Homework 1

Due September 24 at 11 pm

1. (True or False) Prove the following statements or provide a counterexample. Let A, B, and C be events in a probability space.

   (a) If A and B are independent, then so are A^c and B.

   True. If A and B are independent, then $P(A|B) = P(A)$. We know that $P(A^c|B) = 1 - P(A|B) = 1 - P(A) = P(A^c)$, which implies A^c and B are independent.

   (b) If A and B are conditionally independent given C, then they are also conditionally independent given C^c.

   False. Consider the following events,
   A : **Bob passed the exam.**,
   B : **Bob submitted the exam paper on gradescope.** and
   C : **Bob has handed in his exam paper to the instructor directly**.
   Then given C, A and B are independent in that Bob passing the exam doesn't give us clue to where the exam paper is submitted, it could be because of C or B, so no useful information. But given C^c, we know that if Bob passed the exam, he has to submit his paper somewhere, which increases the probability of B happening, which means A and B are not independent given C^c.

   (c) Events in a partition cannot be independent (assume that every event in the partition has nonzero probability).

   True. Consider a partition $\{A_i\}_{i=1}^n$ and that $A_i \cap A_j = \emptyset, \forall i, j \in \{1, 2, \cdots, n\}$ and that $\bigcup_{i=1}^n A_i = \Omega$, $P(A_i) > 0$. Then if A_i's are mutually independent, then we must at least have $P(A_i \cap A_j) = P(A_i)P(A_j) = 0$, which implies $P(A_i)$ or $P(A_j)$ has to be zero, and this is a contradiction. So the original argument holds only when $P(A_i)$ can take zero value.

   (d) If $\mathrm{P}(A|B) = 1$ then $\mathrm{P}(B^c|A^c) = 1$.

   True. We have the following derivations:

$$
\begin{aligned}
P(B^c|A^c) &= \frac{P(A^c \cap B^c)}{P(A^c)} && \text{(By definition)} \\
&= \frac{P((A \cup B)^c)}{P(A^c)} && \text{(Demorgan's Law)} \\
&= \frac{1 - P(A \cup B)}{1 - P(A)} && \text{(Axiom of Probability)} \\
&= \frac{1 - (P(A) + P(B) - P(A \cap B))}{1 - P(A)} && \text{(Inclusion-exclusion)} \\
&= \frac{1 - P(A)}{1 - P(A)} && \text{(P(A given B)=1, P(A and B)=P(B))} \\
&= 1
\end{aligned}
$$

2. (Probability spaces)

(a) Let $(\Omega, \mathcal{C}, \mathrm{P})$ be a probability space. Let A be an event in the collection $\mathcal{C}$, such that $\mathrm{P}(A) \neq 0$, on which we want to condition. We define a collection of events $\mathcal{C}_A$ as the collection of the intersection of A with all the events in $\mathcal{C}$:

$$\mathcal{C}_A = \{A \cap F : F \in \mathcal{C}\}.$$

If we consider a new sample space $\Omega_A := A$, prove that $\mathcal{C}_A$ is a valid collection, and also that the conditional probability measure

$$\mathrm{P}_A(S \cap A) := \frac{\mathrm{P}(S \cap A)}{\mathrm{P}(A)}, \tag{1}$$

where $S \in \mathcal{C}$, is a valid probability measure on $\mathcal{C}_A$.

We first prove that $\mathcal{C}$ is a valid collection.

Proof. i. Since $\emptyset \in \mathcal{C}$, then $A \cap \emptyset = \emptyset \in \mathcal{C}$.

ii. Suppose $E \in \mathcal{C}_A$, then $\exists F_0 \in \mathcal{C}$ s.t. $E = A \cap F_0$, then $E^c = (A \cap F_0)^c = A^c \cup F_0^c$. Since $\mathcal{C}$ is a valid σ-algebra, since $A, F_0 \in \mathcal{C}$, then $A, F_0^c \in \mathcal{C}$, then $A \cap F_0^c \in \mathcal{C}_A$ by definition, which implies that $E^c \in \mathcal{C}_A$.

iii. We pick countable union of the events in $\mathcal{C}_A$, then $A \cap (\cup_{i=1}^{\infty} F_i) = \cup_{i=1}^{\infty} (A \cap F_i) \in \mathcal{C}_A$, thus $\mathcal{C}_A$ is closed under countable union.

. Thus $\mathcal{C}_A$ is a valid collection. □

We examine the following three properties:

i. Since P is a valid probability measure, then $P(S \cap A) \geq 0, P(A) > 0$, thus $P_A(S \cap A) \geq 0, \forall S \cap A \in \mathcal{C}_A$.

ii. We have $P_A(\Omega_A \cap A) = \frac{P(A \cap A)}{P(A)} = \frac{P(A)}{P(A)} = 1$.

iii. We pick countable union of the events in $\mathcal{C}_A$, and we have $P_A(\cup_{i=1}^{\infty}(S_i \cap A)) = \frac{P\left(\cup_{i=1}^{\infty}(S_i \cap A)\right)}{P(A)} = \frac{\sum_{i=1}^{\infty} P(S_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} \frac{P(S_i \cap A)}{P(A)} = \sum_{i=1}^{\infty} P_A(S_i \cap A)$.

, thus P_A is a valid probability measure.

(b) Suppose we have a sample space $\Omega = \{1, \ldots, M\}$ with collection $\mathcal{C} := 2^{\Omega}$, the power set of Ω. To determine P, the probability measure, we employ the following empirical procedure:

   i. Collect N data points taking values in Ω (e.g., N rolls of an M-sided die). Call these observations $x_1, \ldots, x_N$.

   ii. For each $S \subseteq \Omega$, define

$$\mathrm{P}(S) := \frac{\text{number of } i\text{-values such that } x_i \in S}{N}.$$

   As an example, suppose $M = 2$ and we flip a coin $N = 10$ times getting 6 heads and 4 tails, where 1 denotes head and 2 denotes tail. Then

$$\mathrm{P}(\emptyset) = 0, \quad \mathrm{P}(\{1\}) = 0.6, \quad \mathrm{P}(\{2\}) = 0.4, \quad \text{and} \quad \mathrm{P}(\{1, 2\}) = 1.$$

   If P is defined using the above procedure, is it a valid probability measure? Either prove that it is, or give a counterexample.

Yes, it is a valid probability measure.

(a) Because the number on the numerator is always bigger than or equal to zero, $P(S) \geq 0, \forall S \subseteq \Omega$.

(b) $P(\Omega) = \frac{\textbf{number of i-values such that } x_i \in \Omega}{N} = \frac{N}{N} = 1$

(c) $P(\bigcup_i^{\infty} A_i) = \frac{\sum_{i=1}^{\infty} \textbf{number of i-values such that } x_i \in A_i}{N} = \sum_{i=1}^{\infty} \frac{\textbf{number of i-values such that } x_i \in A_i}{N} = \sum_{i=1}^{\infty} P(A_i)$.

, thus P is a valid measure.

3. (Testing) A company with 10 employees decides to test them for COVID-19 before they go back to work in person. From available data, they determine that the probability of each employee being ill is 0.01. The employees have not been in contact with each other for a while, so the events *Employee i is ill*, for $1 \leq i \leq 10$, are modeled as independent. If an employee is ill, the test is positive with probability 0.98. If they are not ill, the test is positive with probability 0.05.

   (a) Is it reasonable to model the events *Test i is positive*, for $1 \leq i \leq 10$, as independent? From now on model them as independent whether you think it is reasonable or not.

   Yes. Because knowing one employee tests positive doesn't give us any clue to whether other employees' tests result. There could be false positive or false negative scenarios. In other words, know one employee's testing results won't increase the probability that another employee will test positive or negative.

   (b) The company tests all employees. What is the probability that there is at least one positive test?

   Denote T_i: Employee i tests positive. I_i: Employee i is ill. By law of total probability, the probability that employee i tests positive is $P(T_i) = P(T_i|I_i)P(I_i) + P(T_i|I_i^c)P(I_i^c) = 0.98 * 0.01 + 0.05 * (1 - 0.01) = 0.0593$. Denote A: At least one positive test, then $P(A) = 1 - P(A^c) = 1 - (1 - 0.0593)^{10} = 0.45736$.

   (c) If there is at least one positive test, what is the probability that nobody is ill? If you make any independence or conditional independence assumptions, please justify them.

   Denote B: No body is ill.
   From the problem we know that $P(B) = 0.99^{10} = 0.9044$. By Bayes rule, we have $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$. Denote N_i: Employee i tests negative. Then $P(A|B) = 1 - P(A^c|B) = 1 - P(\bigcap_{i=1}^{10} N_i|B)$. We assume that N_i's are independent given B. Since given B, knowing the testing results from some employees won't tell us what happens to others, the test results could be affected by lots of factors. Then we have $P(\bigcap_{i=1}^{10} N_i|B) = \prod_{i=1}^{10} P(N_i|B) = (1 - 0.05)^{10} = 0.5987$. Thus $P(A|B) = 1 - 0.5987 = 0.4013$, which means $P(B|A) = \frac{0.4013 \times 0.9044}{0.45736} = 0.83443$.

4. (Streak of heads) In this problem we consider the problem of testing whether a randomly generated sequence is truly random. A certain computer program is supposed to generate independent fair coin flips. When you try it out, you are surprised that it contains long streaks of 1s. In particular, you generate a sequence of length 200, which turns out to contain a sequence of 8 heads in a row.

   (a) Compute the probability that the longest streak of heads that you observe has length x for $x \in \{1, 2, 3, 4, 5\}$ when you flip a fair coin 5 times, and the flips are independent.

   We calculate them one by one. The whole sample space can be regarded as a 5-bit long sequence, and the heads are encoded to be 1 while the tails are encoded to be zeros.

      i. x=0, we only have 00000, so we expect $\frac{1}{32} = 0.03125$
      ii. x=1: We have 10000,01000,00100,00010,00001,10100, 10010,10001, 10101, 01010,01001,00101 so we expect $\frac{12}{2^5} = \frac{3}{8} \approx 0.375$.
      iii. x=2: We have 11011,11010,11001,11000,10110,10011,01101,01100,01011,00011,00110 so we expect $\frac{11}{32} \approx 0.34375$.
      iv. x=3: We have 11100,11101,01110,00111,10111 so we expect $\frac{5}{32} \approx 0.1563$.
      v. x=4: We have 11110, 01111 so we expect $\frac{2}{2^5} = \frac{1}{16} \approx 0.0625$.
      vi. x=5: We have 11111, so we expect $\frac{1}{2^5} = \frac{1}{32} \approx 0.03125$.

(b) Complete the script *streaks.py* to estimate these probabilities using Monte Carlo simulation. Compare it to your answer in the previous question. The script will also apply your code to estimate the probability of streaks of heads with different lengths for 200 flips. Include your code in the answer as well as the figures generated by the script.

We first simulate the case where we flip the coin for 5 times, and we get an array [0.034, 0.366, 0.335, 0.168, 0.071, 0.026], representing the simulated probabilty that $x = i, i = 1, 2, 3, 4, 5$ happens.

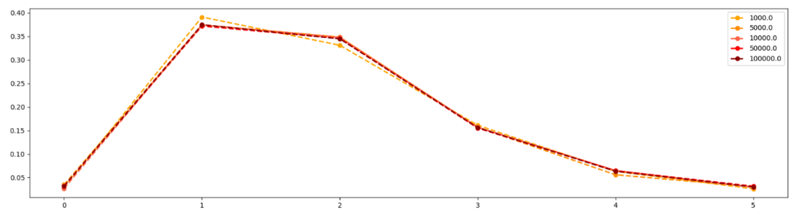

Figure 1: Probability Distribution for Coin Flips - Length of 5

We see that the simulation results are close to the precise computation. Codes on the next page for this question, be sure to check that.

```python
    # Python Codes
    import numpy as np
import matplotlib.pyplot as plt
plt.close("all")
np.random.seed(2017)

def p_longest_streak(n, tries):
    # Write your Monte Carlo code here, n is the length of the sequence and tries is the number
    # of sampled sequences used to produce the estimate of the probability
    # count the number of sequences with longest streaks of number 1,2,,...,n
    bucket = dict([(i, 0) for i in range(n+1)])
    for ind_trial in range(tries):
        seq = []
        for i in range(n):
            r = np.random.random()
            if r >= 0.5:
                seq.append(1)
            else:
                seq.append(0)
        longest_streaks_num = longest_streaks_number(seq)
        bucket[longest_streaks_num] += 1
    return list(map(lambda x: x / tries, bucket.values()))


def longest_streaks_number(seq):
    # Compute the longest streak in a seq list
    longest = 0
    current_streak = 0

    for flip in seq:
        if flip == 1:
            current_streak += 1
            longest = max(longest, current_streak)
        else:
            current_streak = 0

    return longest



n_tries = [1e3,5e3,1e4,5e4,1e5]

n_vals = [5,200]

color_array = ['orange','darkorange','tomato','red', 'darkred', 'tomato', 'purple', 'grey', 'deepskyblue',
               'maroon','darkgray','darkorange', 'steelblue', 'forestgreen', 'silver']
for ind_n in range(len(n_vals)):
    n = n_vals[ind_n]
    plt.figure(figsize=(20,5))
    for ind_tries in range(len(n_tries)):
        tries = n_tries[ind_tries]
        print("tries:␣" + str(tries))
        p_longest_tries = p_longest_streak(n, int(tries))
        plt.plot(range(n+1),p_longest_tries, marker='o',markersize=6,linestyle="dashed",lw=2,
                 color=color_array[ind_tries],
                 markeredgecolor= color_array[ind_tries],label=str(tries))
    plt.legend()

plt.show()

if __name__ == "__main__":
    print("The␣probability␣that␣the␣longest␣streak␣of␣ones␣in␣a␣Bernoulli
␣␣␣␣iid␣sequence␣of␣length␣200␣has␣length␣8␣or␣more␣is␣")
    print(p_longest_streak(200, 1000)) # Compute the probability and print it here
```

(c) What is the estimated probability that the longest streak of heads has length 8 or more for 200 flips? Is the sequence of 8 ones evidence that the program may not be generating truly random sequences?

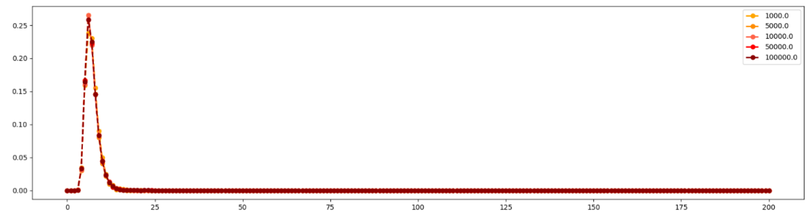

Figure 2: Probability Distribution for Coin Flips - Length of 200

From the simulation, we get the array [0.0, 0.0, 0.0, 0.001, 0.029, 0.159, 0.286, 0.225, 0.143, 0.084, 0.036, 0.015, 0.011, 0.007, 0.003, 0.0, 0.001, 0.0, 0.0, 0.0, 0.0, 0.0,.....] and the probability that the longest streak of heads has length 8 or more for 200 flips is 1-P(longest streak of heads with length less than 8)$=1-(0.001+0.029+0.159+0.286+0.225)=0.3$.
No, since in practive, even a truly random program will generate identical output consecutively, like 8 ones or eight 0's. At least based on our tiny experiment, we cannot conclude that.