DS313: Statistical Foundations of Data Science
Assignment No: 1

Course Instructor: Siddhartha Sarma *Date:* February 1, 2025

Submission deadline: 14 Feb 2025, 10 AM (Bring hardcopies to the classroom)

Related topics:

- **Sets, Sample space, Events, Probability axioms, σ-algebra, Conditional probability, Independence**

Notations:

- A probability space is denoted by a triplet $(\Omega, \mathcal{E}, \Pr)$.
- Ω denotes the sample space corresponding to a *random* experiment and $\emptyset$ represents an empty set.
- $\mathcal{A}, \mathcal{B}$ etc. denote sets/events. The complement of an event $\mathcal{A}$ is denoted by $\mathcal{A}^c$ or $\bar{\mathcal{A}}$.
- $\Pr(\mathcal{A}|\mathcal{B})$ represents conditional probability of event $\mathcal{A}$ given that $\mathcal{B}$ has occurred.

Instructions & Information:

- For the questions prefixed with [PA], you need to solve them on paper and write code to simulate the underlying experiment as well. You are free to choose the programming language for writing your code. As discussed in the last tutorial session, you should plot the normalised frequencies of the events under consideration as bar plots and compare them to their corresponding probabilities calculated analytically. Please refer to the course moodle page for examples.
- *The programming assignments should be submitted through moodle only. You also need to indicate in moodle quiz corresponding to the Assignment#1 the questions you have submitted in hard copy.*
- You must include a table similar to the one below on the front page of your submission to indicate the page numbers of the answer to the assignment questions.

Question no.	1	2	3	4	5	6	7	8	9	10
Page no.										

---

1. Let $\mathcal{A}\Delta\mathcal{B} := (\mathcal{A} \setminus \mathcal{B}) \cup (\mathcal{B} \setminus \mathcal{A})$. Show that (i) $\Pr(\mathcal{A}\Delta\mathcal{B}) = \Pr(\mathcal{A}) + \Pr(\mathcal{B}) - 2\Pr(\mathcal{A}\cap\mathcal{B})$, (ii) $|\Pr(\mathcal{A}) - \Pr(\mathcal{B})| \leq \Pr(\mathcal{A}\Delta\mathcal{B})$ and (iii) $\Pr(\mathcal{A}\Delta\mathcal{C}) \leq \Pr(\mathcal{A}\Delta\mathcal{B}) + \Pr(\mathcal{B}\Delta\mathcal{C})$. *(Since (ii) and (iii) are inequalities, you need to include an example in your answer for which equality holds.)*

2. Prove the following identity for events $\mathcal{A}, \mathcal{B}$ and $\mathcal{C}$ belonging to the same probability space

   $$\Pr(\mathcal{A}\cup\mathcal{B}\cup\mathcal{C}) = \Pr(\mathcal{A})+\Pr(\mathcal{B})+\Pr(\mathcal{C})-\Pr(\mathcal{A}\cap\mathcal{B})-\Pr(\mathcal{B}\cap\mathcal{C})-\Pr(\mathcal{A}\cap\mathcal{C})+\Pr(\mathcal{A}\cap\mathcal{B}\cap\mathcal{C})$$

   and generalize it for N events, $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$.

3. Show that if the events $\mathcal{A}_1, \ldots, \mathcal{A}_n$ from the same probability space are independent and $\mathcal{B}_i$ equals ${\mathcal{A}_i}^c$, then $\mathcal{B}_1, \ldots, \mathcal{B}_n$ are also independent.

4. If $\Omega = \{1, 2, 3, 4, 5\}$, then find the σ-algebra with the least number of elements that contains the sets $\{1\}$, $\{2, 3\}$.

5. If $(\Omega, \mathcal{E}, \Pr)$ is a probability space and $\mathcal{A} \subset \Omega$ such that $\Pr(\mathcal{A}) \neq 0$, show that $\Pr(.|\mathcal{A})$ is a valid probability measure.

6. [PA] An 8-bit register is loaded with binary digits, i.e., `1`s and `0`s, randomly. Describe the probability space for this experiment. Consider the following probability mappings: (a) the probability that ith digit is `0`, where $i = 1$ for the right-most digit, is proportional to 2^{-i} and is independent of the probability of other digits; (b) the probability of a digit located at an even(odd)-numbered position is `0` is $p_{ev}(p_{od})$, and is independent of other digits.

   Mathematically express the following events as a subset of Ω find their probabilities for both mappings mentioned above

   (a) $\mathcal{A}_1 \triangleq$ No two neighbouring digits are same.

   (b) $\mathcal{A}_2 \triangleq$ The register contains exactly four ones.

   (c) $\mathcal{A}_3 \triangleq$ Some cyclic shift of the register contents is equal to `01100110`.

   Also, calculate conditional probabilities, $\Pr(\mathcal{A}_1|\mathcal{A}_3), \Pr(\mathcal{A}_2|\mathcal{A}_3)$.

7. (a) Suppose that an event $\mathcal{A}$ is independent of itself. Show that either $\Pr(\mathcal{A}) = 0$ or $\Pr(\mathcal{A}) = 1$.

   (b) Events $\mathcal{A}$ and $\mathcal{B}$ have probabilities $\Pr(\mathcal{A}) = 0.3$ and $\Pr(\mathcal{B}) = 0.4$. What is $\Pr(\mathcal{A} \bigcup \mathcal{B})$ if $\mathcal{A}$ and $\mathcal{B}$ are independent? What is $\Pr(\mathcal{A} \bigcup \mathcal{B})$ if $\mathcal{A}$ and $\mathcal{B}$ are mutually exclusive (disjoint)?

   (c) Now suppose that $\Pr(\mathcal{A}) = 0.6$ and $\Pr(\mathcal{B}) = 0.8$. In this case, could the events $\mathcal{A}$ and $\mathcal{B}$ be independent? Could they be mutually exclusive?

8. Show that if $\{\mathcal{E}_i\}_{i\in I}$ is any collection of σ-algebra defined on the same Ω, then their intersection $\bigcap\limits_{i\in I} \mathcal{E}_i$ is also a field.

9. [PA] We have two coins—the first is fair and the second two-headed. We pick one of the coins at random, we toss it twice and heads showed both the times. Find the probability that the coin picked is fair.

10. [PA] A particular webserver may be working or not working. If the webserver is not working, any attempt to access it fails. Even if the webserver is working, an attempt to access it can fail due to network congestion beyond the control of the webserver. Suppose that the a priori probability that the server is working is 0.8. Suppose that if the server is working, then each access attempt is successful with probability 0.9, independently of other access attempts. Find the following quantities.

    (a) Pr(first access attempt fails)

    (b) Pr(server is working|first access attempt fails)

    (c) Pr(second access attempt fails|first access attempt fails)

    (d) Pr(server is working|first and second access attempts fail).

A few questions were adapted from the following references

1. Papoulis & Pillai, "Probability, Random Variables and Stochastic Processes", Fourth Ed, McGraw Hill
2. Capiński & Zastwniak, "Probability Through Problems", First Ed, Springer
3. Bruce Hajek, "Random Processes for Engineers", Cambridge University Press, 2015