(\text{first is red, second is red, third is green}) = \frac{4}{8} \times \frac{3}{7} \times \frac{4}{6} = \frac{1}{7}, \\
P(X = 4) &= P(\text{first three is red, fourth is green}) = \frac{4}{8} \times \frac{3}{7} \times \frac{2}{6} \times \frac{4}{5} = \frac{2}{35}, \\
P(X = 5) &= P(\text{first four is red}) = \frac{4 \times 3 \times 2 \times 1}{8 \times 7 \times 6 \times 5} = \frac{1}{70}.
\end{aligned}$$

From the above values, we have

$$\mathbb{E}[X] = 1 \times \frac{1}{2} + 2 \times \frac{2}{7} + 3 \times \frac{1}{7} + 4 \times \frac{2}{35} + 5 \times \frac{1}{70} = 1.8.$$

Q9. For part (a), we apply the law of total probability, i.e.,

$$\begin{aligned}
P(\text{prefer A}) &= P(\text{prefer A} \cap \text{conservative}) + P(\text{prefer A} \cap \text{moderate}) + P(\text{prefer A} \cap \text{liberal}) + P(\text{prefer A} \cap \text{anarchists}) \\
&= 0.4 \times 0.25 + 0.35 \times 0.5 + 0.3 \times 0.2 + 0.1 \times 0.9 = 0.425
\end{aligned}$$

For part (b), let X denote the random variable for the number of students who are conservatives (among the 100 students who prefer candidate A). Then X is a binomial random variable with $n = 100$ trials and "success probability" p where now

$$p = P(\text{conservative} \mid \text{prefer A}) = P(\text{conservative} \cap \text{prefer A})P(\text{prefer A}) = \frac{0.4 \times 0.25}{0.425} = \frac{4}{17} \approx 0.235.$$

That is to say, $X \sim \text{Bin}(n = 200, p = 4/17)$. It is important to note here that we are taking a random sample of $n = 200$ people who prefers candidate A and we are interested in the number of conservatives among these 200 people, as such the probability of success p is now a **conditional probability**, i.e., p is the probability that a person is conservative given that the person prefers candidate A. We therefore have

$$\mathbb{E}[X] = np = 200 \times \frac{4}{17} \approx 47.$$

Q11. For part (a) either the bears are all black bears, of which there are $\binom{12}{6}$ possibilities, or all grizzly bears, of which there are $\binom{8}{6}$ possibilities. There are $\binom{20}{6}$ ways to choose 6 bears and the probability of interest is

$$\frac{\binom{12}{6} + \binom{8}{6}}{\binom{20}{6}} = \frac{952}{38760} \approx 0.02$$

For part (b) the number of ways to choose at most 2 grizzly bears is

$$\binom{12}{6} + \binom{12}{5} \times \binom{8}{1} + \binom{12}{4} \times \binom{8}{2} = 21120$$

and so the probability of interest is $\frac{21120}{38760} \approx 0.54$.

Q12. Let A_i be the event that the ith component works for $i \in \{1, 2, \ldots, 10\}$. We are interested in $P(A_1 \cap A_2 \cap \cdots \cap A_10) = P(A_1) \times P(A_2) \times \cdots \times P(A_10)$ as these events are **mutually independent**. Now $P(A_i) = 1 - 0.1^2 = 0.99$ for any i (as the system fails only when both the main and the backup unit fails, and each happens with probability 0.1 and are independent). Therefore the probability that the system works is

$$(1 - 0.1^2)^{10} \approx 0.905$$