1. Miscellaneous Review

(a) Show that the probability that exactly one of the events $A$ and $B$ occur is $P(A) + P(B) - 2P(A \cap B)$.

(b) If $A$ is independent of itself, show that $P(A) = 0$ or 1.

(c) Find an example of 3 events $A$, $B$, and $C$ such that each pair of them are independent, but they are not mutually independent. Show the calculations.

(d) A bin contains one blue ball and one red ball. One of the balls is drawn at random, then replaced with another ball of the same color. A second ball is drawn at random from the three balls in the box. Given that at least one of the two drawn balls was blue, what is the probability that the first ball was blue?

Solution:

(a) The probability of the event that exactly one of $A$ and $B$ occur is

$$P(A \cap B^c) + P(A^c \cap B) = P(A) - P(A \cap B) + P(B) - P(A \cap B)$$

$$= P(A) + P(B) - 2P(A \cap B).$$

(b) $P(A \cap A) = P(A)P(A)$, so $P(A) = P(A)^2$; this implies that $P(A) \in \{0, 1\}$. Alternatively, suppose for the sake of contradiction that $0 < P(A) < 1$. Then, $P(A | A) = 1 \neq P(A)$, which contradicts the supposed independence of $A$ with itself. Hence, $P(A) \in \{0, 1\}$.

(c) Consider a fair 4-sided die. Let $A$ be the event that 1 or 2 appears in a die roll, $B$ be the event that 1 or 3 appears, and $C$ be the event that 1 or 4 appears. Then, $P(A) = P(B) = P(C) = 1/2$. Furthermore,

$$P(A \cap B) = P(\{1 \text{ appears}\}) = \frac{1}{4} = P(A)P(B).$$
So \(A\) and \(B\) are pairwise independent. Similarly \((A,C)\) and \((B,C)\) are pairwise independent. However,
\[
P(A \cap B \cap C) = P(\{1\ \text{appears}\}) = \frac{1}{4} \neq P(A)P(B)P(C) = \frac{1}{8}.
\]
So these 3 events are not mutually independent.
The answer is not unique; any other valid answer is acceptable.

(d) If \(B_1\) and \(B_2\) denote the events that the first and second balls are blue respectively, then we are looking for \(P(B_1 | B_1 \cup B_2)\). Then,
\[
P(B_1 \cap (B_1 \cup B_2)) = P(B_1) = \frac{1}{2}
\]
and \(P(B_1 \cup B_2) = P(B_1) + P(B_1^c)P(B_2 | B_1^c) = 1/2 + (1/2)(1/3) = 2/3\), so
\[
P(B_1 | B_1 \cup B_2) = (1/2)/(2/3) = 3/4.
\]

2. **Passengers on a Plane**

There are \(N\) passengers in a plane with \(N\) assigned seats (\(N\) is a positive integer), but after boarding, the passengers take the seats randomly. Assuming all seating arrangements are equally likely, what is the probability that no passenger is in their assigned seat? Compute the probability when \(N \to \infty\).

*Hint:* Use the inclusion-exclusion principle: if \(A_1, \ldots, A_N\) are events, then
\[
P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{k=1}^{N} (-1)^{k+1} \sum_{|I|=k} \prod_{i \in I} P(A_i).
\]

**Solution:**

First, let us calculate the probability that at least one passenger sits in his or her assigned seat using inclusion-exclusion. Let \(A_i, i = 1, \ldots, N,\) be the event that passenger \(i\) sits in his or her assigned seat. We first add the probabilities of the single events (of which there are \(N\)), and the probability of each event is \((N-1)!/N!\) (indeed there are \((N-1)!\) ways to permute the remaining passengers once a specific passenger is fixed, and \(N!\) total permutations, so the probability is \((N-1)!/N!\)); next, we subtract the probabilities of the pairwise intersections of events (of which there are \(\binom{N}{2}\)), and the probability of each event is \((N-2)!/N!\) (there are \((N-2)!\) ways to permute the passengers other than the fixed two); continuing on, we see that
\[
P\left(\bigcup_{i=1}^{N} A_i\right) = \sum_{j=1}^{N} (-1)^{j+1} \binom{N}{j} \frac{(N-j)!}{N!} = \sum_{j=1}^{N} (-1)^{j+1} \frac{1}{j!}.
\]
Now, the event that no passenger sits in his or her assigned seat is the complement of the event just discussed:
\[
1 - P\left(\bigcup_{i=1}^{N} A_i\right) = 1 - \sum_{j=1}^{N} (-1)^{j+1} \frac{1}{j!} = \sum_{j=0}^{N} (-1)^{j} \frac{1}{j!}.
\]
Taking the limit as \( N \rightarrow \infty \), the expression converges to \( \sum_{j=0}^{\infty} (-1)^j / j! \), and using the expression for the power series of the exponential function, we conclude that the probability converges to \( \exp(-1) \approx 0.368 \).

### 3. Joint Occurrence

You know that, at least one of the events \( A_r \) (for \( r \in \{1, \ldots, n\} \), where \( n \) is an integer \( \geq 2 \)) is certain to occur but certainly no more than two occur. Show that if the probability of occurrence of any single event is \( p \), and the probability of joint occurrence of any two distinct events is \( q \), we have \( p \geq 1/n \) and \( q \leq 2/(n(n-1)) \).

**Solution:**

Since \( 1 = \mathbb{P}(\bigcup_{r=1}^{n} A_r) \leq \sum_{r=1}^{n} \mathbb{P}(A_r) = np \), we see that \( p \geq 1/n \).

Let \( I := \{(i, j) \in \{1, \ldots, n\}^2 : i < j\} \) be the set of pairs of distinct indices, avoiding repetition. Notice that the events \( \{A_i \cap A_j : (i, j) \in I\} \) are pairwise disjoint, so by countable additivity,

\[
1 \geq \mathbb{P}\left( \bigcup_{(i,j) \in I} (A_i \cap A_j) \right) = \sum_{(i,j) \in I} \mathbb{P}(A_i \cap A_j) = \binom{n}{2} q,
\]

so \( q \leq \frac{1}{\binom{n}{2}^{-1}} = 2/(n(n-1)) \).

### 4. Expanding the NBA

The NBA is looking to expand to another city. In order to decide which city will receive a new team, the commissioner interviews potential owners from each of the \( N \) potential cities (\( N \) is a positive integer), one at a time.

Unfortunately, the owners would like to know immediately after the interview whether their city will receive the team or not. The commissioner decides to use the following strategy: she will interview the first \( m \) owners and reject all of them \( (m \in \{1, \ldots, N\}) \). After the \( m \)th owner is interviewed, she will pick the first city that is better than all previous cities. The cities are interviewed in a uniformly random order. What is the probability that the best city is selected? Assume that the commissioner has an objective method of scoring each city and that each city receives a unique score.

You should arrive at an exact answer for the probability in terms of a summation. Approximate your answer using \( \sum_{i=1}^{n} i^{-1} \approx \ln n \) and find the optimal value of \( m \) that maximizes the probability that the best city is selected.

**Solution:**

Let \( B_i, i = 1, \ldots, N \), be the event that the \( i \)th city is the best of the \( N \) cities, and let \( A \) be the event that the best city is picked by the commissioner. Then,

\[
\mathbb{P}(A) = \sum_{i=1}^{N} \mathbb{P}(A \mid B_i) \mathbb{P}(B_i) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{P}(A \mid B_i)
\]

using the law of total probability. Next, \( \mathbb{P}(A \mid B_i) = 0 \) for \( i = 1, \ldots, m \) because if the best city is among the first \( m \) cities, there is no chance of picking the
best city. Also, \( P(A | B_i) = m/(i - 1) \) for \( i = m + 1, \ldots, N \) because \( P(A | B_i) \) is the probability that second-best city among the first \( i \) cities is within the first \( m \) cities. Therefore,
\[
P(A) = \frac{m}{N} \sum_{i=m+1}^{N} \frac{1}{i-1}.
\]

Now, we turn towards approximation.
\[
P(A) \approx \frac{m}{N} (\ln N - \ln m) = -\frac{m}{N} \ln \frac{m}{N}.
\]

Letting \( x := m/N \), then \( P(A) \approx -x \ln x \), and differentiating with respect to \( x \) suggests that the optimal value is \( x = 1/e \), so we should reject the first \( N/e \) cities. Plugging in this value for \( x \) into \( P(A) = -x \ln x \) gives the optimal probability as \( P(A) \approx 1/e \).

**Note:** This problem is a famous example from optimal stopping theory and is commonly known as the secretary problem (a boss is interviewing secretaries instead of a commissioner interviewing city representatives). In fact, one may use a dynamic programming approach to see why the policy outlined here is in fact the optimal policy. If you are interested, the details of such an approach can be found in *Dynamic Programming and the Secretary Problem* by Beckmann.

5. **Superhero Basketball**

Superman and Captain America are playing a game of basketball. At the end of the game, Captain America scored \( n \) points and Superman scored \( m \) points, where \( n > m \) are positive integers. Supposing that each basket counts for exactly one point, what is the probability that after the start of the game (when they are initially tied), Captain America was always strictly ahead of Superman? (Assume that all sequences of baskets which result in the final score of \( n \) baskets for Captain America and \( m \) baskets for Superman are equally likely.)

**Hint:** Think about symmetry. First, try to figure out which is more likely: there was a tie and Superman scored the first point, or there was a tie and Captain America scored the first point?

**Solution:**

Let \( T \) be the event that Captain America and Superman were tied at least once after the first point. Let \( C \) be the event that Captain America scores the first point, and \( S \) be the event that Superman scores the first point. In fact, \( P(C \cap T) = P(S \cap T) \) for the following reason: given any sequence of baskets where the first point is scored by Captain America and there is a tie, flip all of the baskets until the first tie. This yields a sequence of baskets where the first point is scored by Superman and there is a tie. Thus, the outcomes in \( C \cap T \) are in one-to-one correspondence with the outcomes in \( S \cap T \).

However, since Captain America won the game, \( P(S \cap T) = P(S) \) (if Superman scored the first point, then there must have been a point when Captain America caught up).
So far, we have $P(T) = P(C \cap T) + P(S \cap T) = 2P(S \cap T) = 2P(S)$. We can calculate $P(S) = m/(m + n)$. Thus, $P(T) = 2m/(m + n)$. Finally, the question is asking for $P(T^c) = 1 - 2m/(m + n) = (n - m)/(m + n)$.

**Note:** This is yet another famous problem known as the ballot problem.

6. **[Bonus] Tournament Probabilistic Proof**

The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

In a tournament with $n$ players (where $n$ is a positive integer), each player plays against every other player for a total of $\binom{n}{2}$ games (assume that there are no ties). Let $k$ be a positive integer. Is it always possible to find a tournament such that for any subset $A$ of $k$ players, there is a player who has beaten everyone in $A$? For such a tournament, let us say that every $k$-subset is dominated. For example, Figure 1 depicts the smallest tournament in which every 2-subset is dominated.

![Figure 1: A tournament with 7 vertices such that every pair of players is beaten by a third player.](image)

In fact, as long as $\binom{n}{k}(1 - 2^{-k})^{n-k} < 1$, it is possible to find a tournament of $n$ players such that every $k$-subset is dominated. Prove this fact, and explain why it implies that for any positive integer $k$ there exists a tournament such that every $k$-subset is dominated.

**Solution:**

Let us build a tournament with $n$ players randomly, by deciding each game with an independent fair coin flip.

Let $A_j$, for $j = 1, \ldots, \binom{n}{k}$, be the event that the $j$th subset of $k$ players is not dominated by anyone. What is $P(A_j)$? There are $n - k$ other players, and the probability that a player fails to dominate the $j$th subset is $1 - 2^{-k}$, so the probability that none of the $n - k$ other players dominates the $j$th subset is $P(A_j) = (1 - 2^{-k})^{n-k}$. Hence:

$$P\left(\bigcup_{j=1}^{\binom{n}{k}} A_j\right) \leq \sum_{j=1}^{\binom{n}{k}} P(A_j) = \binom{n}{k}(1 - 2^{-k})^{n-k}.$$
Thus, if \( \binom{n}{k}(1 - 2^{-k})^{n-k} < 1 \), then \( \mathbb{P}(\bigcap_{j=1}^{n} A_j^c) > 0 \), i.e., there is a positive probability that every subset of \( k \) players is dominated, but this means there exists some tournament with \( n \) players such that every \( k \)-subset is dominated.

Now, the question is whether we can pick \( n \) so that \( \binom{n}{k}(1 - 2^{-k})^{n-k} < 1 \). The term \( \binom{n}{k} \) grows roughly at the rate \( n^k \), and the term \( (1 - 2^{-k})^{n-k} \) grows roughly at the rate \( c^n \) for some \( c \in (0, 1) \), and since exponential decay outpaces polynomial growth, we can indeed always pick \( n \) sufficiently large so that \( \binom{n}{k}(1 - 2^{-k})^{n-k} < 1 \) holds.

Can we figure out the order of magnitude of \( n \)? Using the simpler bounds \( \binom{n}{k} \leq n^k \) and \( (1 - 2^{-k})^{n-k} \leq \exp(-(n-k)2^{-k}) \), then it suffices to have \( n^k \exp(-(n-k)2^{-k}) < 1 \), and upon taking logarithms and rearranging we get the condition \( n - k2^k \ln n - k > 0 \). From this we at least need \( n > 2^k \) so \( \ln n > k \) but then \( k2^k \ln n > k^2 k \), which means we need to take \( n > k^2 2^k \). If we let \( n = Ck^2 2^k \) for some constant \( C > 1 \), then \( n - k2^k \ln n - k \approx (C - 1)k^2 2^k > 0 \) for large \( k \), so roughly, a constant multiple of \( k^2 2^k \) players suffices.