Homework 1 (due Thursday May 19th)

1. (25 pts) Consider the Normal distribution $\mathcal{N}(\mu, \sigma^2)$ that we mentioned in class and a sample $X_1, \ldots, X_n$ of size $n$ obtained from this distribution. Which of the following are statistics and why?

   (a) $\bar{X}$
   (b) $\frac{\sqrt{n}(\bar{X}-\mu)}{s}$
   (c) $\prod_{i=1}^{n} X_i$
   (d) $\max\{X_1, \ldots, X_n, \sigma^2\}$
   (e) $1\{X_3>0\}(X_3)$

Extra credit (5pt) What parameter do you think $\bar{X}$ is trying to estimate? Justify your answer.

Solution: (a), (c), and (e) are statistics since they are functions depending only on the data. (b) and (d) are not statistics since they also depend on the parameters $\mu$ and $\sigma^2$. Extra Credit: The sample mean is a good estimate of the actual mean of the distribution $\mu$. This is because, intuitively, we can expect the sample mean of many repeated experiments to well represent the underlying process. We will learn more about this when we learn about the central limit theorem.

2. (25 pts) Show that, for any $A, B_1, \ldots, B_n \subseteq \Omega$ with $\bigcup_{i=1}^{n} B_i = \Omega$, we have $A = (A \cap B_1) \cup \cdots \cup (A \cap B_n)$.

Solution: We have $A = A \cap \Omega$ and $\Omega = \bigcup_{i=1}^{n} B_i$. Thus,

$$A = A \cap \Omega = A \cap \left(\bigcup_{i=1}^{n} B_i\right)$$
Using the distributive property of intersection, we get

$$A = \bigcup_{i=1}^{n} (A \cap B_i).$$

3. (25 pts) Consider the following experiments:

- you have three keys and exactly one of them opens the door that you want to open; at each attempt to open the door, you pick one of the keys completely at random and try to open the door (at each attempt you pick one of the three keys at random, regardless of the fact that you may already have tried some in previous attempts). You count the number of attempts needed to open the door. What is the sample space for this experiment?

- you have an urn with 1 red ball, 1 blue ball and 1 white ball. For two times, you draw a ball from the urn, look at its color, take a note of the color on a piece of paper, and put the ball back in the urn. What is the sample space for this experiment? Write out the set corresponding to the event ‘you observe at least 1 white ball’.

**Solution:**

- The sample space is the set of positive integers $\Omega = \{1, 2, 3, 4, 5, \ldots \}$.
- Let $R$: red, $B$: blue, and $W$: white. The sample space is the set of all possible pairs of observed colors

  $$\Omega = \{RR, RB, RW, BR, BB, BW, WR, WB, WW\}.$$

  Let $A$ be the event ‘you observe at least 1 white ball’. Then $A = \{RW, BW, WR, WB, WW\}$.

4. (25 pts) Think of an example of a random process and describe what the quantity of interest $X_t$ is and what the variable $t$ represents in your example (time, space, . . . ?) **Solution:** Graded according to how specific/believable the example is.

5. **Extra credit** (10 pts) Show that the expression of the sample variance

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

(1)
is equivalent to

\[ S^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right)^2 \right]. \quad (2) \]

**Solution:**

We have

\[
(n - 1)S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2 = \sum_{i=1}^{n} (X_i^2 - 2X_i \bar{X} + \bar{X}^2)
\]

\[= \sum_{i=1}^{n} X_i^2 - 2\bar{X} \sum_{i=1}^{n} X_i + n\bar{X}^2 = \sum_{i=1}^{n} X_i^2 - 2n\bar{X}^2 + n\bar{X}^2 \]

\[= \sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \sum_{i=1}^{n} X_i^2 - n \frac{1}{n^2} \left( \sum_{i=1}^{n} X_i \right)^2. \]

Thus,

\[ S^2 = \frac{1}{n-1} \left[ \sum_{i=1}^{n} X_i^2 - \frac{1}{n} \left( \sum_{i=1}^{n} X_i \right)^2 \right]. \]
Homework 2 (due Thursday, May 19)

1. (12 pts) A secretary types 4 letters to 4 people and addresses the 4 envelopes. If he/she inserts the letters at random, what is the probability that exactly 2 letters will go in the correct envelopes?

Solution:
Call the letters A, B, C, D. What is the probability that he/she correctly address letters A and B? That is \( \frac{1}{4!} \) (There is exactly one case where A and B are matched to their envelopes and the other two aren’t! this case is the four letters addressed to (A,B,D,C) sent to persons (A,B,C,D), in that order. This is one out of the 4! possible orderings). Now, one has \( \binom{4}{2} \) possible choices of 2 letters out of 4 to be addressed correctly. Thus, the probability that exactly 2 letters will go in the correct envelopes is \( \binom{4}{2} \times \frac{1}{4!} = \frac{1}{4} \).

Alternative solution
We use counting. Define F to be the event that there are exactly two chosen correctly, while defining elements of the sample space \( \Omega = \{ABCD,ACBD,ADBC,\ldots\} \) as all addressees on the letters sent to person A,B,C,D, respectively. Then, \(|\Omega| = 4!\), and \(|F| = \binom{4}{2} \times 1\). How did we count \(|F|\)? Every event in F will have two addressees (let’s say A and B) correctly addressed (there are \( \binom{4}{2} \) such scenarios) and, out of the two possible cases in this scenario (\( \{ABCD\} \) and \( \{ABDC\} \)), only one of them is contained in F, since the other has not 2 but 4 correct addressees! The probability of interest, \( P(F) \), is now equal to \( \frac{|F|}{|\Omega|} = \frac{\binom{4}{2}}{4!} \).

2. (12 pts, 6 pts each) A labor dispute has arisen concerning the distribution of 20 laborers to four different construction jobs. The first job is considered very undesirable, and requires 6 laborers. The second, third and fourth utilized 4,5,5 laborers respectively. The accusation was made because four people in the same race out of the twenty were all placed in job 1. In considering whether the assignment represented injustice, an investigation is launched to see the probability of the observed event. (6 points each)

(a) What is the number of points in the sample space?
(b) Find the probability of the observed event, if it is assumed that the laborers are randomly assigned to all jobs

Solution: The number of ways of assigning the 20 laborers to the four jobs is equal to the number of ways of partitioning the \( n = 20 \) into
By a random assignment of laborers to the jobs, we mean that each of the \( N \) sample points has probability equal to \( 1/N \). If \( A \) denotes the event of interest and \( n_a \) the number of sample points in \( A \), the sum of the probabilities of the sample points in \( A \) is \( P(A) = n_a(1/N) = n_a/N \). The number of sample points in \( A \), \( n_a \), is the number of ways of assigning laborers to the four jobs with the 4 members of the ethnic group all going to job 1. The remaining 16 laborers need to be assigned to the remaining jobs. Because there remain two openings for job 1, this can be done in ways. It follows that

\[
n_a = \frac{16!}{2!4!5!5!}, P(A) = \frac{n_a}{N} = 0.0031
\]

Thus, if laborers are randomly assigned to jobs, the probability that the 4 members of the ethnic group all go to the undesirable job is very small. There is reason to doubt that the jobs were randomly assigned.

3. (12 pts) A fair die is tossed 6 times, and the number on the uppermost face is recorded each time. What is the probability that the numbers recorded are in the first four rolls are 1, 2, 3, 4 in any order? It is useful to invoke the counting rules.

**Solution:**

First, notice that all outcomes of the six dice rolls are equally likely.

**The denominator** The number of possible outcomes of the experiment is \( 6^6 \) (in the first roll there are 6 possible outcomes, in the second roll there are again 6 possible outcomes, and so on).

**The numerator** The number of ways in which the numbers 1, 2, 3, 4 can appear in any order is \( 4! \). This can be understood as counting the number of ways of putting 1,2,3 or 4 into four slots exactly one each. The first slot has four options, the second slot has only three options (since the first slot took one of the four candidates), and so forth. Then, the last two slots can take anything, so there are \( 6^2 \) ways of this happening.

**Numerator/denominator** gives \( 4!6^2/6^6 \).
Alternatively, use multiplication of conditional probabilities. The probability of the first dice roll resulting in one of \{1, 2, 3, 4\} is \(\frac{4}{6}\). The second roll can now only take one of the three remaining; this has \(\frac{3}{6}\), and so forth. The last two slots do not matter. So, the probability is 

\[
\frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{1}{6}
\]

Lastly, this can be understood from a series of conditional probabilities multiplied using Bayes’ rule (which we will learn later this week)! Denote \(A_i\)’th roll the event that the \(i\)’th roll fits the description of the event that we are interested in! Then:

\[
P(A_1, A_2, A_3, A_4) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2)P(A_4|A_1, A_2, A_3) = \frac{4}{6} \cdot \frac{3}{6} \cdot \frac{2}{6} \cdot \frac{1}{6}
\]

4. (12 pts, 6 pts each)

(a) What is the number of ways of selecting two random applicants out of five?

(b) What is the probability that exactly one of the two best applicants appear in a random selection of two out of five?

Solution:

(a) \(\binom{5}{2} = \frac{5!}{2!3!} = 10\)

(b) Let \(n_a\) denote the number of sample points in A. Then \(n_a\) equals the number of ways of selecting one of the two best (call this number \(m\)) times the number of ways of selecting one of the three low-ranking applicants (call this number \(n\)). Then \(m = 2\), \(n = 3\), and applying the mn rule,

\[
n_a = \binom{2}{1} \binom{3}{1} = \frac{2!}{1!1!} \cdot \frac{3!}{1!2!} = 6
\]

5. (12 pts) A population of voters contains 40% Republicans and 60% Democrats. It is reported that 30% of the Republicans and 70% of the Democrats favor an election issue. A person chosen at random from this population is found to favor the issue in question. What is the conditional probability that this person is a Democrat?

Solution:

Let \(D\) and \(R\) denote the events ‘the person is a Democrat’ and ‘the
person is a Republican’, respectively. Let \( F \) denote the event ‘the person favors the issue’. We have \( P(D) = 0.6 \), \( P(R) = 0.4 \), \( P(F|D) = 0.7 \), and \( P(F|R) = 0.3 \). Then, using Bayes’ Rule,

\[
P(D|F) = \frac{P(F|D)P(D)}{P(F)} = \frac{P(F|D)P(D)}{P(F|D)P(D) + P(F|R)P(R)}
\]

\[
= \frac{0.7 \times 0.6}{0.7 \times 0.6 + 0.3 \times 0.4} = \frac{7}{9}.
\]

6. (4 pts) Use the axioms of probability to show that for any \( A \subset S \) we have \( P(A^c) = 1 - P(A) \).

**Solution:**
We have \( \Omega = A \cup A^c \). Since \( A \) and \( A^c \) are disjoint, \( P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) \). Thus, \( P(A^c) = P(\Omega) - P(A) \). Since \( P(\Omega) = 1 \), we then have \( P(A^c) = 1 - P(A) \).

7. (8 pts) Let \( B \subset A \). Show that \( P(A) = P(B) + P(A \cap B^c) \).

**Solution:**
Remember from Homework 1 that we have \( A = (A \cap B) \cup (A \cap B^c) \). Furthermore, these two sets are disjoint: \( A \cap B \) is a subset of \( B \), \( A \cap B^c \) is a subset of \( B^c \), and obviously \( B \) and \( B^c \) are disjoint. Therefore, \( P(A) = P(A \cap B) + P(A \cap B^c) \). Since \( B \subset A \), we have \( A \cap B = B \) and thus \( P(A) = P(B) + P(A \cap B^c) \).

8. (10 pts) Show that for \( A, B \subset \Omega \), \( P(A \cap B) \geq \max\{0, P(A) + P(B) - 1\} \).

**Solution:**
Because the probability of any event is a non-negative number, it is clear that \( P(A \cap B) \geq 0 \). Furthermore, recall that in general for two events \( A \) and \( B \) we have

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B)
\]

from which it follows that

\[
P(A \cap B) = P(A) + P(B) - P(A \cup B)
\]

Since \( P(A \cup B) \leq 1 \), we have

\[
P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(A) + P(B) - 1.
\]

Thus, since \( P(A \cap B) \geq 0 \) and \( P(A \cap B) \geq P(A) + P(B) - 1 \), we conclude that \( P(A \cap B) \geq \max\{0, P(A) + P(B) - 1\} \).
9. (10 pts) Let $A, B \subset \Omega$ be such that $P(A) > 0$ and $P(B) > 0$. Show that

- if $A$ and $B$ are disjoint, then they are not independent
- if $A$ and $B$ are independent, then they are not disjoint.

**Solution:**

- Suppose that $A$ and $B$ are independent. Then, $P(A \cap B) = P(A)P(B) > 0$ (since $P(A) > 0$ and $P(B) > 0$ by assumption). However, $A$ and $B$ are disjoint by assumption, i.e. $A \cap B = \emptyset$ which implies that $P(A \cap B) = 0$, and we reach a contradiction.

- Suppose that $A$ and $B$ are disjoint, i.e. $A \cap B = \emptyset$. Then $P(A \cap B) = 0$. However, $A$ and $B$ are independent by assumption, i.e. $P(A \cap B) = P(A)P(B) > 0$ (recall that $P(A) > 0$ and $P(B) > 0$), and again we reach a contradiction.

10. (10 pts) Let $A, B \subset \Omega$ be such that $P(A) > 0$ and $P(B) > 0$. Prove or disprove the following claim: in general, it is true that $P(A \mid B) = P(B \mid A)$. If you believe that the claim is true, clearly explain your reasoning. If you believe that the claim is false in general, can you find an additional condition under which the claim is true?

**Solution:**

The claim is false in general. A simple counter example: take $A$ arbitrary with $0 < P(A) < 1$ and $B = \Omega$. We have $P(A \mid B) = P(A \mid \Omega) = P(A)$ while $P(B \mid A) = P(\Omega \mid A) = 1$. Notice, however, that if $P(A) = P(B)$, then

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A)} = P(B \mid A).$$

Therefore, under this additional condition, the claim is true.

11. (10 pts) Let $A, B \subset \Omega$ be independent events. Show that the following pairs of events are also independent:

- $A, B^c$
- $A^c, B$
- $A^c, B^c$.

**Solution:**
• We have $A = (A \cap B) \cup (A \cap B^c)$ and the two events are disjoint. Therefore,

\[
P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B)
\]

\[
= P(A)[1 - P(B)] = P(A)P(B^c)
\]

which implies that $A$ and $B^c$ are independent.

• To show $A^c$ and $B$ are independent, just switch the roles of $A$ and $B$ in the argument above.

• Once again we can write $A^c = (A^c \cap B^c) \cap (A^c \cap B)$ and the two events are disjoint. Hence, as in the first part,

\[
P(A^c \cap B^c) = P(A^c) - P(A^c \cap B) = P(A^c) - P(A^c)P(B)
\]

\[
= P(A^c)[1 - P(B)] = P(A^c)P(B^c)
\]

implying that $A^c$ and $B^c$ are also independent (notice that here we used the previous result on the independence of $A^c$ and $B$).
Homework 3 (due Tuesday, May 24th)

1. (4 pts) Consider three urns. Urn A contains 2 white and 3 red balls, urn B contains 8 white and 4 red balls and urn C contains 1 white and 3 red balls. If one ball is selected from each urn, what is the probability that the ball chosen from urn A was white, given that exactly 2 white balls were selected?

Solution:
Let $A$ be a random variable representing the color of the ball from urn A. $A$ takes values $\{W, R\}$ for white and red balls, respectively. Let $B$ and $C$ be defined similarly for urns B and C. Let $N_W$ be a random variable that represents the number of white balls chosen. In this case, $N_W$ can be 0, 1, 2, or 3. Using Bayes Rule,


$$P(A = W) = \frac{2}{2+3} = \frac{2}{5}, \quad P(A = R) = \frac{3}{2+3} = \frac{3}{5}$$

since there are 2 white balls and 3 red balls in urn A. There are two ways to get 2 white balls given that $A = W$. Either $B = W$ and $C = R$, or $B = R$ and $C = W$. The probability of this happening is

$$P(N_W = 2 | A = W) = P(B = W \cap C = R) + P(B = R \cap C = W) = P(B = W) * P(C = R) + P(B = R) * P(C = W) = \frac{8}{12} * \frac{3}{4} + \frac{4}{12} * \frac{1}{4} = \frac{7}{12}. $$

Given that $A = R$, the only way to get 2 white balls is if both $B = W$ and $C = W$. The probability of this happening is $P(N_W = 2 | A = R) = P(B = W \cap C = W) = P(B = W) * P(C = W) = \frac{8}{12} * \frac{1}{4} = \frac{1}{6}$. Thus,


$$= \frac{\frac{7}{12} * \frac{2}{5}}{\frac{7}{12} * \frac{2}{5} + \frac{1}{6} * \frac{3}{5}}$$

$$= \frac{7}{10}$$
2. (4 pts) Consider the random variable $X$ with p.m.f.

$$p(x) = \begin{cases} 
0.1 & \text{if } x = 0 \\
0.9 & \text{if } x = 1 \\
0 & \text{if } x \notin \{0, 1\}.
\end{cases}$$

Draw the p.m.f. of $X$. Compute the c.d.f. of $X$ and draw it.

**Solution:**
The c.d.f. of $X$ is

$$F(x) = \begin{cases} 
0 & \text{if } x < 0 \\
p(0) & \text{if } 0 \leq x < 1 \\
p(0) + p(1) & \text{if } x \geq 1.
\end{cases}$$

3. (10 pts) Let $X$ be a continuous random variable with p.d.f.

$$f(x) = 3x^2\mathbb{1}_{[0,1]}(x).$$

Compute the c.d.f. of $X$ and draw it. What is $P(X \in (0.2, 0.4))$? What is $P(X \in [-3, 0.2])$?

**Solution:**

11
The c.d.f. of $X$ is

$$F(x) = \int_{-\infty}^{x} f(x) \, dx = \begin{cases} 0 & \text{if } x < 0 \\ \int_{0}^{x} 3y^2 \, dy & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} = \begin{cases} 0 & \text{if } x < 0 \\ y^3 |_{0}^{x} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

$$= \begin{cases} 0 & \text{if } x < 0 \\ x^3 & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

We have

$$P(X \in (0.2, 0.4)) = F(0.4) - F(0.2) = 0.4^3 - 0.2^3 = 0.056$$

and

$$P(X \in [-3, 0.2)) = F(0.2) - F(-3) = 0.2^3 - 0 = 0.008.$$
Solution:
The c.d.f. of $X$ is
\[
F(x) = \begin{cases} 
0 & \text{if } x < 2 \\
p(2) & \text{if } 2 \leq x < 3 \\
p(2) + p(3) & \text{if } 3 \leq x < 4 \\
p(2) + p(3) + p(4) & \text{if } x \geq 4 
\end{cases}
= \begin{cases} 
0 & \text{if } x < 2 \\
1/6 & \text{if } 2 \leq x < 3 \\
1/6 + 1/3 & \text{if } 3 \leq x < 4 \\
1/6 + 1/3 + 1/2 & \text{if } x \geq 4 
\end{cases}
\]

Thus, we have
\[
P(2 \leq X \leq 5) = F(5) - F(2) + p(2) = 1 - 1/6 + 1/6 = 1
\]
\[
P(X > 3) = 1 - P(X \leq 3) = 1 - F(3) = 1 - 1/2 = 1/2
\]
\[
P(0 < X < 4) = F(3) - F(0) = 1/2 - 0 = 1/2.
\]

5. (10 pts) Consider the following c.d.f. for the random variable $X$:
\[
F(x) = P(X \leq x) = \begin{cases} 
0 & \text{if } x < 0 \\
\frac{1}{8} \sqrt{x} & \text{if } 0 \leq x < 1 \\
-\frac{3}{8} + \frac{1}{2} x & \text{if } 1 \leq x < \frac{3}{2} \\
-\frac{3}{8} + \frac{5}{4} x & \text{if } \frac{3}{2} \leq x < 2 \\
1 & \text{if } x \geq 2.
\end{cases}
\]

What is the p.d.f. of $X$?

Solution:
Notice that $F$ is piecewise differentiable. Therefore, the p.d.f. of $X$ is
\[
f(x) = \left. \frac{d}{dy} F(y) \right|_{y=x} = \begin{cases} 
0 & \text{if } x \leq 0 \\
\frac{1}{16\sqrt{x}} & \text{if } 0 < x < 1 \\
\frac{1}{2} & \text{if } 1 \leq x < \frac{3}{2} \\
\frac{5}{4} & \text{if } \frac{3}{2} \leq x < 2 \\
0 & \text{if } x \geq 2.
\end{cases}
\]

6. (10 pts) Let $\alpha \in (0, 1)$. Derive the $\alpha$-quantile associated to the following c.d.f.
\[
F(x) = \frac{1}{1 + e^{-x}}.
\]
Solution:
We need to solve
\[ F(x_{\alpha}) = \alpha \]
with respect to \( x_{\alpha} \). We have
\[
\frac{1}{1 + e^{-x_{\alpha}}} = \alpha
\]
\[
\frac{1}{\alpha} = 1 + e^{-x_{\alpha}}
\]
\[
- \log \left( \frac{1}{\alpha} - 1 \right) = x_{\alpha}
\]
\[
x_{\alpha} = - \log \frac{1 - \alpha}{\alpha} = \log \frac{\alpha}{1 - \alpha}.
\]

7. (10 pts) Consider the continuous random variable \( X \) with p.d.f. \( f \) and its transformation \( 1_A(X) \) where \( A \subset \mathbb{R} \). Show that \( E(1_A(X)) = P(X \in A) \).

Solution:
Let \( g(x) = 1_A(x) \). By definition, the expected value of the random variable \( g(X) \) is
\[
E(g(X)) = \int_{\mathbb{R}} g(x)f(x) \, dx.
\]
In this case, we have
\[
E(g(X)) = E(1_A(X)) = \int_{\mathbb{R}} 1_A(x)f(x) \, dx = \int_A f(x) \, dx = P(X \in A).
\]

8. (6 pts) Consider a random variable \( X \) and a scalar \( a \in \mathbb{R} \). Using the definition of variance, show that \( V(X + a) = V(X) \).

Solution:
Consider the random variable \( Y = X + a \). We have that \( E(Y) = E(X + a) = E(X) + a \). Now, by definition, \( V(Y) = E[(Y - E(Y))^2] = E[(X + a - (E(X) + a))^2] = E[(X - E(X))^2] = V(X) \).

9. (6 pts) A gambling book recommends the following winning strategy for the game of roulette. It recommends that a gambler bets $1 on red. If red appears (which has probability 18/38), then the gambler should take her $1 profit and quit. If the gambler looses this bet (which has probability 20/38 of occurring), she should make additional $1 bets on red on each of the next two spins of the roulette wheel and then quit. Let \( X \) denote the gamblers winning when she quits.
(a) Find $P(X \geq 0)$

(b) Are you convinced that this is a winning strategy? Explain your answer.

(c) Find $E[X]$.

Solution:

(a) $P(X \geq 0)$ corresponds to (i) the gambler wins at first and quits.
(ii) the gambler keeps gambling after first game, and wins for the last two games. $P(X \geq 0) = \frac{18}{38} + \frac{20}{38} \cdot (\frac{18}{38})^2 \approx 0.592$

(b) No. Winning means winning 1, but losing means losing 1 or 3. The argument is complete following $E(X) < 0$

(c) $P(X = 1) = \frac{18}{38} + \frac{20}{38} \cdot (\frac{18}{38})^2 = 0.592$, $P(X = -1) = \frac{20}{38} \cdot \frac{18}{38} \cdot 2$, $P(X = -3) = (\frac{20}{38})^3$. Thus $E(X) = -0.108$.

10. (10 pts) Let $X$ be a random variable with p.d.f. $f(x) = 1_{[0,1]}(x)$.

- Draw the p.d.f. of $X$.
- Compute the c.d.f. of $X$ and draw it.
- Compute $E(3X)$.
- Compute $V(2X + 100)$.

Solution:

The c.d.f. of $X$ is

$$F(x) = \int_{-\infty}^{x} f(y) \, dy = \begin{cases} 0 & \text{if } x < 0 \\ \int_{0}^{x} dy & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases} = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$
We have that \( E(X) = \int_{\mathbb{R}} x f(x) \, dx = \int_{0}^{1} x \times 1 \, dx = \frac{1}{2} \times 1 = \frac{1}{2} \). Therefore, \( E(3X) = 3E(X) = \frac{3}{2} \). Furthermore,

\[
E(X^2) = \int_{\mathbb{R}} x^2 f(x) \, dx = \int_{0}^{1} x^2 \times 1 \, dx = \frac{1}{3} \times 1 = \frac{1}{3}
\]

hence \( V(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \). Thus, \( V(2X + 100) = V(2X) = 4V(X) = \frac{4}{12} = \frac{1}{3} \).

11. (10 pts) A secretary is given \( n > 0 \) passwords to access a computer file. In order to open the file, she starts entering the password in a random order and after each attempt she discards the password that she used if that password was not the correct one. Let \( X \) be the random variable corresponding to the number of password that the secretary needs to enter before she can open the file.

- Argue that \( P(X = 1) = P(X = 2) = \cdots = P(X = n) = 1/n \).
- Find the expectation and the variance of \( X \).

Potentially useful facts:

- \( \sum_{x=1}^{n} x = n(n+1)/2 \)
- \( \sum_{x=1}^{n} x^2 = n(n+1)(2n+1)/6 \).

Solution:
Let \( R_i \) denote the event ‘the secretary chooses the right password
on the $i$-th attempt. We have $P(X = 1) = P(R_1) = 1/n$ since the secretary is picking one password at random from the $n$ available passwords. The probability that she chooses the right password on the second attempt is

$$P(X = 2) = P(R_1^c \cap R_2) = P(R_2 | R_1^c) = \frac{n - 1}{n - 1} = \frac{1}{n},$$

i.e. it equals the probability that she chooses the wrong password on the first attempt and then she chooses the right one on the second attempt. Similarly,

$$P(X = 3) = P(R_1^c \cap R_2^c \cap R_3^c)
= P(R_1^c) P(R_2^c | R_1^c) P(R_3^c | R_1^c \cap R_2^c)
= \frac{n - 1}{n} \frac{1}{n - 1} \frac{1}{n - 2} = \frac{1}{n},$$

and then it is clear that $P(X = i) = \frac{1}{n}$ for all $i \in \{1, \ldots, n\}$. The expectation of $X$ is therefore

$$E(X) = \sum_{x=1}^{n} x P(X = x) = \sum_{x=1}^{n} x \frac{1}{n}
= \frac{1}{n} \sum_{x=1}^{n} x = \frac{1}{n} \frac{n(n+1)}{2}
= \frac{n+1}{2}.$$  

To compute the variance of $X$ we also need

$$E(X^2) = \sum_{x=1}^{n} x^2 P(X = x) = \frac{1}{n} \sum_{x=1}^{n} x^2
= \frac{1}{n} \frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}.$$  

Thus,

$$V(X) = E(X^2) - [E(X)]^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}
= \frac{2n^2 + n + 2n + 1}{6} - \frac{n^2 + 2n + 1}{4}
= \left(\frac{1}{3} - \frac{1}{4}\right) n^2 + \left(\frac{1}{2} - \frac{1}{2}\right) \frac{n + 1}{6} - \frac{1}{4}
= \frac{1}{12} n^2 - \frac{1}{12} = \frac{n^2 - 1}{12}.  \n$$
12. (10 pts) Consider the positive random variable \( X \) (i.e. \( P(X > 0) = 1 \)) with expectation \( E(X) = 1 \) and variance \( V(X) = 0.002 \), and is symmetric. Use Markov’s inequality for a lower bound, and Tchebysheff’s inequality to provide an upper bounds for the probability of the event \( \{X \leq 0.1\} \). (Hint: think about what allows you to quantify the probability that \( X \leq 0.1 \) or \( X \geq 1.9 \).) Are these bounds useful in this case? Comment on the bounds.

**Solution:**

(a) Lower Bound: Markov’s inequality yields

\[
P(|X| \geq 0.1) \leq \frac{E|X|}{0.1}.
\]

Hence,

\[
P(|X| < 0.1) > 1 - \frac{E|X|}{0.1}.
\]

Since \( X \) is positive, this is equivalent to

\[
P(X < 0.1) > 1 - \frac{E(X)}{0.1} = 1 - 10 = -9.
\]

The bound is valid and correct! The probability of any event is a non-negative number, so it is true that \( P(X < 0.1) > -9 \). However, this bound is uninformative. It gives us no additional information about the distribution of \( X \).

There is no straightforward way to apply Tchebysheff’s inequality for a lower bound, since it quantifies the union of the tail probabilities away from the mean of the random variable (and not zero).

(b) Upper bound: We use the Chebyshev’s inequality with \( k = 0.9 \), to have

\[
P(|X - E(X)| \geq 0.9) = P(|X - 1| \geq 0.9) = P(X \leq 0.1 \text{or} X \geq 1.9) < 0.002/0.9 \approx 1/500
\]

This gives the probability associated with both ‘tails’ of the distribution, \( X \leq 0.1 \) or \( X \geq 1.9 \). Then, we notice that (using the ‘union bound’, from lecture)

\[
P(\{X \leq 0.1\}) \leq P(\{X \leq 0.1\} \cup \{X \geq 1.9\}) \leq 1.111
\]
(or even use that the two events are disjoint, so that $P(\{X \leq 0.1\} \cup \{X \geq 0.9\}) = P(\{X \leq 0.1\}) + P(\{X \geq 1.9\}) \geq P(\{X \geq 0.1\})$). This is

$$P(\{X \leq 0.1\}) \leq P(\{X \leq 0.1\} \cup \{X \geq 1.9\}) \leq 0.002/0.9$$

which seems like an informative bound! The lesson: when you know the variance (i.e. if you have more information), you can use Tchebyshev to better quantify a ‘tail’ probability.
Homework 4 (due Thursday, May 28th)

1. (6 pts) An oil exploration firm is formed with enough capital to finance ten explorations. The probability of a particular exploration being successful is 0.1. Assume the explorations are independent. Find the mean and variance of the number of successful explorations.

**Solution:**

Let $n \in \mathbb{Z}_+$ and $p \in [0,1]$. For any $i \in \{1, \ldots, n\}$, let $X_i$ be the Bernoulli($p$) random variable describing the success or failure of the $i$-th exploration of the oil exploration firm:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th exploration is successful} \\ 0 & \text{otherwise} \end{cases}$$

The number of successful explorations is

$$Y = \sum_{i=1}^{n} X_i.$$

Since the explorations are assumed to be independent, it follows that $Y \sim \text{Binomial}(n, p)$ (recall that the sum of $n$ independent Bernoulli($p$) random variables is a random variable itself and its distribution is $\text{Binomial}(n, p)$). In this exercise $n = 10$ and $p = 0.1$ so we have $E(Y) = np = 10 \times 0.1 = 1$ and $V(Y) = np(1 - p) = 1 \times 0.9 = 0.9$.

2. (8 pts) Ten motors are packaged for sale in a certain warehouse. The motors sell for USD 100 each, but a double-your-money-back guarantee is in effect for any defectives the purchaser may receive (i.e. if the purchaser receives one defective motor, the seller has to pay USD 200 back to the purchaser). Find the expected net gain for the seller if the probability of any one motor being defective is 0.08 (assuming that the quality of any one motor is independent of that of the others).

**Solution:**

Let $n$ be the number of motors that are packaged for sale in the warehouse and $c$ be the price of each motor. For any $i \in \{1, \ldots, n\}$ let $D_i$ denote the Bernoulli($p$) random variable describing whether the $i$-th motor is defective or not:

$$D_i = \begin{cases} 1 & \text{if the } i\text{-th motor is defective} \\ 0 & \text{otherwise} \end{cases}$$
Everytime a motor is sold the seller gains $c$ dollars, but if the motor turns out to be defective the seller has to pay $2c$ dollars back to the buyer because of the double-your-money-back guarantee. Assuming that the quality of each motor is independent of the quality of every other motor in stock, the net gain of the seller is the random variable

$$G = cn - 2c \sum_{i=1}^{n} D_i$$

where $\sum_{i=1}^{n} D_i \sim \text{Binomial}(n, p)$ and the expected net gain is the expectation of $G$:

$$E(G) = E \left( cn - 2c \sum_{i=1}^{n} D_i \right) = cn - 2cE \left( \sum_{i=1}^{n} D_i \right) = cn - 2cnp = cn(1-2p).$$

Since $c = \$100$, $n = 10$ and $p = 0.08$ we have that

$$E(G) = \$100 \times 10 \times (1 - 2 \times 0.08) = \$1000 \times 0.84 = \$840.$$ 

3. (6 pts) Consider again Exercise 11 of Homework 3. Suppose this time that the secretary does not discard a password once she tried it and discovered that the password is wrong. What is the distribution of the random variable $X$ corresponding to the number of attempts needed until she enters the correct password? What are the mean and the variance of $X$?

**Solution:**

We can model the sequence of attempts using a sequence of random variables $X_1, X_2, \ldots \overset{i.i.d.}{\sim} \text{Bernoulli}(p)$ where $p = 1/n$. The number of attempts needed to enter the correct password for the first time then corresponds to a random variable $Y \sim \text{Geometric}(p)$. Its expectation is $E(Y) = 1/p = n$ and its variance is $V(Y) = (1 - p)/p^2 = n(n - 1)$.

4. (8 pts) A geological study indicates that an exploratory oil well should strike oil with probability 0.2.

- What is the probability that the first strike comes on the third well drilled?
- What is the probability that the third strike comes on the seventh well drilled?
- What assumptions did you make to obtain the answers to parts (a) and (b)?
• Find the mean and variance of the number of wells that must be drilled if the company wants to set up three producing wells.

**Solution:**
Let $X$ be the random variable describing the number of the trial on which the $r$-th oil well is observed to strike oil. $X$ is distributed according to $\text{NBinomial}(r, p)$ and its probability mass function is

$$P(X = x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{if } x = r, r+1, \ldots \\ 0 & \text{otherwise.} \end{cases}$$

• The probability that the first strike comes on the third well drilled is

$$P(X = 3) = \binom{3-1}{1-1} p^3 (1-p)^{3-1} = p(1-p)^2$$

which is equal to the probability of observing the first success on the third trial (notice that this is really the same as doing the computation with a Geometric($p$) distribution since here $r = 1$). $p = 0.2$, therefore we have

$$P(X = 3) = 0.2(0.8)^2 = 0.128.$$  

• Here $r = 3$, so the probability that the third strike comes on the seventh well drilled is

$$P(X = 7) = \binom{7-1}{3-1} p^3 (1-p)^{7-3} = \binom{6}{2} p^3 (1-p)^4$$

$$= \binom{6}{2} 0.2^3 \times 0.8^4 \approx 0.049.$$  

• In order to apply the Negative Binomial model we need to make some assumptions. In particular, we have to assume that all the trials are identical and independent Bernoulli trials (i.e. every time a well is drilled the fact that it strikes oil or not does not depend on what happens at other wells and the probability that a well strikes oil is the same for all wells).

• Here we want to compute $E(Y)$ and $V(Y)$ when $r = 3$, namely the expectation and the variance of the number of the trial on which the third well is observed to strike oil. We have

$$E(Y) = \frac{r}{p} = \frac{3}{0.2} = 15$$

$$V(Y) = \frac{r(1-p)}{p^2} = \frac{3 \times 0.8}{0.2^2} = 60.$$  

22
5. (6 pts) Cards are dealt at random and without replacement from a standard 52 card deck. What is the probability that the second king is dealt on the fifth card?

**Solution:**
The probability of the event that the second king is dealt on the fifth card is equivalent to the probability of the intersection of the two events $A$=“exactly one king is dealt in the first four cards” and $B$ = “exactly a king is dealt on the fifth card”. We have

$$P(A) = \frac{\binom{4}{1} \binom{48}{3}}{\binom{52}{4}}$$

and

$$P(B|A) = \frac{\binom{3}{1} \binom{45}{0}}{\binom{48}{1}}$$

so

$$P(A \cap B) = P(A)P(B|A) = \frac{\binom{4}{1} \binom{48}{3} \binom{3}{1} \binom{45}{0}}{\binom{52}{4} \binom{48}{1}} = \frac{4 \times 17296 \times 3 \times 1}{270725 \times 48} \approx 0.016.$$

We can also interpret the problem in the following way. Let $K_1 \sim \text{HGeometric}(r, N, n)$ be the random variable describing the number of kings that are dealt in the first $n$ cards from a deck of $N$ cards containing exactly $r$ kings and let $K_2 \sim \text{Hypergeometric}(s, M, m)$ be a random variable describing the number of kings that are dealt in the first $m$ cards from a deck of $M$ cards containing exactly $s$ kings. Assume moreover that $K_1$ and $K_2$ are independent. Then the probability that the second king is dealt on the fifth hand can be expressed as

$$P(K_1 = 1, K_2 = 1) = P(K_1 = 1)P(K_2 = 1) = \frac{\binom{N}{r-1} \binom{1}{1} \binom{M-s}{m-1}}{\binom{N}{n} \binom{M}{m}}$$

for a particular choice of the parameters. For this question, we know that $N = 52$, $n = 4$ and $r = 4$ and we know how the deck changes after exactly one king and three other cards are dealt: indeed, we know that $M = N - n$, $m = 5 - n$ and $s = r - 1$. Moreover, the independence assumption is appropriate since the cards are always dealt randomly, no matter the composition of the deck. It’s easy to check that plugging this parameters values into the expression for $P(K_1 = 1, K_2 = 1)$ yields
6. (8 pts) A parking lot has two entrances. Cars arrive at entrance I according to a Poisson distribution at an average of three per hour and at entrance II according to a Poisson distribution at an average of four per hour. What is the probability that a total of three cars will arrive at the parking lot in a given hour? Assume that the numbers of cars arriving at the two entrances are independent.

**Solution:**
Let \( C_1 \sim \text{Poisson}(\lambda) \) and \( C_2 \sim \text{Poisson}(\mu) \) denote the number of cars arriving in an hour at entrance I and entrance II of the parking lot respectively. We know that \( \lambda = 3 \) and \( \mu = 4 \) and we assume that \( C_1 \) and \( C_2 \) are independent. At a given hour

\[
C = C_1 + C_2
\]

is the total number of cars arriving at the parking lot. We are interested in computing \( P(C = 3) \). The event \( C = 3 \) can be written as the following disjoint union of events

\[
\{C = 3\} = (\{C_1 = 0\} \cap \{C_2 = 3\}) \cup (\{C_1 = 1\} \cap \{C_2 = 2\}) \cup (\{C_1 = 2\} \cap \{C_2 = 1\}) \cup (\{C_1 = 3\} \cap \{C_2 = 0\})
\]

whose probability is (here we use the fact that the union is a disjoint union and we exploit the independence of \( C_1 \) and \( C_2 \))

\[
P(C = 3) = P(C_1 = 0 \cap C_2 = 3) + P(C_1 = 1 \cap C_2 = 2) + P(C_1 = 2 \cap C_2 = 1) + P(C_1 = 3 \cap C_2 = 0) = e^{-\lambda} \frac{\lambda^3}{3!} + e^{-\lambda} \frac{\mu^2}{2} + e^{-\lambda} \frac{\lambda^2}{2} e^{-\mu} + e^{-\lambda} \frac{\lambda^3}{3!} e^{-\mu} = e^{-(\lambda+\mu)} \left( \frac{\mu^3}{3!} + \frac{\lambda^3}{3!} + \frac{\lambda^2}{2} + \frac{\mu^2}{2} \right) \approx 0.052.
\]
7. (6 pts) Derive the c.d.f. of $X \sim \text{Uniform}(a,b)$.

**Solution:**
Recall that

$$f(x) = \frac{1}{b-a} 1_{[a,b]}(x).$$

We have

$$F(x) = P(X \leq x) = \begin{cases} 
0 & \text{if } x < a \\
\int_a^x \frac{1}{b-a} \, \text{d}x & \text{if } a \leq x < b \\
1 & \text{if } x \geq b.
\end{cases} = \begin{cases} 
0 & \text{if } x < a \\
\frac{x-a}{b-a} & \text{if } a \leq x < b \\
1 & \text{if } x \geq b.
\end{cases}$$

8. (4 pts) Let $X \sim \text{Uniform}(0,1)$. Is $X \sim \text{Beta}(\alpha,\beta)$ for some $\alpha,\beta > 0$?

**Solution:**
Recall the p.d.f. corresponding to the Beta distribution with parameters $\alpha, \beta > 0$:

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} 1_{[0,1]}(x).$$

By comparing to the p.d.f. of $X$

$$f_X(x) = 1_{[0,1]}(x),$$

it is easily seen that $X \sim \text{Uniform}(\alpha, \beta) \equiv \text{Beta}(1,1)$. Notice that in this case $\Gamma(\alpha) = \Gamma(\beta) = \Gamma(1) = 0! = 1$ and $\Gamma(\alpha + \beta) = \Gamma(2) = 1! = 1$.

9. (8 pts) The number of defective circuit boards coming off a soldering machine follows a Poisson distribution. During a specific eight-hour day, one defective circuit board was found. Conditional on the event that a single defective circuit was produced, the time at which it was produced has a Uniform distribution.

- Find the probability that the defective circuit was produced during the first hour of operation during that day.
- Find the probability that it was produced during the last hour of operation during that day.
- Given that no defective circuit boards were produced during the first four hours of operation, find the probability that the defective board was manufactured during the fifth hour.
Solution:
Let $T$ be the random variable denoting the time at which the defective circuit board was produced. $T$ is uniformly distributed in the interval $[0, 8]$ (hours), that is $T \sim \text{Uniform}(0, 8)$. Its probability density function is therefore
\[
f(t) = \begin{cases} 
\frac{1}{8} & \text{if } t \in [0, 8] \\
0 & \text{otherwise},
\end{cases}
\]
and its cumulative distribution function is
\[
F(t) = P(T \leq t) = \begin{cases} 
0 & \text{if } t < 0 \\
\frac{t}{8} & \text{if } t \in [0, 8) \\
1 & \text{if } t \geq 8.
\end{cases}
\]

- The probability that the defective circuit board was produced during the first hour of operation during the day is
\[
P(T \in (0, 1]) = P(0 < T \leq 1) = F(1) - F(0) = \frac{1}{8} - 0 = \frac{1}{8}.
\]

- The probability that the defective circuit board was produced during the last hour of operation during the day is
\[
P(T \in (7, 8]) = P(7 < T \leq 8) = F(8) - F(7) = 1 - \frac{7}{8} = \frac{1}{8}.
\]

- Given that no defective circuits boards were produced in the first four hours (that is we know that the event $T > 4$ occurred), the probability that the defective board was manufactured during the fifth hour is
\[
P(T \in (5, 6) | T > 4) = \frac{P(T \in (5, 6) \cap T > 4)}{P(T > 4)} = \frac{P(T \in (5, 6))}{P(T > 4)} = \frac{F(6) - F(5)}{1 - F(4)} = \frac{6/8 - 5/8}{1 - 1/2} = \frac{1/8}{1/2} = \frac{1}{4}.
\]

10. (6 pts) Let $Z \sim \mathcal{N}(0, 1)$. Show that, for any $z > 0$, $P(|Z| > z) = 2(1 - \Phi(z))$.

Solution:
We have
\[
P(|Z| > z) = P(Z > z \cup Z < -z) = P(Z > z) + P(Z < -z)
= 1 - P(Z \leq z) + P(Z < -z) = 1 - \Phi(z) + \Phi(-z)
= 1 - \Phi(z) + 1 - \Phi(z) = 2(1 - \Phi(z)).
\]
11. (6 pts) Let $Z \sim N(0, 1)$. Express in terms of $\Phi$, the c.d.f. of $Z$, the following probabilities

- $P(Z^2 < 1)$
- $P(3Z/5 + 3 > 2)$.

Compute the exact probability of the event $\Phi^{-1}(\alpha/3) \leq Z \leq \Phi^{-1}(4\alpha/3)$, where $\Phi^{-1}(\alpha)$ is the inverse c.d.f. computed at $\alpha$ (i.e. the $\alpha$-quantile of $Z$).

**Solution:**
We have

- $P(Z^2 > 1) = P(|Z| > 1)$. Based on the result from the previous exercise, $P(|Z| > 1) = 2(1 - \Phi(1))$. Thus, $P(Z < 1) = 1 - P(|Z| > 1) = 1 - 2(1 - \Phi(1)) = 2\Phi(1) - 1$.

- $P(3Z/5 + 3 > 2) = P(3Z/5 > -1) = P(Z > -5/3) = 1 - \Phi(-5/3) = \Phi(5/3)$.

Regarding the last question, we have

\[
P(\Phi^{-1}(\alpha/3) \leq Z \leq \Phi^{-1}(4\alpha/3)) = \Phi(\Phi^{-1}(4\alpha/3)) - \Phi(\Phi^{-1}(\alpha/3)) = 4\alpha/3 - \alpha/3 = \alpha.
\]

12. (8 pts) The SAT and ACT college entrance exams are taken by thousands of students each year. The mathematics portions of each of these exams produce scores that are approximately normally distributed. In recent years, SAT mathematics exam scores have averaged 480 with standard deviation 100. The average and standard deviation for ACT mathematics scores are 18 and 6, respectively.

- An engineering school sets 550 as the minimum SAT math score for new students. What percentage of students will score below 550 in a typical year? Express your answer in terms of $\Phi$.

- What score should the engineering school set as a comparable standard on the ACT math test? You can leave your answer in terms of $\Phi^{-1}$.

**Solution:**
Let $S$ denote the random variable describing the SAT mathematics scores and let $A$ denote the random variable describing the ACT scores.
mathematics scores. Let also \( Z \sim \mathcal{N}(0, 1) \). \( S \) is distributed approximately according to \( \mathcal{N}(\mu_s, \sigma_s^2) \) with \( \mu_s = 480 \) and \( \sigma_s = 100 \) and \( A \) is distributed approximately according to \( \mathcal{N}(\mu_a, \sigma_a^2) \) with \( \mu_a = 18 \) and \( \sigma_a = 6 \).

- Based on the approximations above the percentage of students in the engineering school with SAT math score below 550 is

\[
P(S \leq 550) = P\left( \frac{S - \mu_s}{\sigma_s} \leq \frac{550 - \mu_s}{\sigma_s} \right) \approx P\left( Z \leq \frac{550 - 480}{100} \right) = P(Z \leq 0.7) = \Phi(0.7) \approx 0.758.
\]

- The minimum score that the engineering school should require from new students in the ACT math test in order for the two minimum scores to be two comparable standards is the number \( a \) that satisfies

\[
P(A \leq a) \approx P(S \leq 550) \approx \Phi(0.7).
\]

We have to solve for \( a \) the following equation

\[
P(A \leq a) = P\left( \frac{A - \mu_a}{\sigma_a} \leq \frac{a - \mu_a}{\sigma_a} \right) \approx P\left( Z \leq \frac{a - \mu_a}{\sigma_a} \right) = \Phi(0.7),
\]

namely we need to solve

\[
\frac{a - \mu_a}{\sigma_a} = 0.7
\]

whose solution is \( a = 0.7\sigma_a + \mu_a = 0.7 \times 6 + 18 = 22.2 \).

13. (6 pts) Suppose that a random variable \( X \) has a probability density function given by

\[
f(x) = cx^3e^{-\frac{x}{2}}1_{[0, \infty)}(x).
\]

- Find the value of \( c \) that makes \( f \) a valid p.d.f. (hint: if you think about this carefully, no calculations are needed!)

- Does \( X \) have a Chi-Square distribution? If so, with how many degrees of freedom?

- What are \( E(X) \) and \( V(X) \)?

Solution:
• Given that $f$ is a bona fide p.d.f., $c$ has to be exactly the positive constant that makes the kernel of the p.d.f.

$$x^3e^{-\frac{x}{2}}\mathbb{1}_{[0,\infty)}(x)$$

integrate to 1. We recognize that this kernel is the kernel of a Gamma p.d.f. with parameters $\alpha = 4$ and $\beta = 2$. It follows that $c$ must be the number

$$c = \frac{1}{\beta^\alpha \Gamma(\alpha)} = \frac{1}{2^4 \Gamma(4)} = \frac{1}{16 \times 3!} = \frac{1}{16 \times 6} = \frac{1}{96}.$$

• Recall that $\text{Gamma}(\nu/2, 2) \equiv \chi^2(\nu)$. Therefore, in this case, $\nu = 8$ and the above p.d.f. corresponds to a Chi-Square p.d.f. with $\nu = 8$ degrees of freedom.

• We have $E(X) = \nu = 8$ and $V(X) = 2\nu = 16$.

14. (8 pts) Consider the random variable $X \sim \text{Exponential}(\beta)$. Show that the distribution of $X$ has the memoryless property.

**Solution:**

Let $0 \leq s \leq t$. We have that

$$P(X > s + t | X > s) = \frac{P(X > s + t)}{P(X > s)} = \frac{1 - F(s + t)}{1 - F(s)} = \frac{1 - \left(1 - e^{-\frac{s+t}{\beta}}\right)}{1 - \left(1 - e^{-\frac{s}{\beta}}\right)}$$

$$= e^{-\frac{t}{\beta}} = P(X > t).$$

Therefore, $X$ has the memoryless property.

15. (6 pts) An interesting property of the Exponential distribution is that its associated hazard function is constant on $[0, \infty)$. The hazard function associated to a density $f$ is defined as the function

$$h(x) = \frac{f(x)}{1 - F(x)}. \quad (3)$$

One way to interpret the hazard function is in terms of the failure rate of a certain component. In this sense, $h$ can be thought of as the ratio between the probability density of failure at time $x$, $f(x)$, and the probability that the component is still working at the same time, which is $P(X > x) = 1 - F(x)$.
Show that if \( X \sim \text{Exponential}(\beta) \) is a random variable describing the time at which a particular component fails to work, then the failure rate of \( X \) is constant (i.e. \( h \) does not depend on \( x \)).

**Solution:**
For \( x > 0 \) we have
\[
h(x) = \frac{f(x)}{1 - F(x)} = \frac{1}{\beta} e^{-\frac{x}{\beta}} \frac{1}{1 - \left( 1 - e^{-\frac{x}{\beta}} \right)} = \frac{1}{\beta}.
\]

16. **Extra credit** (10 pts) Consider the Binomial p.m.f.
\[
p(x) = \binom{n}{x} p^x (1 - p)^{n-x} \mathbb{1}_{\{0,1,\ldots,n\}}(x)
\]
with parameters \( n \) and \( p \). Show that
\[
\lim_{n \to \infty} p(x) = e^{-\lambda} \frac{\lambda^x}{x!} \mathbb{1}_{\{0,1,\ldots\}}(x)
\]
where the limit is taken as \( n \to \infty \), \( p \to 0 \), and \( np \to \lambda > 0 \).
(Hint: For three converging sequences \( a_n, b_n, c_n \) whose limits are \( a, b, c \) respectively, \( \lim_{n \to \infty} a_n b_n c_n = abc \).)

**Solution:**
Take any \( x \in \{0, 1, \ldots, n\} \). We have
\[
p(x) = \binom{n}{x} p^x (1 - p)^{n-x} = \frac{n!}{x!(n-x)!} \left( \frac{np}{n} \right)^x \left( 1 - \frac{np}{n} \right)^{n-x}
\]
\[
= \frac{n(n-1)\ldots(n-x+1)}{n^x} \frac{1}{x! (np)^x} \left( \frac{np}{n} \right)^n \left( 1 - \frac{np}{n} \right)^{-x}
\]
Now,
\[
\frac{n(n-1)\ldots(n-x+1)}{n^x} \to 1 \text{ as } n \to \infty
\]
\[
(np)^x \to \lambda^x \text{ as } np \to \lambda
\]
\[
\left( 1 - \frac{np}{n} \right)^n \to \left( 1 - \frac{\lambda}{n} \right)^n \text{ as } np \to \lambda \text{ and } \left( 1 - \frac{\lambda}{n} \right)^n \to e^{-\lambda} \text{ as } n \to \infty
\]
\[
\left( 1 - \frac{np}{n} \right)^{-x} \to \left( 1 - \frac{\lambda}{n} \right)^{-x} \text{ as } np \to \lambda \text{ and } \left( 1 - \frac{\lambda}{n} \right)^{-x} \to 1 \text{ as } n \to \infty.
\]
Thus,
\[
p(x) \to 1 \ast \frac{1}{x!} \ast \lambda^x \ast e^{-\lambda} \ast 1 = e^{-\lambda} \frac{\lambda^x}{x!}
\]
which is the p.m.f. of a Poisson distribution with parameter \( \lambda \).
Extra credit (10 pts) Consider the Normal p.d.f.

\[ f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

with parameters \( \mu \in \mathbb{R} \) and \( \sigma^2 > 0 \). Show that \( \mu - \sigma \) and \( \mu + \sigma \) correspond to the inflection points of \( f \).

**Solution:**

First, we need to compute the second derivative of \( f \). We have

\[ f'(x) = -\frac{1}{\sqrt{2\pi\sigma^2}} \frac{x - \mu}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]

and therefore

\[ f''(x) = -\frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( \frac{1}{\sigma^2} - \frac{(x-\mu)^2}{\sigma^4} \right) \]

\[ = -\frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \left( 1 - \frac{(x-\mu)^2}{\sigma^2} \right) \]

Notice that

\[ -\frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{\sigma^2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} < 0 \]

for any \( x \in \mathbb{R} \), therefore the sign of \( f'' \) depends on

\[ 1 - \frac{(x-\mu)^2}{\sigma^2} = \frac{\sigma^2 - (x-\mu)^2}{\sigma^2} \]

Clearly,

\[ \frac{\sigma^2 - (x-\mu)^2}{\sigma^2} \geq 0 \]

for \( |x-\mu| \leq \sigma \iff \mu - \sigma \leq x \leq \mu + \sigma \),

\[ \frac{\sigma^2 - (x-\mu)^2}{\sigma^2} \leq 0 \]

for \( |x-\mu| \geq \sigma \iff x \geq \mu + \sigma \lor x \leq \mu - \sigma \), and

\[ \frac{\sigma^2 - (x-\mu)^2}{\sigma^2} = 0 \]

for \( x = \mu \pm \sigma \). Therefore \( x = \mu \pm \sigma \) are the two inflection points of \( f \).
18. **Extra credit** (10 pts) Consider the random variable $X \sim \text{Beta}(\alpha, \beta)$ with $\alpha, \beta \in \mathbb{Z}^+$. Show that, for $n \in \mathbb{Z}^+$,

$$
\frac{E(X^{n+1})}{E(X^n)} = \frac{\alpha + n}{\alpha + \beta + n}.
$$

Use this result to show that

$$
E(X^n) = \prod_{i=0}^{n-1} \frac{\alpha + i}{\alpha + \beta + i}.
$$

**Solution:**

For the first claim, fix an arbitrary $n \in \mathbb{Z}^+$. Then

$$
E(X^n) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^n x^{\alpha-1}(1-x)^{\beta-1} \, dx
$$

$$
= \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha+n-1}(1-x)^{\beta-1} \, dx.
$$

Now,

$$
x^{\alpha+n-1}(1-x)^{\beta-1} \mathbb{1}_{[0,1]}(x)
$$

is the kernel of a Beta($\alpha + n$, $\beta$) p.d.f., hence its integral on $[0,1]$ must be equal to

$$
\int_0^1 x^{\alpha+n-1}(1-x)^{\beta-1} \, dx = \frac{\Gamma(\alpha + n)\Gamma(\beta)}{\Gamma(\alpha + \beta + n)}.
$$

Therefore,

$$
E(X^n) = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + n)\Gamma(\beta)}{\Gamma(\alpha)\Gamma(\beta)\Gamma(\alpha + \beta + n)}
= \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + n)}{\Gamma(\alpha + \beta + n)}
$$

Then,

$$
\frac{E(X^{n+1})}{E(X^n)} = \frac{\Gamma(\alpha + \beta)\Gamma(\alpha + n + 1)\Gamma(\alpha)}{\Gamma(\alpha)\Gamma(\alpha + \beta + n + 1)\Gamma(\alpha + \beta)\Gamma(\alpha + n)}
= \frac{\Gamma(\alpha + n + 1)\Gamma(\alpha + \beta + n)\Gamma(\alpha + n)}{\Gamma(\alpha + \beta + n + 1)\Gamma(\alpha + n)\Gamma(\alpha + \beta)}
= \frac{\alpha + n}{\alpha + \beta + n}.
$$
For the second claim, it is enough to notice that

$$E(X^n) = \prod_{i=0}^{n-1} \frac{E(X^{i+1})}{E(X^i)}.$$ 

From the previous result, we have that

$$\frac{E(X^{i+1})}{E(X^i)} = \frac{\alpha + i}{\alpha + \beta + i},$$

thus

$$E(X^n) = \prod_{i=0}^{n-1} \frac{E(X^{i+1})}{E(X^i)} = \prod_{i=0}^{n-1} \frac{\alpha + i}{\alpha + \beta + i}.$$ 

19. **Extra credit** (10 pt) Download the free version of R and RStudio (Desktop edition) on your personal computer. You can use the instructions here: [http://socserv.mcmaster.ca/jfox/Courses/R/ICPSR/R-install-instructions.html](http://socserv.mcmaster.ca/jfox/Courses/R/ICPSR/R-install-instructions.html).

(a) When you’re done, type `version` at the command line and hit enter. Write down the resulting version.string. Next, type `print("Hello, world!")`.

(b) Next, type `mysample = rexp(n=1000,rate=2)` to generate 1,000 i.i.d. samples from an exponential distribution. These are 200 possible wait time (in hours) for a bus that is expected to arrive once every half hour.

(c) Next, type `hist(mysample)` to produce a ‘histogram’ (an estimate of the probability density function (pdf); the outline of the bars will roughly take the shape of the pdf) of the exponential distribution.

(d) Next, use `abline(v=c(0.25,0.75), col=’red’)` to plot two vertical lines. The area between these lines is a good estimate of $P(0.25 \leq X \leq 0.75)$ for an exponential random variable $X$ with parameter $\lambda = 2$; or, in other words, the probability of the wait time being between 15 and 45 minutes.

(e) Lastly, use `dev.copy(png,’mydirectory/histogram.png’); graphics.off()` to save the image as a png file, in your home directory. Include this picture to your homework.
Homework 5 (due June 2nd)

1. (10 pts) **Simple example of multivariate joint distributions, and conditional distributions.** All car accidents that occurred in the state of Florida in 1998 and that resulted in an injury were recorded, along with the seat belt use of the injured person and whether the injury was fatal or not. Total number of recorded accidents is 577,006. This table can be used to describe the joint distribution of two binary random variables $X$ and $Y$, where $X$ is 1 if the passenger does not wear the seat belt and 0 otherwise and $Y$ is 1 if the accident
is fatal or not.

(a) What is the probability that an accident is fatal and the passenger did not wear the seat belt?

(b) What is the probability that an accident is fatal and the passenger wore the seat belt?

(c) What is the marginal distribution of \(X\)? (Hint: just in case you’re wondering what you need to produce, notice \(X\) is a discrete random variable. You should completely produce probability mass function (or CDF) of \(X\)!

(d) What is the conditional distribution of \(Y|X = \text{No Belt}\) and \(Y|X = \text{Belt}\)? (the same hint applies here)

**Solution:**

- Read from the table, \(P(A) = 1601/577006 = 0.0028\)
- \(P(B) = 162527/577006 = 0.28\)
- \(P(X = 1) = (1601 + 510)/577006 = 0.0037\), \(P(X = 0) = (1601 + 510)/577006 = 0.9963\)
- \(P(Y = \text{fatal}|X = \text{No belt}) = \frac{P(Y=\text{fatal},X=\text{No belt})}{P(X=\text{No belt})} = 1601/(1601+510) = 0.76.\)
- \(P(Y = \text{Non-fatal}|X = \text{No belt}) = 510/(1601+510) = 0.24.\)
- \(P(Y = \text{fatal}|X = \text{belt}) = \frac{P(Y=\text{fatal},X=\text{belt})}{P(X=\text{belt})} = 162527/(162527+412368) = 0.28.\)
- \(P(Y = \text{Non-fatal}|X = \text{belt}) = 412368/(162527+412368) = 0.72.\)

2. (10 pts) Consider the following function:

\[
f(x_1, x_2) = \begin{cases} 
6x_1^2x_2 & \text{if } 0 \leq x_1 \leq x_2 \land x_1 + x_2 \leq 2 \\
0 & \text{otherwise.} 
\end{cases}
\]

- Verify that \(f\) is a valid bivariate p.d.f.
• Suppose \((X_1, X_2) \sim f\). Are \(X_1\) and \(X_2\) independent random variables?

• Compute \(P(Y_1 + Y_2 < 1)\)

• Compute the marginal p.d.f. of \(X_1\). What are the values of its parameters?

• Compute the marginal p.d.f. of \(X_2\)

• Compute the conditional p.d.f. of \(X_2\) given \(X_1\)

• Compute \(P(X_2 < 1.1|X_1 = 0.6)\).

**Solution:**

It is clear that \(f_{X_1,X_2}(x_1, x_2) \geq 0\) for any \(x_1, x_2 \in \mathbb{R}\). We just need to check that \(f_{X_1,X_2}\) integrates to 1:

\[
\int_{\mathbb{R}} \int_{\mathbb{R}} f_{X_1,X_2}(x_1, x_2) \, dx_1 \, dx_2 = \int_{0}^{1} \int_{x_1}^{(2-x_1)} 6x_1^2 x_2 \, dx_2 \, dx_1
\]

\[
= 6 \int_{0}^{1} x_1^2 \int_{x_1}^{(2-x_1)} x_2 \, dx_2 \, dx_1 = 3 \int_{0}^{1} x_1^2 x_2^2 |_{x_1}^{2-x_1} \, dx_1
\]

\[
= 3 \int_{0}^{1} x_1^2 [(2-x_1)^2 - x_1^2] \, dx_1 = 12 \int_{0}^{1} x_1^2 (1 - x_1) \, dx_1
\]

\[
= 12 \frac{\Gamma(3) \Gamma(2)}{\Gamma(5)} = 12 \frac{2}{4!} = 24 \frac{24}{24} = 1.
\]

• \(X_1\) and \(X_2\) are not independent since the support of \(f_{X_1,X_2}\) (the union of the green region and the red region in the figure) is not a rectangular region.
\begin{itemize}
  \item The region corresponding to the event $X_1 + X_2 < 1$ is the red region in the figure. We have

  \[ P(X_1 + X_2 < 1) = \int_0^1 \int_{x_1}^{1-x_1} 6x_1^2 x_2 \, dx_2 \, dx_1 \]
  \[ = 6 \int_0^1 x_1^2 \int_{x_1}^{1-x_1} x_2 \, dx_2 \, dx_1 = 3 \int_0^1 x_1^2 \left[ x_2 \right]_{x_1}^{1-x_1} \, dx_1 \]
  \[ = 3 \int_0^1 x_1^2 \left[ (1 - x_1)^2 - x_1^2 \right] \, dx_1 = 3 \int_0^1 x_1^2 (1 - 2x_1) \, dx_1 \]
  \[ = x_1^3 \left[ \frac{1}{2} - \frac{3}{2} x_1^2 \right]_0^1 = \frac{1}{8} - \frac{3}{32} = \frac{1}{32}. \]

  \item Notice that $\text{supp}(X_1) = [0, 1]$. For $x_1 \in [0, 1]$ we thus have

  \[ f_{X_1}(x_1) = \int_{x_1}^{2-x_1} 6x_1^2 x_2 \, dx_2 = 3x_1^2 \left[ x_2 \right]_{x_1}^{2-x_1} \]
  \[ = 3x_1^2 \left[ (2 - x_1)^2 - x_1^2 \right] = 12x_1^2 (1 - x_1) \]

  which we recognize as the p.d.f. of a Beta(3, 2) distribution.

  \item Notice that $\text{supp}(X_2) = [0, 2]$. For $0 \leq x_2 < 1$ we have

  \[ f_{X_2}(x_2) = \int_0^x 6x_1^2 x_2 \, dx_1 = 2x_2 \left[ x_1^3 \right]_0^x \]
  \[ = 2x_2^4. \]

  For $1 \leq x_2 \leq 2$ we have

  \[ f_{X_2}(x_2) = \int_0^{2-x_2} 6x_1^2 x_2 \, dx_1 = 2x_2 \left[ x_1^3 \right]_0^{2-x_2} \]
  \[ = 2x_2 (2 - x_2)^3. \]

  \item The conditional p.d.f. of $X_2$ given $X_1 = x_1$ is only well-defined for $x_1 \in \text{supp}(X_1) = [0, 1]$. For $x_1 \in (0, 1)$,

  \[ f_{X_2|X_1=x_1}(x_2) = \frac{f_{X_1,X_2}(x_1,x_2)}{f_{X_1}(x_1)} = \frac{6x_1^2 x_2 \mathbb{1}_{[x_1,2-x_1]}(x_2)}{12x_1^2 (1 - x_1)} \]
  \[ = \frac{x_2}{2(1 - x_1)} \mathbb{1}_{[x_1,2-x_1]}(x_2). \]
\end{itemize}
• We have

\[ P(X_2 < 1.1 | X_1 = 0.6) = \int_{-\infty}^{1.1} f_{X_2|X_1=0.6}(x_2) \, dx_2 \]

\[ = \int_{-\infty}^{1.1} \frac{x_2}{2(1 - 0.6)} \mathbb{1}_{[0.6,1.4]}(x_2) \]

\[ = \frac{1}{1.6} x_2^{1.1} |_{0.6}^{1.1} = \frac{1.21 - 0.36}{1.6} = 0.53125. \]

3. (10 pts) Consider the following bivariate p.m.f.:

\[ p_{X_1,X_2}(x_1, x_2) = \begin{cases} 
0.38 \text{ if } (x_1, x_2) = (0,0) \\
0.17 \text{ if } (x_1, x_2) = (1,0) \\
0.14 \text{ if } (x_1, x_2) = (0,1) \\
0.02 \text{ if } (x_1, x_2) = (1,1) \\
0.24 \text{ if } (x_1, x_2) = (0,2) \\
0.05 \text{ if } (x_1, x_2) = (1,2). 
\end{cases} \]

• Compute the marginal p.m.f. of \( X_1 \) and the marginal p.m.f. of \( X_2 \)

• Compute \( p_{X_2|X_1=0} \), the conditional p.m.f. of \( X_2 \) given \( X_1 = 0 \)

• Compute \( P(X_1 = 0 | X_2 = 2) \)?

Solution:

• Notice first that \( \text{supp}(X_1) = \{0,1\} \). Thus, we have

\[ p_{X_1}(x_1) = \begin{cases} 
p_{X_1,X_2}(0,0) + p_{X_1,X_2}(0,1) + p_{X_1,X_2}(0,2) \text{ if } x_1 = 0 \\
p_{X_1,X_2}(1,0) + p_{X_1,X_2}(1,1) + p_{X_1,X_2}(1,2) \text{ if } x_1 = 1 \\
0 \text{ if } x_1 \notin \{0,1\}
\end{cases} \]

\[ = \begin{cases} 
0.38 + 0.14 + 0.24 \text{ if } x_1 = 0 \\
0.17 + 0.02 + 0.05 \text{ if } x_1 = 1 \\
0 \text{ if } x_1 \notin \{0,1\}
\end{cases} = \begin{cases} 
0.76 \text{ if } x_1 = 0 \\
0.24 \text{ if } x_1 = 1 \\
0 \text{ if } x_1 \notin \{0,1\}.
\end{cases} \]

Now notice that \( \text{supp}(X_2) = \{0,1,2\} \). We have
• We have

\[
p_{X_2}(x_2) = \begin{cases} 
    p_{X_1,X_2}(0,0) + p_{X_1,X_2}(1,0) & \text{if } x_2 = 0 \\
    p_{X_1,X_2}(0,1) + p_{X_1,X_2}(1,1) & \text{if } x_2 = 1 \\
    p_{X_1,X_2}(0,2) + p_{X_1,X_2}(1,2) & \text{if } x_2 = 2 \\
    0 & \text{if } x_2 \notin \{0, 1, 2\}
\end{cases}
\]

\[
= \begin{cases} 
    0.38 + 0.17 & \text{if } x_2 = 0 \\
    0.14 + 0.02 & \text{if } x_2 = 1 \\
    0.24 + 0.05 & \text{if } x_2 = 2 \\
    0 & \text{if } x_2 \notin \{0, 1, 2\}
\end{cases}
\]

• We have

\[
p_{X_2|X_1=0}(x_2) = \frac{p_{X_1,X_2}(0,x_2)}{p_{X_1}(0)} = \begin{cases} 
    0.38 & \text{if } x_2 = 0 \\
    0.14 & \text{if } x_2 = 1 \\
    0.24 & \text{if } x_2 = 2 \\
    0 & \text{if } x_2 \notin \{0, 1, 2\}
\end{cases}
\]

\[
\approx \begin{cases} 
    0.5 & \text{if } x_2 = 0 \\
    0.184 & \text{if } x_2 = 1 \\
    0.316 & \text{if } x_2 = 2 \\
    0 & \text{if } x_2 \notin \{0, 1, 2\}
\end{cases}
\]

4. (10 pts) Let \(X_1 \sim \text{Uniform}(0, 1)\) and, for \(0 < x_1 \leq 1\),

\[
f_{X_2|X_1=x_1}(x_2) = \frac{1}{x_1} \mathbb{1}_{(0,x_1]}(x_2).
\]

• What named continuous distribution does \(f_{X_2|X_1=x_1}\) seem to resemble? What are the values of its parameters?
• Compute the joint p.d.f. of \((X_1, X_2)\)
• Compute the marginal p.d.f. of \(X_2\).
Solution:

- It is clear that \( f_{X_2|X_1=x_1} \) for \( 0 < x_1 \leq 1 \) is a Uniform p.d.f. over the interval \((0, x_1]\).
- We have
  \[
  f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2|X_1=x_1}(x_2) = \mathbb{1}_{(0,1]}(x_1) \frac{1}{x_1} \mathbb{1}_{(0,x_1]}(x_2)
  = \frac{1}{x_1} \mathbb{1}_{\{0<x_2\leq x_1\leq 1\}}(x_1, x_2).
  \]
- We have, for \( x_2 \in (0, 1] \),
  \[
  f_{X_2}(x_2) = \int_{x_2}^{1} \frac{1}{x_1} \, dx_1 = \log(x_1)|_{x_2}^{1} = -\log(x_2)
  \]
  otherwise \( f_{X_2}(x_2) = 0 \) for \( x_2 \notin (0, 1] \).

5. (5 pts) Let \( X_1 \sim \text{Binomial}(n_1, p) \) and \( X_2 \sim \text{Binomial}(n_2, p) \). Assume that \( X_1 \) and \( X_2 \) are independent. Consider \( W = X_1 + X_2 \). One can show that \( W \sim \text{Binomial}(n_1 + n_2, p) \). Use this result to show that \( p_{X_1|W=w} \) is a Hypergeometric p.m.f. with parameters \( r = n_1 \), \( N = n_1 + n_2 \), and \( n = w \).

Solution:

We have, for \( x_1 \in \{0, 1, \ldots, w\} \)

\[
p_{X_1|W=w}(x_1) = P(X_1 = x_1|W = w) = \frac{P(X_1 = x_1 \cap W = w)}{P(W = w)} = \frac{P(X_1 = x_1 \cap X_2 = w - x_1)}{P(W = w)}
= \frac{P(X_1 = x_1) P(X_2 = w - x_1)}{P(W = w)}
= \binom{n_1}{x_1} p^{x_1} (1-p)^{n_1-x_1} \binom{n_2}{w-x_1} p^{w-x_1} (1-p)^{n_2-w+x_1}
= \binom{n_1}{x_1} \binom{n_2}{w-x_1}
= \binom{n_1}{x_1} \binom{n_2}{w-x_1}
= \binom{n_1+n_2}{w}
and clearly \( p_{X_1|W=w}(x_1) = 0 \) if \( x_1 \notin \{0, 1, \ldots, w\} \).

6. (10 pts) Suppose that you have a coin such that, if tossed, head appears with probability \( p \in [0, 1] \) and tail occurs with probability \( 1-p \). Person A tosses the coin. The number of tosses until the first head appears for
person A is $X_1$. Then person B tosses the coin. The number of tosses until the first head appears for person B is $X_2$. Assume that $X_1$ and $X_2$ are independent. What is the probability of the event $X_1 = X_2$?

**Solution:**

Obviously, $X \sim \text{Geometric}(p)$ and $Y \sim \text{Geometric}(p)$. Because $X$ and $Y$ are independent, we have

$$p_{X_1,X_2}(x_1, x_2) = p_{X_1}(x_1)p_{X_2}(x_2) = (1 - p)^{x_1-1}p\mathbb{1}_{\{1,2,\ldots\}}(x_1)(1 - p)^{x_2-1}p\mathbb{1}_{\{1,2,\ldots\}}(x_2)$$

$$= (1 - p)^{x_1+x_2-2}p^2\mathbb{1}_{\{1,2,\ldots\}}(x_1)\mathbb{1}_{\{1,2,\ldots\}}(x_2).$$

Then,

$$P(X_1 = X_2) = \sum_{x_1 = x_2 = x \in \{1,2,\ldots\}} p_{X_1,X_2}(x_1, x_2) = \sum_{x \in \{1,2,\ldots\}} p_{X_1,X_2}(x, x)$$

$$= p^2 \sum_{x \in \{1,2,\ldots\}} (1 - p)^{2x-2} = p^2 \sum_{x \in \{1,2,\ldots\}} [(1 - p)^2]^{x-1}$$

$$= \frac{p^2}{1 - (1 - p)^2} = \frac{p}{2 - p}.$$ 

7. (10 pts) Let $(X_1, X_2) \sim f_{X_1,X_2}$ with

$$f_{X_1,X_2}(x_1, x_2) = \frac{1}{x_1} \mathbb{1}_{\{0 < x_2 \leq x_1 \leq 1\}}(x_1, x_2).$$

Compute $E(X_1 - X_2)$.

**Solution:**

We have $E(X_1 - X_2) = E(X_1) - E(X_2)$. First we compute the marginal p.d.f.’s. They are

$$f_{X_1}(x_1) = \int_{\mathbb{R}} f_{X_1,X_2}(x_1, x_2) \, dx_2 = \int_{\mathbb{R}} \frac{1}{x_1} \mathbb{1}_{\{0 < x_2 \leq x_1 \leq 1\}}(x_1, x_2) \, dx_2$$

$$= \frac{1}{x_1} \mathbb{1}_{(0,1]}(x_1) \int_{\mathbb{R}} \mathbb{1}_{(0,x_1]}(x_2) \, dx_2$$

$$= \frac{1}{x_1} \mathbb{1}_{(0,1]}(x_1) \int_0^{x_1} \, dx_2$$

$$= \mathbb{1}_{(0,1]}(x_1)$$
and

\[ f_{X_2}(x_2) = \int_{\mathbb{R}} f_{X_1, X_2}(x_1, x_2) \, dx_1 = \int_{\mathbb{R}} \frac{1}{x_1} \mathbb{1}_{\{0 < x_2 \leq x_1 \leq 1\}}(x_1, x_2) \, dx_1 \]

\[ = \int_{\mathbb{R}} \frac{1}{x_1} \mathbb{1}_{\{0, 1\}}(x_2) \mathbb{1}_{[x_2, 1]}(x_2) \, dx_1 \]

\[ = \mathbb{1}_{\{0, 1\}}(x_2) \int_{x_2}^{1} \frac{1}{x_1} \, dx_1 = -\log(x_2) \mathbb{1}_{\{0, 1\}}(x_2). \]

Notice that \( f_{X_1} \) is the p.d.f. of Uniform(0, 1), therefore \( E(X_1) = 1/2 \). On the other hand

\[ E(X_2) = -\int_{0}^{1} x_2 \log(x_2) \, dx_2 = -\left[ \frac{x_2^2}{2} \log(x_2) - \int_{0}^{x_2} \frac{x_2}{2} \, dx_2 \right]_{0}^{1} \]

\[ = \frac{1}{4} \left[ x_2^2 \right]_{0}^{1} = \frac{1}{4}. \]

Therefore \( E(X_1 - X_2) = E(X_1) - E(X_2) = 1/2 - 1/4 = 1/4. \)

8. (10 pts) Consider

\[ f_{X_1, X_2}(x_1, x_2) = 6(1 - x_2) \mathbb{1}_{\{0 \leq x_1 \leq x_2 \leq 1\}}(x_1, x_2). \]

• Are \( X_1 \) and \( X_2 \) independent?
• Are \( X_1 \) and \( X_2 \) correlated (positively or negatively) or uncorrelated?

**Solution:**
• $X_1$ and $X_2$ are not independent. We can see this in two ways:
  1) we draw the support of their joint p.d.f. (which is the set indicated by the indicator function; the red triangle in the figure) and we easily see that it does not correspond to a rectangular region;
  2) we rewrite the indicator function as

$$
\mathbb{1}_{0 \leq x_1 \leq x_2 \leq 1}(x_1, x_2) = \mathbb{1}_{[0, x_2]}(x_1)\mathbb{1}_{[x_1, 1]}(x_2)
$$

and then we easily see that there is no way to factorize the joint p.d.f. into two functions $g$ and $h$ such that $g$ only depends on $x_1$ and $h$ only depends on $x_2$.

• To answer the second question, we need to compute $\text{Cov}(X_1, X_2)$. For that we need $E(X_1)$, $E(X_2)$, and $E(X_1X_2)$. We have

$$
E(X_1X_2) = \int_{\mathbb{R}} \int_{\mathbb{R}} x_1x_2f_{X_1, X_2}(x_1, x_2) \, dx_1 \, dx_2
$$

and then

$$
E(X_1) = \int_{0}^{1} x_1 \left( \int_{0}^{x_2} 6(1 - x_2) \, dx_2 \right) \, dx_1
$$

Then,

$$
E(X_1) = \int_{0}^{1} x_1 \left( \int_{0}^{x_2} \left( \frac{x_1}{2} - x_1^2 + \frac{x_1^3}{3} \right) \, dx_2 \right) \, dx_1
$$

and

$$
E(X_1X_2) = \frac{3}{43}.
$$
and

\[ E(X_2) = \int_0^1 6x_2(1-x_2) \int_0^{x_2} dx_1 dx_2 = 6 \int_0^1 x_2^2(1-x_2) dx_2 = 6 \frac{\Gamma(3) \Gamma(2)}{\Gamma(5)} = \frac{6 \cdot 2}{4!} = \frac{12}{24} = \frac{1}{2}. \]

Thus, \( \text{Cov}(X_1, X_2) = E(X_1X_2) - E(X_1)E(X_2) = \frac{3}{20} - \frac{1}{8} = \frac{3}{20} - \frac{1}{8} = \frac{1}{40}. \) Since the covariance is positive, \( X_1 \) and \( X_2 \) are positively correlated and therefore not independent.

9. (10 pts) Let \( X_1 \) and \( X_2 \) be uncorrelated random variables. Consider \( Y_1 = X_1 + X_2 \) and \( Y_2 = X_1 - X_2. \)

- Express \( \text{Cov}(Y_1, Y_2) \) in terms of \( \text{V}(X_1) \) and \( \text{V}(X_2) \)
- Give an expression for \( \text{Cor}(Y_1, Y_2) \)
- Is it possible that \( \text{Cor}(Y_1, Y_2) = 0? \) If so, when does this happen?

**Solution:**

- We have
  \[ \text{Cov}(Y_1, Y_2) = \text{Cov}(X_1 + X_2, X_1 - X_2) \]
  \[ = \text{Cov}(X_1, X_1) - \text{Cov}(X_1, X_2) + \text{Cov}(X_2, X_1) - \text{Cov}(X_2, X_2) \]
  \[ = \text{V}(X_1) - \text{V}(X_2). \]

- By definition
  \[ \text{Cor}(Y_1, Y_2) = \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{\text{V}(Y_1)\text{V}(Y_2)}}. \]

We have

\[ \text{V}(Y_1) = \text{V}(X_1 + X_2) = \text{V}(X_1) + \text{V}(X_2) + 2\text{Cov}(X_1, X_2) \]
\[ = \text{V}(X_1) + \text{V}(X_2) + 0 = \text{V}(X_1) + \text{V}(X_2) \]

and

\[ \text{V}(Y_2) = \text{V}(X_1 - X_2) = \text{V}(X_1) + \text{V}(X_2) - 2\text{Cov}(X_1, X_2) \]
\[ = \text{V}(X_1) + \text{V}(X_2) + 0 = \text{V}(X_1) + \text{V}(X_2), \]

thus

\[ \text{Cor}(Y_1, Y_2) = \frac{\text{Cov}(Y_1, Y_2)}{\sqrt{\text{V}(Y_1)\text{V}(Y_2)}} = \frac{\text{V}(X_1) - \text{V}(X_2)}{\text{V}(X_1) + \text{V}(X_2)}. \]
• If \( V(X_1) > V(X_2) \), then \( Y_1 \) and \( Y_2 \) are positively correlated.
  If \( V(X_1) < V(X_2) \), then \( Y_1 \) and \( Y_2 \) are negatively correlated.
  Finally, If \( V(X_1) = V(X_2) \), then \( Y_1 \) and \( Y_2 \) are uncorrelated.

10. (10 pts) Let

\[
f_{X_1}(x_1) = \frac{1}{6} x_1^3 e^{-x_1} \mathbb{1}_{[0,\infty)}(x_1)
\]

and

\[
f_{X_2}(x_2) = \frac{1}{2} e^{-\frac{x_2^2}{2}} \mathbb{1}_{[0,\infty)}(x_2)
\]

Let \( Y = X_1 - X_2 \).

- Compute \( E(Y) \)
- Let \( \text{Cov}(X_1, X_2) = \gamma \). Compute \( V(Y) \)
- Suppose that \( X_1 \) and \( X_2 \) are uncorrelated. What is \( V(Y) \) in this case?

Solution:

- It is clear from the inspection of the two p.d.f.'s that \( X_1 \sim \text{Gamma}(4, 1) \) and \( X_2 \sim \text{Exponential}(2) \). Thus, \( E(Y) = E(X_1 - X_2) = E(X_1) - E(X_2) = 4 - 2 = 2 \).
- \( V(Y) = V(X_1 - X_2) = V(X_1) + V(X_2) - 2\text{Cov}(X_1, X_2) = 4 + 4 - 2\gamma = 2(4 - \gamma) \).
- If \( X_1 \) and \( X_2 \) are uncorrelated, then \( \gamma = 0 \) and \( V(Y) = 8 \).

11. (5 pts) Use the definition of variance to show that \( V(X \pm Y) = V(X) + V(Y) \pm 2\text{Cov}(X, Y) \).

Solution: We have

\[
V(X \pm Y) = E[(X \pm Y)^2] - [E(X \pm Y)]^2
= E(X^2 + Y^2 \pm 2XY) - [E(X)\pm E(Y)]^2
= E(X^2) - [E(X)]^2 + E(Y^2) - [E(Y)]^2 \pm 2E(X)E(Y)
= V(X) + V(Y) \pm 2[E(XY) - E(X)E(Y)] = V(X) + V(Y) \pm 2\text{Cov}(X,Y).
\]

12. Extra credit (10 pts) Let \( Z \sim \mathcal{N}(0, 1) \). Show that for any odd \( n \geq 1 \),
\( E(Z^n) = 0 \).
Solution: Let \( f \) denote the p.d.f. associated to \( \mathcal{N}(0, 1) \). Notice that \( f \) is an even function, i.e. \( f(z) = f(-z) \) for any \( z \in \mathbb{R} \). We have

\[
E(Z^n) = \int_{\mathbb{R}} z^n f(z) \, dz
\]

\[
= \int_{-\infty}^{0} z^n f(z) \, dz + \int_{0}^{+\infty} z^n f(z) \, dz
\]

\[
= \int_{-\infty}^{0} z^n (-z) \, dz + \int_{0}^{+\infty} z^n f(z) \, dz
\]

\[
= \int_{0}^{-\infty} (-z)^n f(-z) \, dz + \int_{0}^{+\infty} z^n f(z) \, dz
\]

\[
= \int_{0}^{-\infty} (-z)^n f(-z) \, dz + \int_{0}^{+\infty} z^n f(z) \, dz
\]

Make the change of variable \( x = -z \) in the first integral. We get

\[
\int_{0}^{-\infty} (-z)^n f(-z) \, dz + \int_{0}^{+\infty} z^n f(z) \, dz
\]

\[
= \int_{0}^{+\infty} (x^n) f(x)(-1) \, dx + \int_{0}^{+\infty} z^n f(z) \, dz
\]

\[
= -\int_{0}^{+\infty} (x^n) f(x) \, dx + \int_{0}^{+\infty} z^n f(z) \, dz = 0.
\]

13. Extra Credit: (10 pts) Show that for two random variables \( X \) and \( Y \), \( Cor(X, Y) \in [-1, 1] \).

Hint: consider \( X' = X - E(X) \) and \( Y' = Y - E(Y) \) and study the quadratic function \( h(t, X', Y') = (tX' - Y')^2 \) as a function of \( t \). Notice that \( h(t, X', Y') \geq 0 \), therefore \( E(h(t, X', Y')) \geq 0 \) which implies that the discriminant of \( h(t, X', Y') \) as a quadratic function of \( t \) must be non-positive! Try to use this fact to prove the claim that \( Cor(X, Y) \in [-1, 1] \).

Solution: We have \( E(h(t, X', Y')) \geq 0 \). This means

\[
E[(tX' - Y')^2] = E(t^2 X'^2 + Y'^2 - 2tX'Y')
\]

\[
= t^2 E(X'^2) - 2t E(X'Y') + E(Y'^2) \geq 0
\]

This can only happen if the discriminant is non-positive, i.e. if

\[
4[E(X'Y')]^2 - 4E(X'^2)E(Y'^2) \leq 0,
\]
i.e.

\[
[E(X'Y')]^2 \leq E(X'^2)E(Y'^2)
\]

or, equivalently,

\[
\frac{[E(X'Y')]^2}{E(X'^2)E(Y'^2)} \leq 1.
\]

Notice that

\[
\frac{[E(X'Y')]^2}{E(X'^2)E(Y'^2)} = \frac{[E((X - E(X))(Y - E(Y)))]^2}{E((X - E(X))^2)E((Y - E(Y))^2)}
\]

\[
= \frac{\text{Cov}(X,Y)^2}{V(X)V(Y)} = [\text{Cor}(X,Y)]^2.
\]

Thus, it follows that \( |\text{Cor}(X,Y)| \leq 1 \) or, equivalently, \(-1 \leq \text{Cor}(X,Y) \leq 1\).

14. **Extra Credit** (10 pts): In an earlier homework, you downloaded R and RStudio and printed Hello, world!, and generated some samples from a given distribution. We will use R again for a fun new experiment. The Hypergeometric distribution can be approximated by a Binomial distribution. Specifically, \( \text{HG}(n,N,r) \approx \text{Bin}(n,p = r/N) \) when \( N \geq 20n \), as a rule of thumb! (**Fun exercise:** first think about intuitively why this is true, then prove that the former distribution goes to the latter as \( N \to \infty \) and \( p = r/N! \))

Conduct a simulation in R to demonstrate that this is true. Start with the code below and make modifications as you see fit. Among other things, you may need to increase the number of simulations (n.sample), adjust the parameters \( (n,N,r) \) so that we have a valid scenario where the Binomial distribution can approximate the Hypergeometric distribution, and adjust the plotting parameters so that you get a meaningful, easy to read graphic that shows that the two simulated distributions are approximately the same.

```r
# Initialize some useful information
n.sample <- 100
n <- 500
N <- 10000
r <- 750
p <- r/N

# Generate random sample from binomial distribution
# Dont confuse n and size here. The folks who wrote this function
```

47
# used some unfortunately confusing parameter names
bin.sample <- rbinom(n = n.sample, size = n, prob = r/N)
# Generate random sample from hypergeometric distribution
# Don't get confused here -- the folks who write this function
# used a different parameterization of the hypergeometric distribution
hg.sample <- rhyper(nn = n.sample, m = r, n = N-r, k = n)
# Plot the results
par(mfrow = c(2,1))
hist(bin.sample, breaks = 50,
     main = "Simulated Binomial Distribution", col = 2,
     xlab = "Hey there, fix my axis label!!")
hist(hg.sample, breaks = 50,
     main = "Simulated Hypergeometric Distribution", col = 3,
     xlab = "Master Chief")
Homework 6 (due June 9th)

1. (5 pts) Let $X \mid Q = q \sim \text{Binomial}(n, q)$ and $Q \sim \text{Beta}(\alpha, \beta)$. Compute $E(X)$.

**Solution:**

We have

$$E(X) = E[E(X \mid Q)] = E(nQ) = nE(Q) = \frac{n\alpha}{\alpha + \beta}.$$

2. (5 pts) Let $X \mid Q = q \sim \text{Binomial}(n, q)$ and $Q \sim \text{Beta}(3, 2)$. Find the marginal p.d.f. of $X$. (Notice that here we are mixing discrete and continuous random variables, but the usual definitions still apply!)

**Solution:**

The support of $X$ is supp($X$) = \{0, 1, \ldots, n\}. For $x \in \text{supp}(X)$ and $q \in (0, 1)$ we have

$$p_X(x) = \int_0^1 p_{X,Q}(x,q) \, dq = \int_0^1 p_{X\mid Q=q}(x)p_Q(q) \, dq$$

$$= \int_0^1 \binom{n}{x} q^x (1-q)^{n-x} \frac{\Gamma(3 + 2)}{\Gamma(3)\Gamma(2)} q^2 (1-q) \, dq$$

$$= 12 \binom{n}{x} \int_0^1 q^{2+x}(1-q)^{n-x+1} \, dq$$

$$= 12 \binom{n}{x} \frac{\Gamma(3 + x)\Gamma(n - x + 2)}{\Gamma(3 + x + n - x + 2)}$$

$$= 12 \binom{n}{x} \frac{\Gamma(3 + x)\Gamma(n - x + 2)}{\Gamma(5 + n)}.$$

For any $x \notin \text{supp}(X)$, of course $p_X(x) = 0$.

3. (5 pts) Let $X \sim \text{Exponential}(4)$. Consider $Y = 3X + 1$. Find the p.d.f. of $Y$ and $E(Y)$.

**Solution:**

We have $g(x) = 3x + 1$ thus the support of $Y$ is $[1, \infty)$, $g^{-1}(x) = \frac{x-1}{3}$, and

$$\frac{d}{dx} g^{-1}(x) = \frac{1}{3}.$$
It follows that, for $y \in [1, \infty)$,

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dx} g^{-1}(x) \right|_{x=y} = \frac{1}{4} e^{-\frac{y-1}{12}}$$

otherwise $f_Y(y) = 0$. Now,

$$E(Y) = \int_1^\infty y \frac{1}{12} e^{-\frac{y-1}{12}} dy = \int_0^\infty (12x + 1)e^{-x} dx$$

$$= 12 \int_0^\infty xe^{-x} dx + \int_0^\infty e^{-x} dx = 12 + 1 = 13.$$

4. (10 pts) We mentioned in class that if $X \sim \text{Gamma}(\alpha, \beta)$ then, for $c > 0$, $Y = cX \sim \Gamma(\alpha, c\beta)$. Prove it.

**Solution:**

We have $g(x) = cx$, $g^{-1}(x) = x/c$, and

$$\frac{d}{dx} g^{-1}(x) = \frac{1}{c}.$$

The image of the support of $X$ through $g$ is still $[0, \infty)$. Then, for $y \in [0, \infty)$,

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dx} g^{-1}(x) \right|_{x=y}$$

$$= \frac{1}{\beta^\alpha \Gamma(\alpha)} \left( \frac{y}{c} \right)^{\alpha-1} e^{-\frac{y}{c\beta}} \frac{1}{c}$$

otherwise $f_Y(y) = 0$. This is clearly the p.d.f. of a $\text{Gamma}(\alpha, c\beta)$ distribution.

5. (10 pts) Let $F$ be a given continuous and strictly increasing c.d.f.. Let $U \sim \text{Uniform}(0, 1)$. Find a function $g$ such that $g(U) \sim F$. You can assume that $g$ is strictly increasing.
Solution:
Let \( Y = g(U) \). We have

\[
F(y) = P(Y \leq y) = P(g(U) \leq y) = P(U \leq g^{-1}(y)) = \begin{cases} 
0 & \text{if } g^{-1}(y) < 0 \\
g^{-1}(y) & \text{if } g^{-1}(y) \in [0, 1) \\
1 & \text{if } g^{-1}(y) \geq 1.
\end{cases}
\]

Thus, it is enough to set \( g^{-1} = F \), i.e. \( g = F^{-1} \), and then \( Y = g(U) \sim F \).

6. (10 pts) Let \( X \sim f_X \) with\n
\[
f_X(x) = \frac{b}{x^2} 1_{[b, \infty)}(x)
\]

and let \( U \sim \text{Uniform}(0, 1) \). Find a function \( g \) such that \( g(U) \) has the same distribution of \( X \).

Solution:
Using the result of exercise 5, we need to set \( g = F_X^{-1} \). We have

\[
F_X(x) = \begin{cases} 
0 & \text{if } x < b \\
\frac{x}{b} \int_b^x f_X(y) \, dy & \text{if } x \geq b
\end{cases} = \begin{cases} 
0 & \text{if } x < b \\
\frac{x}{b} \int_b^x \frac{b}{y^2} \, dy & \text{if } x \geq b
\end{cases} = \begin{cases} 
0 & \text{if } x < b \\
1 - \frac{b}{x} & \text{if } x \geq b.
\end{cases}
\]

It follows that

\[
g(x) = F_X^{-1}(x) = \frac{b}{1 - x}.
\]

7. (10 pts) Let \( \theta > 0 \) and \( X \sim f_X \) with\n
\[
f_X(x) = \frac{2x}{\theta} e^{-\frac{x^2}{\theta}} 1_{[0, \infty)}(x).
\]

Consider \( Y = X^2 \). What is \( f_Y \), the p.d.f. of \( Y \)? What are \( E(Y) \) and \( V(Y) \).

Solution:
Notice that the function \( g(x) = x^2 \) is invertible over the support of
$X, [0, \infty)$. Furthermore, $g$ is such that $\text{supp}(Y) = \text{supp}(X)$. We have $g^{-1}(x) = \sqrt{x}$ and
\[
\frac{d}{dx}g^{-1}(x) = \frac{1}{2\sqrt{x}}.
\]
Therefore
\[
f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dx}g^{-1}(x) \right|_{x=y} = \frac{2\sqrt{y}}{\theta} e^{-\frac{y}{\theta}} 1_{[0,\infty)}(\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\theta} e^{-\frac{y}{\theta}} 1_{[0,\infty)}(y).
\]

It follows that $Y \sim \text{Exponential}(\theta)$. Thus, $E(Y) = \theta$ and $V(Y) = \theta^2$.

8. **Extra credit** (10pts) Let $X \sim \text{Beta}(\alpha, \beta)$. Consider $Y = 1 - X$. What is the probability distribution of $Y$? Now, let $\alpha = \beta = 1$. Claim that $X$ and $Y$ have the same distribution. Are $X$ and $Y$ also independent?

**Solution:**
Here $g(x) = 1 - x$, so that $g^{-1}(x) = g(x)$ and $\frac{d}{dx}g^{-1}(x) = -1$. Notice that, for this $g$, $\text{supp}(Y) = \text{supp}(X) = [0,1]$. Then, for $y \in [0,1]$ we have
\[
f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dx}g^{-1}(x) \right|_{x=y} = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} (1 - y)^{\alpha - 1} y^{\beta - 1} \cdot 1
\]

otherwise $f_Y(y) = 0$. It is clear then that $Y \sim \text{Beta}(\beta, \alpha)$. If $\alpha = \beta = 1$, then $X$ and $Y$ share the exact same distribution, which is $\text{Uniform}(0,1)$. However, since $Y$ is a deterministic function of $X$ (in particular, $Y = 1 - X$), it is obviously false that $X$ and $Y$ are independent random variables.

9. (5 pts) Let $X \sim F_X$. What is $E(X|X)$?

**Solution:**
For $x \in \text{supp}(X)$, we have $E(X|X = x) = E(x|X = x) = E(x) = x$, i.e. $E(X|X) = X$. 52
10. (10 pts) Let $X, Y, Z$ be three random variables such that $E(X|Z) = 3Z$, $E(Y|Z) = 1$, and $E(XY|Z) = 4Z$. Assume furthermore that $E(Z) = 0$. Are $X$ and $Y$ uncorrelated?

**Solution:**
We have

$$
Cov(X, Y) = E[Cov(X, Y|Z)] + Cov[E(X|Z), E(Y|Z)]
$$

$$
= E[E(XY|Z) - E(X|Z)E(Y|Z)] + Cov(3Z, 1)
$$

$$
= E(4Z - 3Z * 1) + 3Cov(Z, 1) = E(Z) + 3 * 0 = 0 + 0 = 0.
$$

It follows that $X$ and $Y$ are uncorrelated.

11. **Extra credit** (5 pts) Let $Y$ be a continuous random variable with distribution Uniform($a, b$), for some $a < b$. Find the cdf of $Y$.

**Solutions:** Since $Y \sim \text{uniform}(a, b)$, cdf of $Y$ is:

$$
F_Y(x) = \begin{cases} 
0 & \text{if } x < a \\
\frac{x-a}{b-a} & \text{if } a \leq x < b \\
1 & \text{if } x \geq b
\end{cases}
$$

12. **Extra credit** (10 pts) Suppose we want to simulate from a continuous random variable $X$ having a strictly increasing c.d.f. $F_X$. However, the computer at our disposal can only simulate from a Uniform(0, 1) distribution.

(a) Show that if $U \sim \text{Uniform}(0, 1)$, its c.d.f. is

$$
F_U(u) = \begin{cases} 
0 & \text{if } u < 0 \\
u & \text{if } u \in [0, 1] \\
1 & \text{if } u > 1
\end{cases}
$$

(b) Consider the random variable $Y = F_X^{-1}(U)$, where $U \sim \text{Uniform}(0, 1)$. Show that $Y$ has the same distribution as $X$ by proving that

$$
P(Y \leq x) = F_X(x), \quad \text{for all } x \in \mathbb{R}.
$$

*Hint: use part (a) and the monotonicity of $F_X$.*

(c) Using the same computer, how would you go about generating an Exp($\lambda$) random variable (you can use the following fact, without proving it: if $X \sim \text{Exp}(\lambda)$, then $F_X(x) = 1 - e^{-x/\lambda}$)?

53
Solutions:

(a) Since, $U \sim Uniform(0, 1)$,

$$f_U(u) = \begin{cases} 
1 & \text{if } 0 \leq u \leq 1; \\
0 & \text{otherwise}.
\end{cases}$$

Therefore, integrating $f_U(u)$ we have,

$$F_U(u) = \begin{cases} 
0 & \text{if } u < 0; \\
u & \text{if } 0 \leq u \leq 1; \\
1 & \text{if } u > 1.
\end{cases}$$

(b) $P(Y \leq x) = P(F_X^{-1}(U) \leq x) = P(U \leq F_X(x)) = F_X(x)$ using the fact that $F_X$ is strictly increasing (monotonicity) and $U$ has Uniform(0,1) distribution.

(c) The exponential CDF has the form $F(x) = 1 - e^{-x/\lambda}$ for $x \geq 0$. Thus, to generate values of $X$, we should generate values $u \in (0, 1)$ of a uniformly distributed random variable $U$, and set $X$ to the value for which $1 - e^{-x/\lambda} = u \Rightarrow x = -\lambda\ln(1 - u)$.

13. (10 points, 2 pts each) We will study estimation for a normally distributed sample.

(a) Obtain the maximum likelihood estimator for the mean in a Normal model (assume $\sigma^2 > 0$ is known):

$$X_1, \ldots, X_n \sim N(\mu, \sigma^2)$$

(Hint: write down the likelihood function $L(\mu, \sigma^2 | y) = \prod_i f(y_i | \mu, \sigma^2)$.)

(b) Obtain the maximum likelihood estimator for the noise parameter in the Normal model. (Hint: same hint)

(c) Is the maximum likelihood estimate of $\mu$ unbiased?

(d) Is the maximum likelihood estimate of $\sigma^2$ unbiased?

(e) What is an unbiased estimate of $\sigma^2$? (Two hints: (1) consider multiplying a constant, (2) $n$ is a constant i.e. there is nothing random about the sample size.)

Solution:
(a) In finding the estimators, the first thing we’ll do is write the probability density function as a function of $\theta_1 = \mu$ and $\theta_2 = \sigma^2$.

$$f(x_i; \theta_1, \theta_2) = \frac{1}{\sqrt{\theta_2 \sqrt{2\pi}}} \exp \left[ -\frac{(x_i - \theta_1)^2}{2\theta_2} \right]$$

for $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$.

Now, that makes the likelihood function:

$$L(\theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2) = \theta_2^{-n/2}(2\pi)^{-n/2} \exp \left[ -\frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2 \right]$$

and therefore the log of the likelihood function:

$$\log L(\theta_1, \theta_2) = -\frac{n}{2} \log \theta_2 - \frac{n}{2} \log(2\pi) - \frac{1}{2\theta_2} \sum_{i=1}^{n} (x_i - \theta_1)^2$$

(b) Taking the partial derivative of the log likelihood with respect to $\theta_1$, and setting to 0,

$$\sum x_i - n\theta_1 = 0$$

Thus $\hat{\theta}_1 = \hat{\mu} = \frac{\sum x_i}{n} = \bar{x}$.

Taking the partial derivative of the log likelihood with respect to $\theta_2$, and setting to 0,

$$-n\theta_2 + \sum (x_i - \theta_1)^2 = 0$$

Thus $\hat{\theta}_2 = \hat{\sigma^2} = \frac{\sum (x_i - \bar{x})^2}{n}$

(c) $E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^{n} X_i\right) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{n} (n\mu) = \mu$.

Unbiased.
Reorganizing the expression, $\hat{\sigma}^2 = \left( \frac{1}{n} \sum_{i=1}^{n} X_i^2 \right) - \bar{X}^2$. Therefore,

$$E(\hat{\sigma}^2) = E \left[ \frac{1}{n} \sum_{i=1}^{n} X_i^2 - \bar{X}^2 \right] = \frac{1}{n} \sum_{i=1}^{n} E(X_i^2) - E(\bar{X}^2)$$  \hspace{1cm} (4)

$$= \frac{1}{n} \sum_{i=1}^{n} (\sigma^2 + \mu^2) - \left( \frac{\sigma^2}{n} + \mu^2 \right)$$  \hspace{1cm} (5)

$$= \frac{1}{n} \left( n\sigma^2 + n\mu^2 \right) - \frac{\sigma^2}{n} - \mu^2$$  \hspace{1cm} (6)

$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{na^2 - \sigma^2}{n} = \frac{(n-1)\sigma^2}{n}$$  \hspace{1cm} (7)

$$= \sigma^2$$  \hspace{1cm} (8)

Biased.

(e) Notice that we are only $\frac{n}{n-1}$ off here. Thus

$$\hat{\sigma}_{ubiased}^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

14. (10 points, 2.5 pts each) We will now study a Chi-square random variable and the sample standard deviation $S^2$.

(a) Prove that the random variable

$$U = (n - 1)S^2 / \sigma^2$$

for $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$, is distributed as a Chi square random variable (i.e. $Z \sim \chi^2(n - 1)$).

(b) What is expectation $E(U)$?

(c) What is the expectation of $E(S^2)$?

(d) What is the bias of $E(S^2)$? Is it unbiased?

**Solution:**

(a) Recall that if $X_i$ is a normally distributed random variable with mean $\mu$ and variance $\sigma^2$, then:

$$\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$
(b) Recall that the expected value of a chi-square random variable is its degrees of freedom. That is, if \( X \sim \chi^2(r) \), then \( E(X) = r \). Therefore,

\[
E\left(\frac{(n-1)S^2}{\sigma^2}\right) = n - 1
\]

(c) \( E(S^2) = E\left[\frac{\sigma^2}{n-1} \cdot \frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n-1} E\left[\frac{(n-1)S^2}{\sigma^2}\right] = \frac{\sigma^2}{n-1} \cdot (n-1) = \sigma^2 \)

(d) 0. Unbiased.

**Note:** There is no extra credit for the following problem, because it is a duplicate with 13. The problem is reproduced here for reference:

15. **Extra Credit** (10 pts) Let’s say it is reasonable that an \( n \)-sized random sample of body weights in the age range of 17-25:

\[
X_1, \ldots, X_n
\]

is i.i.d. distributed as a Normal distribution with mean \( \mu \) and variance \( \sigma^2 \). A public health official would like to know what the ‘spread’ (variance) of the body weights of this population – we already learned from class that a maximum likelihood estimator of the two parameters, \((\hat{\mu}_{\text{MLE}}, \hat{\sigma}^2_{\text{MLE}})\) is

\[
\left( \frac{\sum_{i=1}^{n} X_i}{n}, \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n} \right)
\]

(a) Is the above estimator for \( \sigma^2 \) biased? If so, what is the bias? (7 pts)

(b) Also, can you guess what an unbiased estimator is? (3 pts)
Homework 7 (due June 14th)

(For this section of the class, I highly recommend trying some problems in the WMS book, for the relevant sections as written in the lecture notes.)

1. (20 pts, 5 points each) Maximum likelihood estimation
   
   (a) What is the MLE of the variance of $X_i$ in a setting where we have $n$ i.i.d. observations $X_1, \cdots, X_n \sim \text{Bernoulli}(\rho)$?
   
   (b) What is the MLE of the variance of $X$ in a setting where we have one binomial observation?
   
   (c) What is the MLE of $\lambda$ in a setting where we have samples $X_1, \cdots, X_n \sim \text{exp}(\theta)$, where the exponential distribution is defined differently as
   $$f_{X_1}(x) = \theta e^{-\theta x}.$$ 
   Is it easy to see that this estimator is unbiased? Explain.
   
   (d) The MLE has the remarkable property that it converges in distribution to a normal random variable (under some mild conditions that we will not worry about) in the following way:
   $$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta) \xrightarrow{d} N(0, \frac{1}{I(\theta)})$$
   for the Fisher’s information number:
   $$I(\theta) = V \left[ \frac{d}{d\theta} \log f(X; \theta) \right] = -E \left[ \frac{d^2}{d\theta^2} \log f(X; \theta) \right]$$
   Continuing with the setting of (c) and using this result about the asymptotic normality of the maximum likelihood estimator, what is a good approximate distribution of $\hat{\theta}_{\text{MLE}}$, when $n$ is large?

Solution:

- Book Example 9.17. It is not easy, because $E(\frac{1}{X})$ is not something that is easily calculable; we do not have a good immediate method of calculating $E(g(Y))$ for some function $g$.
- The answer is $X$. This is because if you take the derivative of the joint pmf of the sample, which is just the pmf of a single sample,
   $$\log p_X(x|p) = x \log p + (1 - x) \log(1 - p)$$
   $$\frac{d}{dp} p_x(x|p) = \frac{1}{p} - \frac{(1 - x)}{(1 - p)}$$
Set the last expression equal to zero, to get that $\hat{p}_{\text{MLE}} = X$.

- Find the MLE that minimizes

$$\log L(\theta | \mathbf{X}) = f_{x_1, \ldots, x_n}(x_1, \ldots, x_n | \theta) = e^{-\theta \sum_i X_i + n \log \theta}$$

then calculate

$$I(\theta) = E[-\frac{d^2}{d\theta^2} \log L(\theta | \mathbf{X})] = \frac{n}{\theta^2}$$

. The answer follows that $\hat{\theta} \overset{d}{\rightarrow} N(\theta, \frac{\theta^2}{n})$

- $\hat{\theta}_n \overset{d}{\rightarrow} N(0, \frac{1}{I(\theta)})$

2. (30 pts) Let $X_1, \ldots, X_{10} \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. We don’t know the value of $\sigma^2$, but we want to decide whether, based on the data $X_1, \ldots, X_{10}$, $\sigma^2 \leq \sigma_0^2 = 3$ or not. We observe from the data that $S^2 = s^2 = 10$.

- Use the sampling distribution of $S^2$ (read lecture 13) to compute the probability of the event $\{S^2 \geq s^2\}$ under the assumption that the true unknown variance is $\sigma^2 = \sigma_0^2$ (you will need to use a statistical software and use the c.d.f. of a $\chi^2$ with a certain number of degrees of freedom to compute this probability).

- Based on the probability that you computed, explain why the data $X_1, \ldots, X_{10}$ suggest that it is likely not true that $\sigma^2 \leq \sigma_0^2$.

Solution:

- We know that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1).$$

Here, $\sigma^2$ is unknown. However, let’s assume that $\sigma^2$ is at least as large as $\sigma_0^2$ and let’s therefore plug $\sigma^2 = \sigma_0^2$ in the above result

\[1\] We can safely plug-in $\sigma_0^2$ because

$$\sup_{0 < \sigma^2 \leq \sigma_0^2} P_{\sigma^2}(S^2 > s^2) = P_{\sigma_0^2}(S^2 > s^2).$$

Here the notation $P_{\sigma_0^2}(A)$ should be interpreted as the ‘probability of the event $A$ when we assume that the true unknown variance is $\sigma_0^2$’. 
about the sampling distribution of $S^2$. It then follows that,

$$P(S^2 \geq s^2) = P\left(\frac{(n-1)S^2}{\sigma_0^2} \geq \frac{(n-1)s^2}{\sigma_0^2}\right)$$

$$= P\left(\chi^2(9) \geq 30\right) \approx 0.0004$$

(with a little abuse of notation).

- We thus discovered that, if the true unknown variance was not larger than $\sigma_0^2 = 3$ as we conjectured, then the probability of observing a value of the sample variance $s^2$ as extreme as the one we observed from the data ($s^2 = 10$) is extremely small. Based on the probability that we computed, it looks therefore very unlikely that the true unknown variance $\sigma^2$ is not larger than $\sigma_0^2 = 3$.

3. (20 pts) The times a cashier spends processing individual customer’s orders are independent random variables with mean 2.5 minutes and standard deviation 2 minutes. What is the approximate probability that it will take more than 4 hours to process the orders of 100 people? You can leave your answer in terms of $\Phi$.

Solution:

Let $Z \sim \mathcal{N}(0, 1)$ and let $X_1, \ldots, X_n, \ldots$ denote the times spent by the cashier to process the 100 individual customer’s orders. Without any other assumption on the distribution of the $X$’s other than independence, finite mean and finite variance, we can only rely on the Central Limit Theorem. In this case, $n = 100$ is likely large enough to give a good approximation to the event of interest, which is

$$\sum_{i=1}^{n} X_i > 240$$
(notice that there are exactly 240 minutes in 4 hours). Thus, we have

\[
P \left( \sum_{i=1}^{n} X_i > 240 \right) = P \left( \bar{X}_n > \frac{240}{n} \right)
\]

\[
= P \left( \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} > \frac{\sqrt{n} \left( \frac{240}{n} - \mu \right)}{\sigma} \right)
\]

\[
\approx P \left( Z > \frac{\sqrt{n} \left( \frac{240}{n} - \mu \right)}{\sigma} \right) = 1 - P \left( Z \leq \frac{\sqrt{n} \left( \frac{240}{n} - \mu \right)}{\sigma} \right)
\]

\[
= 1 - \Phi \left( \frac{\sqrt{n} \left( \frac{240}{n} - \mu \right)}{\sigma} \right) = 1 - \Phi \left( \frac{10 \left( \frac{240}{100} - 2.5 \right)}{2} \right) = 1 - \Phi \left( \frac{1}{2} \right) = \Phi \left( \frac{1}{2} \right).
\]

4. (15 pts) Let \( X_1, \ldots, X_n \) be iid random variables with c.d.f. \( F \), finite expectation \( \mu \) and finite variance \( \sigma^2 \). Consider the random variable \( Y = a\bar{X}_n + b \) for \( a, b \in \mathbb{R} \) and \( a > 0 \). Assume that \( n \) is ‘large’. Give an approximate expression for the \( \alpha \)-quantile of \( Y \). Your answer will be in terms of \( \Phi^{-1} \).

Solution:

Let \( Z \sim \mathcal{N}(0,1) \). Without further information about \( F \), we can only use the Central Limit Theorem. By definition, the \( \alpha \)-quantile of \( Y \) is the value \( q \in \mathbb{R} \) such that \( P(Y \leq q) = \alpha \). We can use the Central Limit Theorem to approximate this number when \( n \) is large.

\[
\alpha = P(Y \leq q) = P(a\bar{X}_n + b \leq q) = P \left( \bar{X}_n \leq \frac{q - b}{a} \right)
\]

\[
= P \left( \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \leq \frac{\sqrt{n} \left( \frac{q - b}{a} - \mu \right)}{\sigma} \right)
\]

\[
\approx P \left( Z \leq \frac{\sqrt{n} \left( \frac{q - b}{a} - \mu \right)}{\sigma} \right) = \Phi \left( \frac{\sqrt{n} \left( \frac{q - b}{a} - \mu \right)}{\sigma} \right).
\]

Thus,

\[
\sqrt{n} \frac{\left( \frac{q - b}{a} - \mu \right)}{\sigma} \approx \Phi^{-1}(\alpha),
\]
i.e.
\[ q \approx b + a \left( \mu + \frac{\sigma \Phi^{-1}(\alpha)}{\sqrt{n}} \right). \]

5. (15 pts) Let \( X_1, \ldots, X_n \) iid \( \sim \) Bernoulli\((p)\) be a sequence of binary random variables where
\[
X_i = \begin{cases} 
1, & \text{if the Pittsburgh Pirates won the } i\text{-th game against the Cincinnati Reds} \\
0, & \text{if the Pittsburgh Pirates lost the } i\text{-th game against the Cincinnati Reds}
\end{cases}
\]
and \( p \in (0, 1) \). Suppose that we are interested in the odds that the Pittsburgh Pirates win a game against the Cincinnati Reds. To do that, we need to consider the function
\[ g(p) = \frac{p}{1 - p} \]
representing the odds of the Pirates winning against the Reds. It is natural to estimate \( p \) by means of \( \bar{X}_n \). Provide a distribution that can be used to approximate probability statements about \( g(\bar{X}_n) \).

**Solution:**
We have \( E(X_i) = p \in (0, 1) \) and \( V(X_i) = p(1 - p) \in (0, 1) \) for any \( i \in \{1, \ldots, n\} \) so both the expectation and the variance of the \( X \)'s exist and are finite. Consider the function \( g(x) = \frac{x}{1 - x} \) defined for \( x \in (0, 1) \). We have
\[ g'(x) = \frac{1}{(1 - x)^2}. \]
Notice that \( g'(x) > 0 \) for any \( x \) in \( (0, 1) \). Thus, \( g'(E(X_i)) \neq 0 \). The Delta Method allows us to write
\[
\frac{\sqrt{n}[g(\bar{X}_n) - g(E(X_i))] - \sqrt{n} \left( \frac{\bar{X}_n}{1 - \bar{X}_n} - \frac{p}{1 - p} \right)}{|g'(E(X_i))| \sqrt{V(X_i)}} \overset{d}{\to} \mathcal{N}(0, 1).
\]
Hence, we can use the distribution
\[
\mathcal{N} \left( \frac{p}{1 - p}, \frac{p}{n(1 - p)^2} \right)
\]
to approximate probability statements about \( g(\bar{X}_n) = \frac{\bar{X}_n}{1 - \bar{X}_n} \).
6. **Extra credit** (10 pts) Let $U_1, U_2, \ldots, U_n, \ldots$ be a sequence of random variables such that $U_n \xrightarrow{d} \text{Uniform}(0,1)$. Consider the function $g(x) = -\log x$ and the new sequence of random variables $X_1, X_2, \ldots, X_n, \ldots$ with $X_i = -\log U_i$. Prove that $X_n \xrightarrow{d} \text{Exponential}(1)$. Hint: use the method of the c.d.f. from lecture.

**Solution:**

We first want to compute the c.d.f. of the $n$-th transformed variable $X_n$ in the sequence. We have

$$F_{X_n}(x) = P(X_n \leq x) = P(-\log U_n \leq x) = P(U_n \geq e^{-x})$$

Because $U_n \xrightarrow{d} \text{Uniform}(0,1)$, it follows that

$$P(U_n \geq e^{-x}) \rightarrow \begin{cases} 0 & \text{if } x < 0 \\ 1 - e^{-x} & \text{if } x \geq 0 \end{cases} = F(x).$$

We recognize $F$ this as the c.d.f. of an Exponential(1) distribution. Thus, we conclude that $X_n \xrightarrow{d} \text{Exponential}(1)$. 
Homework 8 (due June 16th)

1. (10 pts) Prove the weak law of large numbers under the additional assumption that the variance $\sigma^2$ of the iid sequence $X_1, X_2, \ldots, X_n, \ldots$ exists and is finite. Hint: use Tchebysheff’s inequality. (Go back to earlier lecture notes and homework for how tchebysheff was used)

**Solutions:**
take any $\epsilon > 0$. by tchebysheff’s inequality we have that

$$p(|\bar{x}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

clearly,

$$0 \leq \lim_{n \to \infty} p(|\bar{x}_n - \mu| > \epsilon) \leq \lim_{n \to \infty} \frac{\sigma^2}{n\epsilon^2} = 0,$$

hence the result.

2. (10 pts) consider the sequence of random variables $x_1, x_2, \ldots, x_n, \ldots$ with $x_n \sim \mathcal{N}(\mu, 3/(1+\log n))$. show that $x_n \xrightarrow{p} \mu$. (hint: use tchebysheff again)

**Solutions:**
take any arbitrary $\epsilon > 0$. tchebysheff’s inequality implies

$$p(|x_n - \mu| > \epsilon) \leq \frac{v(x_n)}{\epsilon^2} = \frac{3}{(1+\log n)\epsilon^2}.$$ 

clearly,

$$0 \leq \lim_{n \to \infty} p(|x_n - \mu| > \epsilon) \leq \lim_{n \to \infty} \frac{3}{(1+\log n)\epsilon^2} = 0,$$

and thus $x_n \xrightarrow{p} \mu$.

3. (15 pts, 5 pts each)

(a) What is the mgf of $X_i \sim \text{Bernoulli}(p)$?

(b) What is the mgf of $X = \sum X_i$?

(c) What is the distribution of $X$? (We knew this, but now we know a convenient way of proving it!)

**Solutions:**
(a) The moment generating function of $X_i$ is

$$m_{X_i}(t) = E(e^{tX_i}) = \sum_{x=0,1} e^{tx}p^x(1-p)^{1-x} = 1 - p + e^tp$$

(b) The mgf of the sum of $X_i$ is the multiple of the mgfs!

$$m_X(t) = \prod_{i=1}^{n} m_{X_i}(t) = (1 - p + e^tp)^n$$

(c) We will actually approach this in the more general case, where for $X_i$ which follow a Binomial $(n_i, p)$ distribution, the sum $X = \sum_i X_i$ follows a Binomial $(\sum_i n_i, p)$ distribution.

(d) By the convolution formula, $Z = X_1 + X_2$ has probability distribution:

$$f_Z(z) = \sum_{x=0}^{z} \binom{n_1}{x} p^x(1-p)^{n_1-x} \binom{n}{z-x} p^{z-x}(1-p)^{n_2-(z-x)}$$

$$= p^z(1-p)^{n_1+n_2-z} \sum_{x=0}^{z} \binom{n_1}{x} \binom{n_2}{z-x}$$

$$= \binom{n_1+n_2}{z} p^z(1-p)^{n_1+n_2-z}$$

which is the pmf of a Binomial distribution with parameters $n_1 + n_2$ and $p$. The general case for $k > 1$ variables follows by induction.

4. (10 pts) Sum of Poissons:

(a) Prove the sum of Poisson distributed $X,Y$ each with parameters $\lambda_1, \lambda_2$ is a Poisson random variable. (What is the parameter?) (Hint: the following formula – sometimes called the discrete convolution formula for $Z = X_1 + X_2$

$$P(X_1 + X_2 = z) = f_Z(z) = \sum_{x=0}^{z} f_{X_1}(x) f_{X_2}(z-x)$$

may come in handy)

(b) Prove the above with moment generating functions.
Solutions:

(a) By the convolution formula, $Z = X + Y$ has probability distribution:

$$f_Z(z) = \sum_{x=0}^{\infty} 1 \cdot \frac{1}{x!} e^{-\lambda_1} \frac{1}{(z-x)!} e^{-\lambda_2}$$

$$= e^{-(\lambda_1+\lambda_2)} \sum_{x=0}^{\infty} \frac{\lambda_1^x \lambda_2^{z-x}}{x! (z-x)!}$$

Now, notice that you can use the so-called binomial formula with $N = z, a = \lambda_1, b = \lambda_2$:

$$(a + b)^N = \sum_{x=0}^{N} \binom{N}{x} a^x b^{N-x}$$

to get that (9) is equal to

$$e^{-(\lambda_1+\lambda_2)} \sum_{x=0}^{\infty} \frac{(\lambda_1 + \lambda_2)^z}{z!}$$

You can recognize a new Poisson pmf with parameter $\lambda_1 + \lambda_2$ – this completes the proof. (You can also see this generalizes easily to the sum of many Poissons)

(b) Each of $X$ and $Y$ have the moment generating function

$$m_X(t) = e^{\lambda_1(e^t-1)}, \quad m_Y(t) = e^{\lambda_2(e^t-1)}$$

and the mgf of $X + Y$ is

$$m_{X+Y}(t) = m_X(t)m_Y(t) = e^{\lambda_1+\lambda_2(e^t-1)}$$

which is easily seen as the mgf of a Poisson random variable with parameter $\lambda_1 + \lambda_2$. (This is much easier than (a), wouldn’t you agree?)

5. (10 pts) Assume that $X_1, \cdots, X_n$ have finite moment generating functions; i.e. $m_{X_i}(t) < \infty$ for all $t$, for all $i = 1, \cdots, n$. Prove the central limit theorem again, using moment generating functions. (Hint: come to office hours for a hint.)
Solutions:
Let \( T_n = \frac{S_n - n\mu}{\sigma \sqrt{n}} \), for the sum \( S_n = \sum_i X_i \). Also denote the mean and variance of \( X \) as \( E(X_i) = \mu \) and \( V(X_i) = \sigma^2 \). Then, we want to prove that
\[
P(T_n \leq x) \xrightarrow{n \to \infty} P(Z \leq x)
\]
Denote as the ‘standardized’ version of \( X_i \) as \( Y_i = \frac{X_i - \mu}{\sigma} \) - we see that \( Y_i \) are still i.i.d. but with mean \( E(Y_i) = 0 \) and \( V(Y_i) = 1 \), and
\[
T_n = \frac{\sum_{i=1}^{n} Y_i}{\sqrt{n}}
\]
Now, the central limit theorem is to show that the moment generating function of \( T_n \) goes to that of a standard normal random variable \( Z \sim \mathcal{N}(0, 1) \), whose mgf is \( \exp\left(\frac{t^2}{2}\right) \).

\[
m_{T_n}(t) = E(e^{tT_n})
= \left(E(e^{\frac{t}{\sqrt{n}}Y_i})\right)^n
= \left(1 + \frac{t}{\sqrt{n}} EY + \frac{1}{2} \frac{t^2}{n} E(Y^2) + \frac{1}{6} \frac{t^3}{n^{3/2}} E(Y^3) + \cdots\right)^n
\]
\[
\approx \left(1 + \frac{t^2}{2n}\right)^2
\rightarrow e^{\frac{t^2}{2}}
\]

6. (10 pts) Using the central limit theorem for Poisson random variables, compute the value of
\[
\lim_{n \to \infty} e^{-n} \sum_{i=0}^{n} \frac{n^i}{i!}
\]
**Solutions:** If you denote \( X_1, \cdots, X_n \) each i.i.d. Poisson random variables with parameter 1. Then, \( S_n \) is Poisson(\( n \)), and the expectation of \( S_n \) is \( E(S_n) = n \). The central limit theorem states that the sum of Poissons should go to \( P(S_n \leq n) \rightarrow \frac{1}{2} \). Lastly, notice that the quantity of interest is exactly \( P(S_n \leq n)! \)

7. (15 pts) Roll a fair die \( n \) times, call the outcomes \( X_i \). Then, estimate the probability that \( S_n = \sum_i X_i \) exceeds its expectation by at least \( n \), for \( n = 100 \) and \( n = 1000 \).
• (5 pts) Use Tchebysheff’s inequality.
• (5 pts) Obtain the mgf of $S_n$.
• (5 pts) Use large deviation bounds. (Hint: you will need to use a computer program of your choice, to get a component in the large deviation bound.)

Solutions:

• The Tchebysheff’s upper bound is $\frac{35}{12n}$. This can be found by
  
  $P(S_n - E(S_n) \geq k \cdot \sqrt{V(S_n)}) \leq \frac{1}{k^2}$

  for the appropriate $k$ that matches our problem. Indeed, $k \cdot \sqrt{V(S_n)} = k \cdot \sqrt{\frac{35}{12} \cdot n}$ which gives $k$ such that $\frac{1}{k^2} = \frac{35}{12n}$.

• The mgf is
  
  $m_{X_i}(t) = \frac{1}{6} \sum_{i=1}^{6} e^{it} = \frac{e^{it}(e^{6t} - 1)}{6(e^t - 1)}$.

• Observe the $E(X_i) = 3.5$ and $E(S_n) = 3.5n$. We want to find an upper bound on $P(S_n - 3.5n \geq n) = P(S_n \geq 4.5n)$. We know that the mgf of $X_i$ is
  
  $m_{X_i}(t) = \frac{1}{6} \sum_{i=1}^{6} e^{it} = \frac{e^{it}(e^{6t} - 1)}{6(e^t - 1)}$.

  from part (b). Then, we need to compute the maximum of
  
  $I(t) = \sup_{t \geq 0} 4.5t - \log(m_{X_i}(t))$

  which can be solved using a numerical solver of your choice. The (unique) maximizer can be verified to be $t \simeq 0.37105$, so that $I(4.5) \simeq 0.178$. The resulting large deviation bound is:
  
  $P(S_n \geq 4.5n) \leq e^{-0.178n}$

  for, say, $n = 100$, this is $1.08 \cdot 10^{-8}$, while the Tchebysheff’s bound is $0.03$ – in short, the large deviation is shown to be a much tighter upper bound on the quantity of interest.
8. (10 pts) In each game a gambler wins the dollars he bets with probability \( p \) and looses with probability \( 1 - p \). If he has less than $3, he will bet all that he has. Otherwise, since his goal is to have $5, he will only bet the difference between $5 and what he has. He continues to bet until wither he has either $0 or $5. Let \( X_n \) denote the amount he has right after the \( n \)-th bet.

(a) Draw the transition graph.
(b) Classify the states. (We haven’t covered this material yet! Full points given)

**Solutions:**

9. The possible states are \( \{0, 1, 2, 3, 4, 5\} \). The probability transition matrix \( P \) is

\[
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
1 - p & 0 & p & 0 & 0 & 0 \\
1 - p & 0 & 0 & 0 & p & 0 \\
0 & 1 - p & 0 & 0 & 0 & p \\
0 & 0 & 0 & 1 - p & 0 & p \\
0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

and the transition probability matrix is

10. The states 0-5 are recurrent, and all others are transient.

11. (10 pts) The number of failures \( X_t \), which occur in a computer network over the time interval \( [0, t] \), can be described by a homogeneous Poisson process \( \{X_t, t \geq 0\} \). On average, there is a failure after every 4 hours, i.e. the intensity of the process is equal to \( \lambda = 0.25 \).

(a) What is the probability of at most 1 failure in \( [0, 8) \), at least 2 failures in \( [8, 16) \), and at most 1 failure in \( [16, 24) \) (time unit: hour)?
(b) What is the probability that the third failure occurs after 8 hours?

**Solutions:**

(a) 

\[
P[T_9 - T_0 \leq 1 \& T_{16} - T_8 \geq 2 \& T_{24} - T_{16} \leq 1] = P[T_9 - T_0 \leq 1 \& T_{16} - T_8 \geq 2 \& T_{24} - T_{16} \leq 1]
\]

\[
= P[T_8 \leq 1]P[T_8 \geq 2]P[T_8 \leq 1] \cdots \text{By independence}
\]

69
Now, notice
\[ P(T_8 \leq 1) = P(T_8 = 0) + P(T_8 = 1) = e^{-25.8} + 0.25 \cdot 8^{-0.25} = 0.406 \]
and
\[ P(T_8 \geq 2) = 1 - P(T_8 \leq 1) = 0.594 \]
So, the answer is
\[ 0.406 \cdot 0.594 \cdot 0.406. \]

12. **Extra credit** (10 pts) Prove the large deviation bound. (Hint: start with the Markov inequality)

**Solutions:**
Denote \( S_n = X_1 + \cdots + X_n \). Then, for any \( t > 0 \), by Markov’s Inequality:
\[ P(S_n \geq a \cdot n) = P(e^{tS_n-tan} \geq 1) \leq E(e^{tS_n-tan} = e^{tan} m_{X_i}(t)^n = \exp(-n(at-\log m_{X_i}(t))) \]
Since this holds for any \( t > 0 \), we will simply minimize this over \( t > 0 \) – this is accomplished by the sup(\cdot) operation – this gives the desired expression.
Homework 9 (due June 23rd)

1. (15 pts) Consider a Bernoulli process $X$ and its inter-arrival times $\{T_i\}_{i=1}^{\infty}$. Show that $T_r = \sum_{i=1}^{r} T_i$ with $r \in \{1, 2, \ldots \}$ is a Negative Binomial random variable with parameters $r$ and $p$, where $p$ is the probability of success of each of the random variables $\{X_i\}_{i=1}^{\infty}$.

**Solutions:** The inter-arrival times of a Bernoulli process are iid Geometric($p$) random variables. The m.g.f. of a Geometric distribution of parameter $p$ is

$$m(t) = \frac{pe^t}{1 - (1-p)e^t}$$

for $t < -\log(1-p)$. Consider the random variable $T_r = \sum_{i=1}^{r} T_i$. We have

$$m_{T_r}(t) = m^r(t) = \left[ \frac{pe^t}{1 - (1-p)e^t} \right]^r$$

for $t < -\log(1-p)$, which is the m.g.f. of the Negative Binomial distribution with parameters $r$ and $p$. Thus, $T_r \sim \text{NBinomial}(r, p)$.

2. (15 pts) A rat is wondering inside a maze, which has rooms labeled as F, 2, 3, 4, 5 and S. If a room has $k$ doors, then the probability that the rat chooses any particular door is $1/k$. The rat always moves to a different room at each time step. If the rat reaches room F, which contains food, or room S, which gives it a harmless electrical shock, the rat remains there and the experiment terminates.

(a) Compute by hand the probability that the rat gets to room S in 4 steps given that it starts in room 3.

(b) Compute the n-step ahead transition probability matrix for $n = 2, n = 4, n = 10, n = 50$. Hint: you definitely want to use a computer for this.

(c) Identify the classes of transient and recurrent states. (i.e. classify the states as such)

(d) Does this chain admit a stationary distribution? Explain.

(e) Now suppose that the rules of the experiment change: the experiment terminates only when the rat reaches room S for the second time (in particular, the rat can visit room F multiple times). Derive the transition probability matrix for this process.

**Solutions**
(a) (3 points) Firstly, for the states: \{2, 3, 4, 5, F, S\} we obtain the
matrix of transition probabilities as
\[
P = \begin{pmatrix}
0 & 0 & 1/2 & 0 & 1/2 & 0 \\
0 & 0 & 1/3 & 1/3 & 1/3 & 0 \\
1/3 & 1/3 & 0 & 0 & 0 & 1/3 \\
0 & 1/2 & 0 & 0 & 0 & 1/2 \\
0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
\end{pmatrix}
\]
Now, there are 7 possible paths, indicated below along with their probabilities

\[
\begin{array}{c|c}
\text{Path} & \text{Probability} \\
3 - 5 - 3 - 4 - S & 1/54 \\
3 - 5 - 3 - 5 - S & 1/36 \\
3 - 5 - S - S - S & 1/6 \\
3 - 4 - 2 - 4 - S & 1/54 \\
3 - 4 - 3 - 4 - S & 1/81 \\
3 - 4 - 3 - 5 - S & 1/54 \\
3 - 4 - S - S - S & 1/9. \\
\end{array}
\]

The probabilities are obtained using the transition probability matrix. For example, the path 3 - 5 - S - S - S has probability
\[
(1/3)(1/2)(1)(1) = 1/6
\]
and the path 3 - 4 - 3 - 4 - S has probability
\[
\]
So the sum is \(\frac{121}{324}\).

(b) (3 points) Using R: Please check Fig. 3.

(c) (3 points) The states F and S are recurrent. The other states are transient.

(d) (3 points) This chain does not admit a stationary distribution since it has two recurrent states.

(e) (3 points) Consider the following state space
\[
\{(F, 0), (2, 0), (3, 0), (4, 0), (5, 0), (S, 1), (F, 1), (2, 1), (3, 1), (4, 1), (5, 1), (S, 2)\}
\]

72
where each state consist of a pair, the first element denoting the
room and the second the number of visits to room $S$.
Then, the sequence of rooms visited by the rat under the new
rule forms a first order Markov chain having this enlarged state
space. The transition probability matrix is Fig. 4. Notice that
the state $(S, 2)$ is the only recurrent state (i.e. it is an absorbing
state).

3. (15 pts) Suppose a student in 36-217 can either be up-to-date (U)
with them material covered in class or behind (B). The probability of
a student being up-to-date or behind on a particular week depends
on whether he/she has been behind or up-do-date in the previous two
weeks. In particular

- If behind both this week and last week, the student will be behind
  next week as well with probability 0.8
- If up-to-date both this week and last week, the student will be
  up-to-date next week as well with probability 0.9.
- If behind last week and up-to-date this week, the student will be
  behind with probability 0.5.
- If up-to-date last week and behind this week, the student will be
  behind with probability 0.7.

(a) Is this a first-order Markov chain? Why?
(b) Explain how you can enlarge the state space and obtain a first-
order Markov chain.
(c) More generally, if you have a k-th order Markov chain on a state
space of cardinality $m$, explain how you can always derive a first-
order Markov chain on a larger state space and find the cardinal-
ity if this enlarged state space.

Solutions:

(a) (2 points) This is not a first-order Markov chain since the prob-
ability of being in a given state depends on the states of the two
previous steps.
(b) (3 points) We can consider the state space $\{(U, U), (U, B), (B, U), (B, B)\}$
which is the set of paths that matter for determining the states
of the next step. We consider the new random variable $Y_t =$
In order to define the probabilities of the new first-order Markov chain we define
\[
P(Y_t = (X_t, X_{t-1}) = (i, j)|Y_{t-1} = (X_{t-1}, X_{t-2}) = (k, l)) = \begin{cases} 
P(X_t = i|X_{t-1} = k, X_{t-2} = l) & \text{if } j = k \\
0 & \text{otherwise}
\end{cases}
\]

Thus, in this case we obtain the probability transition matrix for the states \{\(U, U\), \(U, B\), \(B, U\), \(B, B\)\}

\[
\begin{pmatrix}
0.9 & 0.1 & 0 & 0 \\
0 & 0 & 0.3 & 0.7 \\
0.5 & 0.5 & 0 & 0 \\
0 & 0 & 0.2 & 0.8
\end{pmatrix}
\]

(c) (4 points) In general if we have a state space \(\{E_1, \ldots E_m\}\) we can consider the new state space \(\{E_1, \ldots E_m\}^k\) with cardinality \(m^k\) which contents the set of paths that matter for determining the states of the next step. We consider the new random variable \(Y_t = (X_t, X_{t-1}, \ldots, X_{t-(k-1)})\). In order to define the probabilities of the new first-order Markov chain we define
\[
P(Y_t = (X_t, X_{t-1}, \ldots, X_{t-(k-1)}) = (i_0, i_1, \ldots, i_{k-1})|Y_{t-1} = (X_{t-1}, X_{t-2}, \ldots, X_{t-k}) = (j_1, j_2, \ldots, j_k)) = \begin{cases} 
P(X_t = i_0|X_{t-1} = j_1, X_{t-2} = j_2, \ldots, X_{t-k} = j_k) & \text{if } i_1 = j_1, \ldots, i_{k-1} = j_{k-1} \\
0 & \text{otherwise}
\end{cases}
\]

4. (10 pts) For a Markov chain with state space \(\{a, b, c, d, e, f\}\) and transition probability matrix
\[
P = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 \\
0.3 & 0.1 & 0.1 & 0 & 0.2 & 0.2 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0.3 & 0.1 & 0 & 0.4 & 0.4 \\
0.2 & 0.4 & 0.3 & 0 & 0 & 0.1 \\
0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

Classify the states (recurrent and transient). Does the chain admit a limiting distribution? **Solutions:** The chains contains 3 recurrent classes: \(\{a\}, \{b, g\}\) and \(\{d\}\). The others are transient. This chain does not admit a limiting distribution because there are more than one recurrent class (i.e. the limiting distribution depends on the initial state).
5. (15 pts) Let $X_0, X_1, \ldots$ be a Markov Chain with state space \{1, 2, 3\}, initial distribution $p_{X_0} = (1/5, 2/5)$ and transition probability matrix

$$P = \begin{bmatrix} 1/5 & 4/5 & ? \\ 2/5 & 1/2 & ? \\ 0 & 1/10 & ? \end{bmatrix}$$

Fill in the entries for $P$ and $p_{X_0}$, and answer the following:

(a) Compute $P(X_1 = 1 | X_0 = 2)$.

(b) The row vector $p_{X_0}$ describes the distribution of $X_0$. What is the row vector describing the distribution of $X_1$?

(c) What is $P(X_1 = 3)$?

(d) What is the row vector describing the distribution of $X_2$?

(e) What is $P(X_2 = 1)$?

**Solutions:**

The probabilities along the rows of the transition matrix must sum to 1, as must the initial distribution. Hence we have

$$p_{X_0} = (1/5, 2/5, 2/5),$$

and

$$P = \begin{bmatrix} 1/5 & 4/5 & 0 \\ 2/5 & 1/2 & 1/10 \\ 0 & 1/10 & 9/10 \end{bmatrix}$$

(a) (3 points) $P(X_1 = 1 | X_0 = 2) = P_{21} = 2/5 = .4$

(b) (3 points) The row vector describing $X_1$ is obtained by matrix multiplication:

$$p_{X_1} = p_{X_0} \ast P = (.2, .4, .4) \begin{bmatrix} .2 & .8 & 0 \\ .4 & .5 & .1 \\ 0 & .1 & .9 \end{bmatrix} = (.2, .4, .4)$$

Note that this matrix multiplication implements the law of total probability for each value that $X_1$ could take, namely:

$$P(X_1 = k) = \sum_{i=1}^{3} P(X_1 = k | X_0 = i) P(X_0 = i)$$
(c) (3 points) \( P(X_1 = 3) = p_{X_1}[3] = .4 \) using the distribution calculated in the previous part.

(d) (3 points) To get the row vector describing \( X_2 \) we need to use the 2-step transition matrix \( P_2 \) which is just the square of \( P \):

\[
P_2 = P \ast P = \begin{bmatrix}
.2 & .8 & 0 \\
.4 & .5 & .1 \\
0 & .1 & .9
\end{bmatrix} \begin{bmatrix}
.2 & .8 & 0 \\
.4 & .5 & .1 \\
0 & .1 & .9
\end{bmatrix} = \begin{bmatrix}
.28 & .58 & .14 \\
.36 & .56 & .08 \\
.04 & .14 & .82
\end{bmatrix}
\]

Now we compute the probability distribution for \( X_2 \) using the law of total probability again:

\[
P(X_2 = k) = \sum_{i=1}^{3} P(X_2 = k | X_0 = i)P(X_0 = i).
\]

Or in matrix form:

\[
p_{X_2} = p_{X_0} \ast P_2 = (.2, .4, .4) \begin{bmatrix}
.28 & .58 & .14 \\
.36 & .56 & .08 \\
.04 & .14 & .82
\end{bmatrix} = (.2, .4, .4)
\]

(e) (3 points) \( P(X_2 = 1) = p_{X_2}[1] = .2 \) using the distribution calculated in the previous part.

6. (15 pts) Let \( X_n \) be the maximum reading obtained in the first \( n \) throws of a fair die and \( Y_n \) be the minimum reading obtained in the first \( n \) throws of a fair die.

(a) Explain why \( X_1, X_2, X_3, \cdots \) forms a Markov chain.

(b) Find the transition probability matrix for the chain and draw the transition graph.

(c) Is \( \{X_n\} \) irreducible or reducible? Explain.

(d) Suppose that the fair die is 3- rather than 6-faced. Let \( Z_n \) denote the maximum reading obtained in the first \( n \) throws. Does \( \{Z_n\} \) have a limiting (equilibrium) distribution? If yes, find the limiting distribution. If not, explain why not (do not give a mathematical reason, but rather explain what makes this Markov chain special).

(e) Is \( Y_1, Y_2, Y_3, \cdots \) a Markov chain? Explain.
Solutions:

(a) (3 points) First, \(X_1, X_2, X_3, \ldots\) a sequence of random variables that take values in the same set of \(\{1, 2, 3, 4, 5, 6\}\).

Second, \(X_1, X_2, X_3, \ldots\) has the Markov Property, that is, for \(n = 1, 2, 3, \ldots\)

\[ P\{X_{n+1} = s_{n+1} | X_n = s_n, \ldots, X_1 = s_1\} = \begin{cases} \frac{1}{6} & s_n < s_{n+1} \leq 6 \\ \frac{s_n}{6} & s_{n+1} = s_n \\ 0 & s_{n+1} < s_n \end{cases} \]

That is, given the current state \(s_n\) and past states \(s_{n-1}, \ldots, s_1\), the distribution of the next state \(X_{n+1}\) depends only on \(s_n\), the value of the current state. So \(X_1, X_2, X_3, \ldots\) forms a Markov chain with 6 states \(\{1, 2, 3, 4, 5, 6\}\).

(b) (3 points) Its transition probability matrix is:

\[ P = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 2/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 3/6 & 1/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 4/6 & 1/6 & 1/6 \\ 0 & 0 & 0 & 0 & 5/6 & 1/6 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \]

and the graph is

(c) (3 points) \(\{X_n\}\) is reducible because one state, state 6, is the absorbing state, i.e., you cannot go to any other state once you get to state 6.

(d) (3 points) \(\{Z_n\}\) is a Markov chain with states \(\{1, 2, 3\}\) and its transition probability matrix is

\[ P = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 2/3 & 1/3 \\ 0 & 0 & 1 \end{bmatrix} \]

Since one of the states, state 3, is absorbing, \(\{Z_n\}\) is reducible. The limiting distribution is the trivial probability distribution that assign probability one to state 3.

(e) (3 points) \(Y_1, Y_2, Y_3, \ldots\) is also a Markov chain.
First, \( Y_1, Y_2, Y_3, \ldots \) a sequence of random variables that take values in the same set of \( \{1, 2, 3, 4, 5, 6\} \).

Second, \( Y_1, Y_2, Y_3, \ldots \) has the Markov Property, that is, for \( n = 1, 2, 3, \ldots \)

\[
P\{Y_{n+1} = s_{n+1} | Y_n = s_n, \ldots, Y_1 = s_1\} = \begin{cases} 
\frac{1}{6} & 1 \leq s_{n+1} < s_n \\
1 - \frac{s_n}{6} & s_{n+1} = s_n \\
0 & s_{n+1} > s_n
\end{cases}
\]

That is, given the current state \( s_n \) and past states \( s_{n-1}, \ldots, s_1 \), the distribution of the next state \( Y_{n+1} \) depends only on \( s_n \), the value of the current state.

So \( Y_1, Y_2, Y_3, \ldots \) forms a Markov chain with 6 states \( \{1, 2, 3, 4, 5, 6\} \).

7. A few more to be added.

**Practice problems** (Each extra credit of 5 pts; some of these have been solved/sketched in class!)

1. Consider a Poisson process \( \mathcal{X} = \{X_t\}_{t \geq 0} \). Fix a time \( t_0 \geq 0 \). Show that \( \mathcal{Y} \) satisfies the fresh-start property, i.e. show that the process \( \mathcal{Y} = \{Y_t\}_{t \geq 0} \) with \( Y_t = X_{t+t_0} - X_{t_0} \) is also a Poisson process and it is independent from \( \{X_t\}_{0 \leq t < t_0} \). Hint: show that \( \mathcal{Y} \) satisfies the properties that describe a Poisson process according to the first definition of Poisson process that we discussed in class.

**Solution:**
We need to check that the process \( \mathcal{Y} \) satisfies the four properties that define a Poisson process. We have

- \( Y_0 = X_{t_0+t_0} - X_{t_0} = X_0 = 0 \)
- for any two time intervals \( (s, t] \) and \( (u, v] \) such that \( (s, t] \cap (u, v] = \emptyset \), we have \( Y_t - Y_s = X_{t_0+t} - X_{t_0+s} \) and \( Y_u - Y_u = X_{t_0+u} - X_{t_0+u} \); since \( \mathcal{X} \) is a Poisson process and \( (t_0+s, t_0+t] \cap (t_0+u, t_0+u] = \emptyset \) if \( (s, t] \cap (u, v] = \emptyset \), it follows that \( Y_t - Y_s \) and \( Y_u - Y_u \) are independent
- for any \( s, t \geq 0 \) we have \( Y_{t+s} - Y_s = X_{t+s+t_0} - X_{s+t_0} \sim \text{Poisson}(\lambda(t+s+t_0-(s+t_0)) \equiv \text{Poisson}(\lambda t). \)
Notice that, because $\mathbb{X}$ is a Poisson process, for any $t \geq 0$, $Y_t = X_{t+t_0} - X_{t_0}$ is independent of any increment $X_u - X_v$ with $0 \leq u < v < t_0$. Thus, $Y$ is independent from $\{X_t\}_{0 \leq t < t_0}$. We thus conclude that Poisson processes satisfy the fresh-start property.

2. Consider a Poisson process $\mathbb{X} = \{X_t\}_{t \geq 0}$ and the time of the first arrival $T_1$. Show that, conditional on the event $\{X_t = 1\}$ (i.e. there was only one arrival until time $t > 0$), $T_1$ is uniformly distributed on $[0,t]$. Use the definition of conditional probability and the express the event $\{T_1 \leq s \cap X_t = 1\}$ in terms of Poisson process count increments to show that the function $s \mapsto P(T_1 \leq s|X_t = 1)$ corresponds to a $\text{Uniform}(0,t)$ c.d.f..

**Solution:**

We want to compute $P(T_1 \leq s|X_t = 1)$. We have

$$P(T_1 \leq s|X_t = 1) = \frac{P(T_1 \leq s \cap X_t = 1)}{P(X_t = 1)}$$

$$= \begin{cases} 
0 & \text{if } s < 0 \\
\frac{P(X_s - X_0 = 1 \cap X_t - X_s = 0)}{P(X_t = 1)} & \text{if } 0 \leq s < t \\
1 & \text{if } s \geq t 
\end{cases}$$

since, when $0 \leq s < t$, the intersection of the events ‘the first arrival was before time $s$’ and ‘there was an arrival before time $t$’ is ‘there was an arrival in the time interval $[0,s]$ and there were no other arrivals in the time interval $(s,t]$’.

For $0 \leq s < t$, the independent independence and the stationarity of the increments of a Poisson process imply that

$$\frac{P(X_s - X_0 = 1 \cap X_t - X_s = 0)}{P(X_t = 1)} = \frac{P(X_s - X_0 = 1)P(X_t - X_s = 0)}{P(X_t = 1)} = \frac{e^{-\lambda s} \lambda s e^{-\lambda (t-s)}}{e^{-\lambda t} \lambda t} = \frac{s}{t}.$$

It follows that

$$P(T_1 \leq s|X_t = 1) = \begin{cases} 
0 & \text{if } s < 0 \\
\frac{s}{t} & \text{if } 0 \leq s < t \\
1 & \text{if } s \geq t 
\end{cases}$$

which is the c.d.f. of a $\text{Uniform}(0,t)$ distribution.
3. Show that if \( X_i \sim \text{Gamma}(\alpha_i, \beta) \) for \( \{1, \ldots, n\} \) and \( X_1, \ldots, X_n \) are independent, then \( X = \sum_{i=1}^{n} X_i \sim \text{Gamma}(\sum_{i=1}^{n} \alpha_i, \beta) \). Hint: the m.g.f. of \( Y \sim \text{Gamma}(\alpha, \beta) \) is

\[ m_Y(t) = (1 - \beta t)^{-\alpha} \]

which is well-defined and finite for \( t < 1/\beta \).

**Solution:**

For \( t < 1/\beta \), we have

\[ m_X(t) = \prod_{i=1}^{n} m_{X_i}(t) = \prod_{i=1}^{n} (1 - \beta t)^{-\alpha_i} = (1 - \beta t)^{-\sum_{i=1}^{n} \alpha_i}. \]

We easily see that the m.g.f. of \( X \) thus corresponds to the m.g.f. of a \( \text{Gamma}(\sum_{i=1}^{n} \alpha_i, \beta) \) distribution.

4. We showed in class that if \( X \sim \mathcal{N}(0, 1) \), then \( X^2 \sim \chi^2(1) \). We mentioned that if \( X_1, \ldots, X_n \overset{i.i.d.}{\sim} \mathcal{N}(0, 1) \), then \( Y = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n) \). Use the result that you obtained from the previous exercise to prove that \( Y \sim \chi^2(n) \).

**Solution:**

Since \( X_i^2 \sim \chi^2(1) \), we have

\[ m_{X_i}(t) = (1 - 2t)^{-\frac{1}{2}} \]

for any \( i \in \{1, \ldots, n\} \). By applying the result from the previous exercise we obtain

\[ m_Y(t) = (1 - 2t)^{-\frac{1}{2}} \sum_{i=1}^{n} \frac{1}{2} = (1 - 2t)^{-\frac{n}{2}} \]

for \( t < 1/2 \), which corresponds to the m.g.f. of a \( \text{Gamma}(n/2, 2) \equiv \chi^2(n) \) distribution.

5. Let \( X \sim \text{Binomial}(n_1, p) \) and \( Y \sim \text{Binomial}(n_2, p) \) with \( X \) and \( Y \) independent random variables. What is the distribution of \( Z = X + Y \)? (Hint: you have proven the Bernoulli version of this problem in a previous homework)

**Solution:**

We have

\[ m_Z(t) = m_X(t)m_Y(t) = [pe^t + 1 - p]^{n_1}[pe^t + 1 - p]^{n_2} = [pe^t + 1 - p]^{n_1+n_2} \]

which clearly corresponds to the m.g.f. of a \( \text{Binomial}(n_1 + n_2, p) \) distribution.
AFirstLook at Long-Term Behavior

Define $P(\infty)_{ij} = \lim_{n \to \infty} P(n)_{ij}$, if the limit exists. If the limit exists for all states $i$ and $j$, then put the entries in the matrix $P(\infty)$.

Exercise 4. [Exists?] Construct a transition matrix $P$ such that $P(\infty)$ does not exist.

Exercise 5. [The Rat Maze] A rat is placed in the top maze. If the rat is in one of the other four rooms, it chooses one of the doors at random with equal probability. The rat stays put once it reaches either the food (F) or the shock (S).

Write out the transition matrix $P$ and the limiting transition matrix $P(\infty)$ for the rat maze example. If the rat starts in room “2,” what is the probability it reaches the food before the shock?

Figure 2: The maze.
Figure 3: The solution for Problem 2 Part b

```
P=rbind(c(0, 0, 1/2, 0, 1/2, 0),
  + c(0, 0, 1/3, 1/3, 1/3, 0),
  + c(1/3, 1/3, 0, 0, 0, 1/3),
  + c(0, 1/2, 0, 0, 0, 1/2),
  + c(0, 0, 0, 1, 0),
  + c(0, 0, 1))

> "%^%^"<function(A,n){
  + if(n==1) A else {B<-A; for(i in (2:n)){A<-A%^%B}}; A
  + }
>
  > round(P %^% 2, 4)
[1,] 0.1667 0.1667 0.0000 0.0000 0.5000 0.1667
[2,] 0.1111 0.2778 0.0000 0.0000 0.3333 0.2778
[3,] 0.0000 0.0000 0.2778 0.1111 0.2778 0.3333
[4,] 0.0000 0.0000 0.1667 0.1667 0.1667 0.5000
[5,] 0.0000 0.0000 0.0000 0.0000 1.0000 0.0000
[6,] 0.0000 0.0000 0.0000 0.0000 0.0000 1.0000
>
> round(P %^% 4, 4)
[1,] 0.0463 0.0741 0.0000 0.0000 0.6389 0.2407
[2,] 0.0494 0.0957 0.0000 0.0000 0.4815 0.3735
[3,] 0.0000 0.0000 0.0957 0.0494 0.3735 0.4815
[4,] 0.0000 0.0000 0.0741 0.0463 0.2407 0.6389
[5,] 0.0000 0.0000 0.0000 0.0000 1.0000 0.0000
[6,] 0.0000 0.0000 0.0000 0.0000 0.0000 1.0000
>
> round(P %^% 10, 4)
[1,] 0.0021 0.0039 0.0000 0.0000 0.7105 0.2834
[2,] 0.0026 0.0047 0.0000 0.0000 0.5669 0.4258
[3,] 0.0000 0.0000 0.0047 0.0026 0.4258 0.5669
[4,] 0.0000 0.0000 0.0039 0.0021 0.2834 0.7105
[5,] 0.0000 0.0000 0.0000 0.0000 1.0000 0.0000
[6,] 0.0000 0.0000 0.0000 0.0000 0.0000 1.0000
>
> round(P %^% 50, 4)
[1,] 0 0 0 0 0.7143 0.2857
[2,] 0 0 0 0 0.5714 0.4286
[3,] 0 0 0 0 0.4286 0.5714
[4,] 0 0 0 0 0.2857 0.7143
[5,] 0 0 0 0 1.0000 0.0000
[6,] 0 0 0 0 0.0000 1.0000
```
<table>
<thead>
<tr>
<th></th>
<th>$(F, 0)$</th>
<th>$(2, 0)$</th>
<th>$(3, 0)$</th>
<th>$(4, 0)$</th>
<th>$(5, 0)$</th>
<th>$(S, 1)$</th>
<th>$(F, 1)$</th>
<th>$(2, 1)$</th>
<th>$(3, 1)$</th>
<th>$(4, 1)$</th>
<th>$(5, 1)$</th>
<th>$(S, 2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(F, 0)$</td>
<td>1/2</td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(2, 0)$</td>
<td>1/2</td>
<td></td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(3, 0)$</td>
<td>1/3</td>
<td></td>
<td>1/3</td>
<td>1/3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(4, 0)$</td>
<td></td>
<td>1/3</td>
<td>1/3</td>
<td>1/3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(5, 0)$</td>
<td></td>
<td></td>
<td>1/2</td>
<td>1/2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(S, 1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>$(F, 1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(2, 1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(3, 1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1/2</td>
<td>1/2</td>
</tr>
<tr>
<td>$(4, 1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1/3</td>
</tr>
<tr>
<td>$(5, 1)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(S, 2)$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Figure 4: Maze Solution