INSTRUCTIONS:

1. The exam has three parts: Part I, Part II, and Part III. Part I and II constitute the regular credit portion of the exam worth 130 points, and Part III is for extra credit (worth 70 points). Part I contains 8 problems, Part II contains 5 problems, and Part III contains 3 problems. You have 180 minutes (3 hours) to solve the problems. You may write short solutions for problems in Part I.

2. You may leave your answer in terms of factorials, binomial coefficients, and/or powers of numbers.

3. Make sure you write your solutions ONLY in the space provided below each problem. There is plenty of space for each problem. You can ask for extra sheets for doing scratchwork (you don’t need to submit the scratchwork).

4. You may refer to physical copies of any books or lecture notes during the exam. However, the use of any electronic devices will lead to the cancellation of your exam and a zero score, with the possibility of the authorities getting involved.

5. Make sure you write your name and NetID in the space provided above.

6. If we catch you cheating, or later suspect that your answers were copied from someone else, you will be given a zero on the exam, and might even be reported to the authorities!
Part I

Total points: 80
Number of problems: 8
Recommended duration: 30 mins

Problem 1. [10 pts]
How many integers are there between 1 and 1000 that are multiples of 4 or 5?

Proof. Let $A$ be the integers between 1 and 1000 that are divisible by 4, and let $B$ the integers in that range divisible by 5. We want to find $|A \cup B|$. Using inclusion-exclusion for two sets,

$$|A \cup B| = |A| + |B| - |A \cap B|.$$

$|A| = \frac{1000}{4} = 250$, $|B| = \frac{1000}{5} = 200$. Note that $A \cap B$ is basically all integers in the given range that are divisible by 20, and so $|A \cap B| = 50$. This means that

$$|A \cup B| = 250 + 200 - 50 = 400.$$
Problem 2. [10 pts]
What is the probability that 6 cards chosen at random from a complete deck of 52 cards contain an odd number of spades?

Proof. There are three disjoint cases/events to be considered here: either we pick exactly 1 spade, or exactly 3 spades, or exactly 5 spades.

\[
P(\text{exactly one spade is picked}) = \binom{13}{1} \binom{39}{5} \cdot \binom{52}{6}.
\]

\[
P(\text{exactly three spades are picked}) = \binom{13}{3} \binom{39}{3} \cdot \binom{52}{6}.
\]

\[
P(\text{exactly five spades are picked}) = \binom{13}{5} \binom{39}{1} \cdot \binom{52}{6}.
\]

So the probability of picking an odd number of spades is

\[
\binom{13}{1} \binom{39}{5} + \binom{13}{3} \binom{39}{3} + \binom{13}{5} \binom{39}{1} \cdot \binom{52}{6}.
\]

\]
Problem 3. [10 pts]
How many distinct strings can be formed by permuting the letters of the word “GRAMMARIAN”?

Proof. The given string contains 3 ‘A’s, 2 ‘M’s, 2 ‘R’s, and one occurrence each of ‘I’, ‘N’, and ‘G’. Using the permutations with repetitions formula or BOOKKEEPER’S rule, we have that the total number of distinct strings is

$$\frac{10!}{3!(2!)^2}.$$
Problem 4. [10 pts]
An urn contains 10 balls with the number ‘5’ printed on them, 6 balls with ‘4’ printed on them, and 4 balls with ‘2’ printed on them. Consider the experiment of picking a random ball from the urn and observing the number printed on the ball. If $X$ is the number observed, what is $E[X]$?

Proof. Range$(X) = \{2, 4, 5\}$. There are 20 balls in total. 10 of them have a ‘5’ printed and them, and so

$$P(X = 5) = \frac{10}{20} = \frac{1}{2}.$$  

Since there are 6 balls with ‘4’ printed on them,

$$P(X = 4) = \frac{6}{20} = \frac{3}{10}.$$  

Similarly,

$$P(X = 2) = \frac{4}{20} = \frac{1}{5}.$$  

Using the formula for expectation, we know that

$$E[X] = \sum_{a \in \text{Range}(X)} aP(X = a) = 5 \cdot \frac{1}{2} + 4 \cdot \frac{3}{10} + 2 \cdot \frac{1}{5} = \frac{41}{10} = 4.1.$$  

$\square$
Problem 5. [10 pts]
What is the coefficient of $x^{87}$ in $x^{45}(4x - 3)^{100}$?

Proof. We will use the binomial theorem to solve this. Recall the binomial theorem says that the coefficient of $x^k$ in $(ax + b)^n$ is $\binom{n}{k} a^k b^{n-k}$. We want to find the coefficient of $x^{87}$ in $x^{45}(4x - 3)^{100}$, which will basically be the coefficient of $x^{42}$ in $(4x - 3)^{100}$, which using the binomial theorem is

$$\binom{100}{42} 4^{42} (-3)^{58}.$$
Problem 6. [10 pts]
A palindromic binary string is a binary string which reads the same backward as forward. For example, 1010101, 110010011, 11100111 are all palindromic binary strings. If we pick a string at random from the set of all binary strings of length 11 what is the probability that it is a palindromic binary string?

Proof. The sample space is the set of all binary strings of length 11, and so $|\Omega| = 2^{11}$. We are interested in palindromic binary strings of length 11. The last 5 bits of such strings are basically the reverse of the first 5 bits, and so the last 5 bits are completely determined by the first 5 bits. This means that in order to count the number of palindromic strings of length 11, we only need to count the total number of possibilities for the first 6 bits, which is $2^6$, and so the probability that we pick a palindromic string is

$$\frac{2^6}{2^{11}} = \frac{1}{2^5}.$$
Problem 7. [10 pts]
Your friend tossed a fair coin when you weren’t around. She has a habit of messing with you every once in a while, and so there is a $1/3$ chance that she will lie to you about the result of the coin toss (report heads as tails and tails as heads), and $2/3$ chance that she will tell you the true result of the coin toss. Suppose she claims that the coin toss resulted in heads. What’s the probability that she’s lying?

Proof. Let $A$ be the event that your friend claims that the coin toss resulted in heads, and let $B$ be the event that the coin toss is actually heads. Then,

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c).$$

$P(A|B) = 2/3$ since your friend is truthful with probability $2/3$ and will report a heads as heads with probability $2/3$. As for $P(A|B^c)$, if the outcome is tails then there is a $1/3$ chance that your friend will lie to you and report a tails as heads, and so $P(A|B^c) = 1/3$. Thus, using the law of total probability,

$$P(A) = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{2}.$$

Now, given that she is reporting/claiming a heads, what is the probability that she is lying, i.e the probability that the actual outcome was tails given that she is claiming it to be heads. This is just $P(B^c|A)$, and we can compute this using Bayes rule:

$$P(B^c|A) = \frac{P(B^c)P(A|B^c)}{P(A)} = \frac{\frac{1}{2} \cdot \frac{1}{3}}{\frac{1}{2}} = \frac{1}{3}.$$

(The use of the tree method to solve this problem is also acceptable).
Problem 8. [10 pts]
Consider the following Venn diagram depicting three events A, B, and C inside a probability space $(\Omega, P)$:

(The circle on the left represents the event $C$, the bigger circle on the right is the event $B$, and the smallest circle in the diagram is the event $A$). If $P(A), P(B), P(C) > 0$, what is $P(C|B)$ and $P(B|A)$?

Proof. Since $A \subseteq B$ (from the Venn diagram), $A \cap B = A$, and so

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)}{P(A)} = 1.$$ 

Now, $B \cap C = \emptyset$ (from the Venn diagram), and so,

$$P(C|B) = \frac{P(B \cap C)}{P(B)} = \frac{P(\emptyset)}{P(C)} = 0.$$ 

\qed
Problem 9. [10 pts]
Consider a building in Manhattan with $n+1$ floors (The bottom most floor of the building is floor 1). Suppose $k$ people enter an elevator on the first floor, one after the other, and each of them independently presses a random floor button on the panel (excluding the first floor button, i.e. none of them press the first floor button; they only press buttons for floors 2 to $n+1$). What is the expected number of floors the elevator makes a stop at, not counting the first floor?

Proof. Define random variables $X_i$, for $2 \leq i \leq n+1$, such that $X_i$ is 1 if at least one of the $k$ persons presses the button for floor $i$, and 0 otherwise. Also note that the elevator stops at floor $i$ if and only if at least one of the $k$ persons presses the button for floor $i$. Thus, if $X$ is the number of stops (not including floor one) that the elevator makes, the $X = \sum_{i=2}^{n+1} X_i$.

Using linearity of expectation, we know that
\[
\mathbb{E}[X] = \sum_{i=2}^{n+1} \mathbb{E}[X_i].
\]

Also, $\mathbb{E}[X_i] = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = P(X_i = 1)$, so we basically need to compute $P(X_i = 1)$ which is the same the probability that at least one of the $k$ persons presses the button for floor $i$. We will instead find the probability that none of the $k$ persons press the button for floor $i$, i.e. $P(X_i = 0)$, and then use the fact that $P(X_i = 1) = 1 - P(X_i = 0)$.

Note that all $k$ persons press a random floor button between 2 and $n+1$, and they do so independently. Thus, we can represent the outcomes of this process as strings of length $k$, where every entry in the string is a number between 2 and $n+1$, and the $j^{th}$ entry of this string represents the floor chosen/pressed by the $j^{th}$ person. Clearly, the total number of outcomes is $n^k$. Out of these, the outcomes that correspond to the event $X_i = 0$ i.e. the event that none of the $k$ persons press/choose floor $i$ are all sequences/strings of length $k$ that don’t contain the number $i$, and the number of such strings is $(n-1)^k$, so
\[
P(X_i = 0) = \frac{(n-1)^k}{n^k},
\]
and so
\[
P(X_i = 1) = 1 - \frac{(n-1)^k}{n^k}.
\]

We can conclude that
\[
\mathbb{E}[X] = n \left(1 - \frac{(n-1)^k}{n^k}\right).
\]
More space for Problem 9
**Problem 10.** [10 pts]

Find the number of non-negative integer solutions (i.e., the solutions where all the variables are greater than or equal to zero) to the equation

\[ x + y + z + w = 15 \]

that satisfy the constraints \( x, y, z \leq 5 \).

**Hint:** Let \( A_x \) be all solutions which obey \( x \leq 5 \), \( A_y \) all solutions that obey \( y \leq 5 \), and \( A_z \) all solutions that obey \( z \leq 5 \). Then you want to find \( |A_x \cap A_y \cap A_z| \). Use inclusion-exclusion to do this.

**Proof.** Let \( S \) denote the set of all nonnegative integer solutions to the equation \( x + y + z + w = 15 \). This is \( \binom{15+4-1}{4-1} = \binom{18}{3} \). We want to find \( |A_x \cap A_y \cap A_z| \), but instead we will first compute

\[ |S \setminus (A_x \cap A_y \cap A_z)| = |(A_x \cap A_y \cap A_z)^c| = |A_x^c \cup A_y^c \cup A_z^c|. \]

Using inclusion-exclusion for three sets,

\[ |A_x^c \cup A_y^c \cup A_z^c| = |A_x^c| + |A_y^c| + |A_z^c| - |A_x^c \cap A_y^c| - |A_x^c \cap A_z^c| - |A_y^c \cap A_z^c| + |A_x^c \cap A_y^c \cap A_z^c|. \]

We make a couple of observations:

1. \( |A_x^c \cap A_y^c \cap A_z^c| \) is the set of all nonnegative integer solutions in which \( x > 5 \), \( y > 5 \), and \( z > 5 \), which is zero since the RHS is 15 and \( w \) is nonnegative.

2. By symmetry, \( |A_x^c| = |A_y^c| = |A_z^c| \). Furthermore, \( |A_x^c| \) is number of nonnegative integer solutions in which \( x > 5 \) or \( x \geq 6 \), which is just \( \binom{9+4-1}{4-1} = \binom{12}{3} \).

3. By symmetry, \( |A_x^c \cap A_y^c| = |A_x^c \cap A_z^c| = |A_y^c \cap A_z^c| \). Furthermore, \( |A_x^c \cap A_y^c| \) is the number of nonnegative integer solutions in which \( x > 5 \) and \( y > 5 \), i.e. \( x, y \geq 6 \), and the number of such solutions is \( \binom{3+4-1}{4-1} = \binom{6}{3} \).

Combining all these observations with the inclusion-exclusion expression, we get

\[ |A_x^c \cup A_y^c \cup A_z^c| = 3 \cdot \binom{12}{3} - 3 \cdot \binom{6}{3}, \]

and so, using the difference method,

\[ |A_x \cap A_y \cap A_z| = |S| - |A_x^c \cup A_y^c \cup A_z^c| = \binom{18}{3} - \left( 3 \cdot \binom{12}{3} - 3 \cdot \binom{6}{3} \right). \]

**Note:** This problem can also be solved using a hack. Notice that we can directly count the number of nonnegative integer solutions that satisfy \( x, y, z \leq 5 \). We let \( x, y, z \) take values in \( \{0, 1, 2, \ldots, 5\} \) independently, and set \( w = 15 - (x + y + z) \) to obtain a possible solution for the equation obeying all constraints (there is no constraint on \( w \) other than \( w \geq 0 \)). Thus, the total number of solutions that obey the given conditions is \( 6^3 \).

\( \square \)
More space for Problem 10
Problem 11. [10 pts]
A game consists of 100 independent rounds, where each round proceeds as follows:

1. Toss a fair coin.
2. If the outcome is heads, pick a coin with two heads, otherwise pick a fair coin.
3. Toss the coin that was picked in Step 2.
4. If the outcome of the coin toss in Step 3 is heads, you lose $4, otherwise you win $8 if the toss in Step 3 results in tails.

Let $X$ be your net gain, i.e. the total amount of money won by you minus the total amount of money lost by you at the end of the game. What is $E[X]$ and $Var(X)$?

**Proof.** Using the law of total probability (or the tree method) we can compute the probability of winning money in any given round, which is the same as the probability of the final coin toss resulting in a tails:

$$P(\text{the final coin toss results in tails}) = \frac{1}{2} \cdot \frac{1}{2} + \frac{1}{2} \cdot 0 = \frac{1}{4}.$$ 

Let $X_i$ be the net gain for round $i$, i.e. $X_i$ is money won in round $i$ minus money lost in round $i$. Then the net gain at the end of the game $X$ is $\sum_{i=1}^{100} X_i$. Using linearity of expectation,

$$E[X] = \sum_{i=1}^{100} E[X_i].$$

The expected net gain in round $i$, i.e. $E[X_i]$, is

$$E[X_i] = P(\text{final toss is tails}) \cdot 8 + P(\text{final toss is heads})(-4)$$

$$= \frac{1}{4} \cdot 8 + \frac{3}{4} \cdot (-4) = 2 - 3 = -1,$$

and so $E[X] = -100.$

Since $X_1, \ldots, X_n$ are mutually independent,

$$Var(X) = \sum_{i=1}^{100} Var(X_i),$$

so we only need to compute $Var[X_i]$. We will the formula for variance:

$$Var(X_i) = E[X_i^2] - (E[X_i])^2.$$ 

We already know the second term. For the first term, i.e. $E[X_i^2]$, we can do:

$$E[X_i^2] = P(\text{final toss is tails}) \cdot 64 + P(\text{final toss is heads})(16)$$

$$= \frac{1}{4} \cdot 64 + \frac{3}{4} \cdot 16 = 28.$$ 

Thus, $Var(X_i) = 28 - (E[X_i])^2 = 28 - 1 = 27$, and so $Var(X) = 2700.$

\[\square\]
More space for Problem 11
**Problem 12.** [10 pts]
Suppose we randomly permute the letters of the word “ABRUPT” such that that every permutation is equally likely. Let $X$ be the distance between ‘A’ and ‘T’ in the random permutation. For example, the distance between ‘A’ and ‘T’ in “ABRUPT” is 4, in “ATBRUP” is 0, and in “RBAPUT” is 2. Compute $\mathbb{E}[X]$.

**Proof.** Let us represent ‘A’ and ‘T’ using $\ast$, and the rest of the letters using $+$. So for example, “ABRUPT” becomes $\ast + + + + +$. Notice that the distance between ‘A’ and ‘T’ in a permutation of “ABRUPT” is the same as the distance between the two $\ast$ in the permutation after we convert the letters into $\ast$ and $+$. The next thing to notice is that every pattern of 2 $\ast$ and 4 $+$ corresponds to $4!2!$ different permutations of “ABRUPT”: there are $2!$ ways of replacing the two $\ast$ with ‘A’ and ‘T’, and $4!$ ways of replacing the four $+$ with ‘R’, ‘U’, ‘P’, ‘B’. Since each pattern of 2 $\ast$ and 4 $+$ corresponds to the same number of permutations of “ABRUPT”, we can completely forget about the actual letters. Thus, the problem reduces to the following:

“Suppose we pick a uniformly random pattern consisting of 2 $\ast$ and 4 $+$. Let $X$ be the distance between the two $\ast$. What is $\mathbb{E}[X]$?”

To solve the above problem, we note that $\text{Range}(X) = \{0, 1, 2, 3, 4\}$, and so For the event $[X = 4]$, there is only pattern: $\ast + + + +$, for the event $[X = 3]$, there are two possible patterns: $\ast + + + +$ and $+ + + + +$. For $[X = 2]$, we have three: $\ast + + * + +$, $+ * + + * +$, and $+ + + * +$. As for $[X = 1]$, there are four possible patterns: $\ast * + + +$, $+ * * + +$, $+ + + * +$, and $+ + + + *$. As for $[X = 0]$, there are five possible patterns: $\ast * * * +$, $+ * * * +$, $+ + * * +$, $+ + + * +$, and $+ + + + *$. Thus, the total number of patterns is $1 + 2 + 3 + 4 + 5 = 15$, and so

\[
\begin{align*}
P(X = 1) &= \frac{5}{15} \quad \text{(for [X = 4])} \\
P(X = 1) &= \frac{4}{15} \quad \text{(for [X = 3])} \\
P(X = 2) &= \frac{3}{15} \quad \text{(for [X = 2])} \\
P(X = 3) &= \frac{2}{15} \quad \text{(for [X = 1])} \\
P(X = 4) &= \frac{1}{15}. \\
\end{align*}
\]

Recall the definition of expectation:

\[
\mathbb{E}[X] = \sum_{i=0}^{5} i P(X = i) = 0 \cdot P(X = 0) + 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) + 4 \cdot P(X = 4).
\]

Putting in values, we get

\[
\mathbb{E}[X] = 1 \cdot \frac{4}{15} + 2 \cdot \frac{3}{15} + 3 \cdot \frac{2}{15} + 4 \cdot \frac{1}{15} = \frac{20}{15} = \frac{4}{3}.
\]
More space for Problem 12
Problem 13. [10 pts]
How can we choose a uniform random number between 1 and 100 (i.e., each number between 1 and 100 is equally likely to be chosen) using only coin tosses? Consider the following algorithm:

1. Toss a fair coin 7 times. After every toss, if the result is heads write down a ‘1’, otherwise write down a ‘0’. At the end of the 7 tosses, a binary string of length 7 has been written down. For example, if the 7 tosses are H,T,T,H,H,T,T then the binary string we get is 1001100.

2. Interpret the binary string as an integer between 0 and 127 by thinking of the binary string as a binary representation of a number. For example, as in the above example, if we had the tosses as H,T,T,H,H,T,T, then the string we get is 1001100, which can be interpreted as the binary representation of the integer 76. Let’s call the integer obtained in this step $X$.

3. If $1 \leq X \leq 100$, then output $X$ and stop the algorithm, otherwise if $X > 100$ or $X = 0$, go back to Step 1.

If $T$ is number of times the algorithm has to repeat the above three steps, what is $E[T]$?

Proof. Let’s call doing Steps 1, 2, and 3 as one “iteration”. Notice that in every iteration, there are $2^7$ possible outcomes for the first step, each outcome being a unique binary string of length 7. This means that each of the $2^7$ binary strings is equally likely in the first step. For the second step, notice that each of the $2^7$ binary strings of length 7 is the binary representation for a unique integer between 0 and 127, and in fact each integer between 0 and 127 has a unique binary representation of length 7. This means that $X$ is equally likely to be equal to any of the 128 integers between 0 and 127, i.e. for every $i \in \{0,1,\ldots,127\}$,

$$P(X = i) = \frac{1}{128}.$$ 

Thus, we have the uniform distribution over the range of $X$.

If $1 \leq X \leq 100$ that counts as “success” for us. Since $X$ is uniformly distributed over $\{0,1,\ldots,127\}$,

$$P(1 \leq X \leq 100) = \frac{100}{128}.$$ 

Thus, the probability of success in any given iteration is $\frac{100}{128}$. The entire algorithm can be modelled as a geometric distribution: perform steps 1, 2 and 3, if we “succeed” (i.e., $1 \leq X \leq 100$) at the end of Step 3, then we stop, otherwise we go back to Step 1 and do another iteration. Since the probability of “success” in any given round/iteration is $\frac{100}{128}$, this is a geometric distribution with parameter $p = \frac{100}{128}$. This means that the expected number of steps to see a “success”, i.e. the expected value of $T$, is

$$E[T] = \frac{1}{p} = \frac{128}{100} = 1.28.$$ 

\[ \square \]
More space for Problem 13
Part III (Extra credit)

Total points: 70  
Number of problems: 3  
Recommended duration: 90 minutes

Problem 14. [20 pts]
Consider a coin with bias $p$, i.e. the probability of heads is $p$, and that of tails is $1 - p$. Suppose we perform $n$ independent coin tosses using this coin. Let $X$ be the total number of heads observed, and $Y$ be the total number of tails observed. Compute $Cov(X,Y)$.

**Hint:** Think how $X$ and $Y$ are related to each other, or if they are independent.

**Proof.** Observe that $X + Y = n$ (since the total number of heads plus the total number of tails should be $n$, the total number of tosses). Thus, $Y = n - X$. Recall that

$$ Cov(X,Y) = E[XY] - E[X]E[Y], $$

and so

$$ Cov(X,Y) = E[X(n - X)] - E[X]E[n - X] $$

$$ = nE[X] - E[X^2] - nE[X] + (E[X])^2 $$

$$ = -E[X^2] + (E[X])^2 = - (E[X^2] - (E[X])^2) = -Var(X), $$

where the last equality uses the formula for variance.

Also note that $X$ is a binomial random variable, and so $Var(X) = np(1 - p)$. This means that

$$ Cov(X,Y) = -np(1 - p). $$

□
More space for Problem 14:
Problem 15. \[10 + 10 + 5 = 25 \text{ pts}\]
The goal of this problem is to prove the following statement:

Let \(X_1, \ldots, X_n\) be pairwise independent random variables. Then

\[
\text{Var}(X_1 + X_2 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i).
\]

1. Prove using induction, that for all \(n \geq 2\),

\[
(X_1 + X_2 + \ldots X_n)^2 = \sum_{i=1}^{n} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n} X_i X_j \right).\]

2. Use the result from the first part along with the definition of variance to show that

\[
\text{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i) + 2 \left( \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \right).\]

3. Use the pairwise independence of \(X_1, \ldots, X_n\) to complete the proof.

Proof. 1. For \(n = 2\), we have \((X_1 + X_2)^2 = (X_1 + X_2)(X_1 + X_2) = X_1^2 + X_1 X_2 + X_2 X_1 + X_2^2 = X_1^2 + X_2^2 + 2X_1 X_2 = \sum_{i=1}^{2} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq 2} X_i X_j \right).\) This completes the base case.

For the inductive hypothesis, assume that \((X_1 + X_2 + \ldots + X_n)^2 = \sum_{i=1}^{n} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n} X_i X_j \right).\) We will show that the relation also holds true for \(n + 1\) variables.

We want to show that

\[
(X_1 + X_2 + \ldots X_n + X_{n+1})^2 = \sum_{i=1}^{n+1} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n+1} X_i X_j \right).
\]

Let’s start with the LHS. We can think of it as

\[
((X_1 + X_2 + \ldots + X_n) + X_{n+1})^2.
\]

Let \(Y = (X_1 + X_2 + \ldots + X_n)\). Then using the formula for the \(n = 2\) case, we get

\[
(Y + X_{n+1})^2 = Y^2 + X_{n+1}^2 + 2YX_{n+1}.
\]

Putting back the value of \(Y\), we get

\[
(X_1 + X_2 + \ldots + X_n + X_{n+1})^2 = (X_1 + X_2 + \ldots + X_n)^2 + X_{n+1}^2 + 2(X_1 + X_2 + \ldots + X_n)X_{n+1},
\]

which, using the induction hypothesis,

\[
= \sum_{i=1}^{n} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n} X_i X_j \right) + X_{n+1}^2 + 2(X_1 X_{n+1} + X_2 X_{n+1} + \ldots + X_n X_{n+1}).
\]

Combining terms, we get that

\[
(X_1 + X_2 + \ldots + X_n + X_{n+1})^2 = \sum_{i=1}^{n+1} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n+1} X_i X_j \right).
\]
2. Recall that

\[\text{Var}(X_1 + \ldots + X_n) = \mathbb{E}[(X_1 + \ldots + X_n)^2] - (\mathbb{E}[X_1 + \ldots + X_n])^2.\]

Using part 1, the RHS is

\[= \mathbb{E} \left[ \sum_{i=1}^{n} X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n} X_i X_j \right) \right] - (\mathbb{E}[X_1 + \ldots + X_n])^2\]

\[= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2 \left( \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j] \right) - (\mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n])^2\]

\[= \sum_{i=1}^{n} \mathbb{E}[X_i^2] + 2 \left( \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i X_j] \right) - \sum_{i=1}^{n} (\mathbb{E}[X_i])^2 - 2 \left( \sum_{1 \leq i < j \leq n} \mathbb{E}[X_i] \mathbb{E}[X_j] \right),\]

where the last step involves applying part 1 of the problem to \((\mathbb{E}[X_1] + \mathbb{E}[X_2] + \ldots + \mathbb{E}[X_n])^2\).

Rearranging terms, we get

\[\text{Var}(X_1 + \ldots + X_n)\]

\[= \sum_{i=1}^{n} (\mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2) + 2 \left( \sum_{1 \leq i < j \leq n} (\mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]) \right)\]

\[= \sum_{i=1}^{n} \text{Var}(X_i) + 2 \left( \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j) \right),\]

where the last step follows from the formula for variance and covariance.

3. If \(X_1, \ldots, X_n\) are pairwise independent, then for every \(1 \leq i < j \leq n\) we have that \(X_i\) and \(X_j\) are independent, and so \(\text{Cov}(X_i, X_j) = 0\). Using this observation in the equation proved in part 2, we get

\[\text{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \text{Var}(X_i).\]
More space for Problem 15:
Problem 16. \[5 + 5 + 15 = 25 \text{ pts}\]

Let \([n] = \{1, \ldots, n\}\), i.e. the first \(n\) positive integers. A bijection \(f : [n] \to [n]\) is said to have a fixed point at \(i \in [n]\) if \(f(i) = i\). Let \(S_n\) denote the set of all bijections with domain \([n]\) and codomain \([n]\), i.e. all bijections of the form \(f : [n] \to [n]\). Consider the experiment of picking a random bijection \(f : [n] \to [n]\) from \(S_n\) such that every bijection is \(S_n\) is equally likely to be picked. Let \(X\) be the total number of fixed points of \(f\). For \(1 \leq i \leq n\), define the indicator random variable \(X_i\) which takes value 1 if and only if \(i\) is a fixed point of \(f\), i.e. \(f(i) = i\). Clearly, \(X = \sum_{i=1}^{n} X_i\).

1. What is \(E[X]\)?

2. We will now compute \(Var(X)\). Note that \(Var(X) = Var(X_1 + \ldots + X_n)\), and so we might be tempted to use linearity to compute \(Var(\sum_{i=1}^{n} X_i)\). Unfortunately, this does not work in this case since \(X_1, \ldots, X_n\) are not even pairwise independent. Can you provide an argument in support of this fact, i.e. in support of the fact that \(X_1, \ldots, X_n\) are not pairwise independent?

3. Nevertheless, we can still compute \(Var(X) = Var(X_1 + \ldots + X_n)\). We know that \(Var(X) = E[X^2] - (E[X])^2\). We already computed the second term. To compute \(E[X^2]\), we can note that

\[
E[X^2] = E[(X_1 + \ldots + X_n)^2].
\]

Now use the result from part 1 of Problem 15 to compute \(E[X^2]\). You might want to keep in mind that \(X_1, \ldots, X_n\) are indicator random variables when computing quantities like \(E[X_i^2]\) and \(E[X_i X_j]\). Also, don’t forget: \(X_1, \ldots, X_n\) are not pairwise independent!

Proof. 1. Using of linearity of expectation, \(E[X] = \sum_{i=1}^{n} E[X_i]\), so we need to compute \(E[X_i]\). Since \(X_i\) is an indicator random variable, \(E[X_i] = P(X_i = 1) = P(f(i) = i)\). The sample space is the set of all bijections, and we will use the uniform probability measure on this sample space since each bijective function is equally likely. The size of the sample space is \(n!\). The number of bijections in which \(f(i) = i\) is \((n-1)!\) — any such bijection maps \(i\) to \(i\), and the rest of the elements can be permuted arbitrarily and there are \((n-1)!\) ways of doing this. This means that

\[
P(X_i = 1) = P(f(i) = i) = \frac{(n-1)!}{n!} = \frac{1}{n}.
\]

Note that this is true for every \(1 \leq i \leq n\), and so

\[
E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \frac{1}{n} = 1.
\]

2. We will first now compute \(P(X_1 = 1 \cap X_2 = 1)\). Note that \([X_1 = 1] \cap [X_2 = 1]\) is the event that contains all bijective functions where \(f(1) = 1\) and \(f(2) = 2\). The number of such bijective functions is \((n-2)!\) since any such function maps 1 to 1 and 2 to 2, and the rest \((n-2)\) elements can be permuted arbitrarily, and there are \((n-2)!\) ways of doing that. This means that

\[
P(X_1 = 1 \cap X_2 = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)}.
\]

However, \(P(X_1 = 1)P(X_2 = 1) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}\) using part 1 (the fact that \(P(X_i = 1) = 1/n\)). This means \(P(X_1 = 1 \cap X_2 = 1) \neq P(X_1 = 1)P(X_2 = 1)\), which means that \(X_1\) and \(X_2\)
are not independent random variables, and so this implies that $X_1, \ldots, X_n$ are not pairwise independent.

3. Using part 1 of Problem 15,

$$
E[(X_1 + \ldots + X_n)^2] = E \left[ \sum_{i=1}^n X_i^2 + 2 \left( \sum_{1 \leq i < j \leq n} X_i X_j \right) \right]
$$

$$
= \sum_{i=1}^n E[X_i^2] + 2 \left( \sum_{1 \leq i < j \leq n} E[X_i X_j] \right).
$$

Let’s first compute $E[X_i^2]$. We know that

$$
E[X_i^2] = \sum_{a \in \text{Range}(X_i)} a^2 P(X_i = a) = 1 \cdot P(X_i = 1) + 0 \cdot P(X_i = 0) = P(X_i = 1).
$$

From part 1 of this problem, we know that $P(X_i = 1) = \frac{1}{n}$. Thus,

$$
\sum_{i=1}^n E[X_i^2] = \sum_{i=1}^n \frac{1}{n} = 1.
$$

Next, for $1 \leq i < j \leq n$, we want to compute $E[X_i X_j]$. Let $Z_{ij} = X_i X_j$, then $\text{Range}(Z_{ij}) = \{0, 1\}$, and so $E[Z_{ij}] = P(Z_{ij} = 1) = P(X_i X_j = 1)$. Note that that $X_i X_j = 1$ if and only if $X_i = 1$ and $X_j = 1$, and so $E[X_i X_j] = E[Z_{ij}] = P(X_i = 1 \cap X_j = 1)$. Note that $[X_i = 1] \cap [X_j = 1]$ is the event that contains all bijective functions where $f(i) = i$ and $f(j) = j$. The number of such bijective functions is $(n-2)!$ since any such function maps $i$ to $i$ and $j$ to $j$, and the rest $(n-2)$ elements can be permuted arbitrarily, and there are $(n-2)!$ ways of doing that. This means that

$$
P(X_i = 1 \cap X_j = 1) = \frac{(n-2)!}{n!} = \frac{1}{n(n-1)},
$$

and so $E[X_i X_j] = \frac{1}{n(n-1)}$ and

$$
\sum_{1 \leq i < j \leq n} E[X_i X_j] = \sum_{1 \leq i < j \leq n} \frac{1}{n(n-1)} = \binom{n}{2} \frac{1}{n(n-1)} = \frac{1}{2}
$$

where the last two equalities follow from the observation that the number of pairs $i, j$ that satisfy $1 \leq i < j \leq n$ is $\binom{n}{2}$.

Putting everything together,

$$
E[X^2] = E[(X_1 + \ldots + X_n)^2]
$$
\[
\sum_{i=1}^{n} E[X_i^2] + 2 \left( \sum_{1 \leq i < j \leq n} E[X_i X_j] \right)
\]

\[
= 1 + 2 \cdot \frac{1}{2} = 1 + 1 = 2.
\]

Thus, \( Var(X) = E[X^2] - (E[X])^2 = 2 - 1 = 1. \)

\( \square \)