AMIT GOYAL

PROBABILITY

Contents

List of Figures

List of Tables

1 Probability

Definition 1.1: Experiment

An experiment is an act whose outcome is not predictable with certainty. For example: Tossing a fair coin, selecting a student by drawing a name from the box, rolling a fair dice, drawing 5 cards from a well shuffled deck of 52 cards etc.

Definition 1.2: Sample Space

Consider an experiment. The set of all possible outcomes of an experiment is known as the sample space of the experiment and is denoted by S. For example - if the experiment consists of flipping two coins, then the sample space consists of four elements: $S = \{(H, H), (H, T), (T, H), (T, T)\}$.

Definition 1.3: Events

Any subset E of the sample space S is known as an event. Two ways to describe events:

1. In words
2. As sets

Definition 1.4: High-school definition of Probability

$$\Pr(E) = \frac{\text{Number of outcomes in } E}{\text{Number of outcomes in } S}$$

Note that this definition assumes all outcomes are equally likely and the sample space is a finite set.

Theorem 1.1: The basic principle of counting

Suppose an experiment has two stages. Then if stage 1 can result in any of m possible outcomes and if, for each outcome of stage 1, there are n possible outcomes of stage 2, then together there are mn possible outcomes of the experiment.

Theorem 1.2: Permutations

Suppose we have n objects. There are $n!$ different arrangements of objects possible. Each arrangement is known as a permutation. There are $(n-1)!$ circular permutations.

Theorem 1.3: Combinations

Suppose we have n objects. How many different groups of r objects can be formed from a total of n objects? We define $\binom{n}{r}$, for $r \leq n$, by $\binom{n}{r} = \frac{n!}{r!(n-r)!}$. We say that $\binom{n}{r}$ represents the number of possible combinations of n objects taken r at a time.

Theorem 1.4: Sampling: Choosing k objects out of n

1. Order Matters
   (a) With Replacement
   $$n^k$$
   (b) Without Replacement
   $$^nP_k = \frac{n!}{(n-k)!}$$
2. Order does not matter
   (a) With Replacement
   $$^{n+k-1}C_k = \binom{n+k-1}{k} = \frac{(n+k-1)!}{k!(n-1)!}$$
   (b) Without Replacement
   $$^nC_k = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Theorem 1.5

The following equalities hold:

1. $\binom{n}{k} = \binom{n}{n-k}$
2. $n\binom{n-1}{k-1} = k\binom{n}{k}$
3. $\binom{m+n}{k} = \sum_{j=0}^{k} \binom{m}{j}\binom{n}{k-j}$
4. $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$
5. $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$

Definition 1.5: Probability

Let $\mathcal{F}$ denotes the set of all events of an experiment. Probability is function $\Pr : \mathcal{F} \to [0,1]$ that satisfy the following axioms:

1. $\Pr(\emptyset) = 0, \Pr(S) = 1$
2. $\Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i)$ for a disjoint collection $\{A_i \subset S : i \in \mathbb{N}\}$ of events.

Theorem 1.6

1. $\Pr(A^c) = 1 - \Pr(A)$
2. If $A \subset B$, then $\Pr(A) \leq \Pr(B)$.
3. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$
4. $\Pr(A \cup B \cup C) = \Pr(A) + \Pr(B) + \Pr(C) - \Pr(A \cap B) - \Pr(A \cap C) - \Pr(B \cap C) + \Pr(A \cap B \cap C)$
5. $\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) + \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) - \sum_{i<j<k<l} \Pr(A_i \cap A_j \cap A_k \cap A_l) + \cdots + (-1)^{n+1} \Pr\left(\bigcap_{i=1}^{n} A_i\right)$

Definition 1.6: Independence of events

- Independence of two events. Events A and B are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$.
- Independence of three events. Events A, B and C are independent if $\Pr(A \cap B) = \Pr(A)\Pr(B)$, $\Pr(A \cap C) = \Pr(A)\Pr(C)$, $\Pr(B \cap C) = \Pr(B)\Pr(C)$ and $\Pr(A \cap B \cap C) = \Pr(A)\Pr(B)\Pr(C)$.
- Independence of n events. A collection $\{A_i \subset S : 1 \leq i \leq n\}$ of events are said to be independent if $\forall I \subset \{1, 2, \ldots n\}$, $\Pr\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \Pr(A_i)$

Solved Problems

Example 1.1: [Click]

Out of 5 men and 2 women, a committee of 3 is to be formed. In how many ways can it be formed if at least one woman is included in each committee?

Solution 1.1

$\binom{7}{3} - \binom{5}{3}$

Example 1.2: [Click]

An urn containing 5 red, 5 black and 10 white balls. If balls are drawn without replacement. What is the probability that in first 7 draws, at least one ball of each colour is drawn?

Solution 1.2

Let A_r be the event that there is at least one red ball drawn in the first seven balls. Likewise, A_b be the event that there is at least one black ball drawn in the first seven balls, and A_w be the event that there is at least one white ball in the seven draws. We want to find the probability of the event that at least one ball of each color is drawn in the seven draws which is
$\Pr(A_r \cap A_b \cap A_w) = 1 - \Pr(A_r^c \cup A_b^c \cup A_w^c)$.
So to find $\Pr(A_r \cap A_b \cap A_w)$, we just need to find $\Pr(A_r^c \cup A_b^c \cup A_w^c)$. By inclusion-exclusion principle,

$$\begin{aligned}
& \Pr(A_r^c \cup A_b^c \cup A_w^c) \\
= & \Pr(A_r^c) + \Pr(A_b^c) + \Pr(A_w^c) - \Pr(A_r^c \cap A_b^c) \\
& - \Pr(A_r^c \cap A_w^c) - \Pr(A_b^c \cap A_w^c) + \Pr(A_r^c \cap A_b^c \cap A_w^c) \\
= & \frac{\binom{15}{7}}{\binom{20}{7}} + \frac{\binom{15}{7}}{\binom{20}{7}} + \frac{\binom{10}{7}}{\binom{20}{7}} - \frac{\binom{10}{7}}{\binom{20}{7}} - 0 - 0 + 0 \\
= & \frac{\binom{15}{7}}{\binom{20}{7}} + \frac{\binom{15}{7}}{\binom{20}{7}} \\
= & 2 \times \frac{6435}{77520} \\
\approx & 0.166
\end{aligned}$$

Therefore, $\Pr(A_r \cap A_b \cap A_w) = \frac{64650}{77520} \approx 0.834$

Example 1.3: [Click]

The probability of a contractor getting a plumbing contract is $2/3$ and the probability of him getting an electricity contract is $5/9$. The probability of getting at least one contract is $4/5$. What's the probability that he he will get both contracts?

Solution 1.3

Let P be the event that the contractor gets the plumbing contract, E be the event that he gets the electricity contract. We are given
$\Pr(P) = \frac{2}{3}$, $\Pr(E) = \frac{5}{9}$ and $\Pr(P \cup E) = \frac{4}{5}$.
To find $\Pr(P \cap E)$, we will use the following equality
$\Pr(P \cap E) = \Pr(P) + \Pr(E) - \Pr(P \cup E) = \frac{2}{3} + \frac{5}{9} - \frac{4}{5} = \frac{19}{45}$

Example 1.4: [Click]

Five players are dealt 3 cards each. What is the probability of a player getting 3 aces?

Solution 1.4

Let E_j denotes the event that player j gets three aces, where $j \in \{1,2,3,4,5\}$. We want to find the probability $\Pr(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5)$. Since E_1, E_2, E_3, E_4, E_5 are mutually disjoint,
$\Pr(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5) = \Pr(E_1) + \Pr(E_2) + \Pr(E_3) + \Pr(E_4) + \Pr(E_5)$
By symmetry, $\Pr(E_1) = \Pr(E_2) = \Pr(E_3) = \Pr(E_4) = \Pr(E_5)$ holds.
Therefore,

$$\Pr(E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5) = 5\Pr(E_1) = 5 \times \frac{\binom{4}{3}}{\binom{52}{3}} = \frac{1}{1105}$$

Example 1.5: [Click]

What is the probability that no two have the same face value in a poker hand of 5 cards?

Solution 1.5

$$\frac{\binom{13}{5}4^5}{\binom{52}{5}} \approx 0.50708$$

Example 1.6: [Click]

How many ways can 12 boys and 14 girls be arranged in a line?

Solution 1.6

26!

Example 1.7: [Click]

10 balls are thrown into 5 bins uniformly, at random, and independently, what is the probability that there are no empty bins?

Solution 1.7

Let the five bins be numbered 1, 2, 3, 4, and 5, respectively, and let A_i be the event that bin i is not empty. We want to find the probability of the event that no bin is empty i.e. $A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5$. Now $\Pr(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) = 1 - \Pr(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c \cup A_5^c)$ and by inclusion-exclusion principle,

$$\begin{aligned}
& \Pr(A_1^c \cup A_2^c \cup A_3^c \cup A_4^c \cup A_5^c) \\
= & \binom{5}{1} \Pr(A_1^c) - \binom{5}{2} \Pr(A_1^c \cap A_2^c) + \binom{5}{3} \Pr(A_1^c \cap A_2^c \cap A_3^c) \\
& - \binom{5}{4} \Pr(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c) \\
& + \binom{5}{5} \Pr(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c \cap A_5^c) \\
= & \binom{5}{1} \left(\frac{4}{5}\right)^{10} - \binom{5}{2} \left(\frac{3}{5}\right)^{10} + \binom{5}{3} \left(\frac{2}{5}\right)^{10} - \binom{5}{4} \left(\frac{1}{5}\right)^{10} \\
& + \binom{5}{5} \left(\frac{0}{5}\right)^{10} \\
\approx & \; 0.477
\end{aligned}$$

Therefore, $\Pr(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) \approx 1 - 0.477 = 0.523$

Example 1.8: [Click]

In how many ways can you put 15 identical balls in 3 distinct boxes such that each box contains at least 1 and at most 10 balls?

Solution 1.8

Notice that the number of ways to put 15 identical balls in 3 distinct boxes such that each box contains at least 1 and at most 10 balls is equal to the number of positive integer solutions to the following system of equations/inequalities:

$$\begin{aligned} x_1 + x_2 + x_3 &= 15 \\ 1 \le x_1 &\le 10 \\ 1 \le x_2 &\le 10 \\ 1 \le x_3 &\le 10 \end{aligned}$$

Equivalently, we can define $y_1 = x_1 - 1$, $y_2 = x_2 - 1$ and $y_3 = x_3 - 1$, rewrite the system as

$$\begin{aligned} y_1 + y_2 + y_3 &= 12 \\ 0 \le y_1 &\le 9 \\ 0 \le y_2 &\le 9 \\ 0 \le y_3 &\le 9 \end{aligned}$$

and find the number of non-negative integer solutions to this system. Doing so gives us: $\binom{14}{12} - 3 \times \binom{4}{2}$ where $\binom{14}{12}$ is the number of non-negative integer solutions to $y_1 + y_2 + y_3 = 12$ and $3 \times \binom{4}{2}$ is the number of non-negative integer solutions to $y_1 + y_2 + y_3 = 12$ with the property that either y_1 or y_2 or y_3 is greater than or equal to 10.
Therefore, the number of ways to put 15 identical balls in 3 distinct boxes such that each box contains at least 1 and at most 10 balls is equal to $\binom{14}{12} - 3 \times \binom{4}{2} = 73$.

Example 1.9: [Click]

Does it imply that the two not independent events are mutually exclusive events?

Solution 1.9

No. Consider any event A with the property that $0 < \Pr(A) < 1$. A is neither independent of itself nor A and A are mutually exclusive. However, two mutually exclusive events A and B with the property that $0 < \Pr(A) \le 1$ and $0 < \Pr(B) \le 1$ can't be independent. This is because $\Pr(A \cap B) = 0 \neq \Pr(A)\Pr(B)$.

Example 1.10: [Click]

A drawer has 5 brown socks and 4 green socks. A man takes out 2 socks at random. What is the probability that they match?

Solution 1.10

$$\frac{\binom{5}{2} + \binom{4}{2}}{\binom{9}{2}}.$$

Example 1.11: [Click]

Tickets numbered 1 to 10 are mixed up and two tickets are drawn at random. What is the probability that they are multiples of 3?

Solution 1.11

There are $\binom{10}{2} = 45$ ways to select two tickets from the set of tickets numbered 1 to 10. There are $\binom{3}{2} = 3$ ways to select two tickets from the set of tickets that are numbered multiples of 3. So the required probability is $\frac{3}{45} = \frac{1}{15}$.

Example 1.12: [Click]

If $\Pr(A) \geq 0.8$ and $\Pr(B) \geq 0.8$, then $\Pr(A \cap B) \geq l$. Find the largest value of l for which the above implication is true.

Solution 1.12

$$\begin{aligned}\Pr(A \cap B) &= \Pr(B) - \Pr(A^c \cap B) \\ &\geq \Pr(B) - \Pr(A^c) \\ &= 0.8 - 0.2 \\ &= 0.6\end{aligned}$$

Therefore, $\Pr(A \cap B) \geq 0.6$.

2 | Conditional Probability

Definition 2.1: Conditional Probability

Conditional Probability of event A given that event B has occurred is defined as:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \quad \text{if } \Pr(B) > 0$$

Theorem 2.1

1. $\Pr(A \cap B) = \Pr(B)\Pr(A|B) = \Pr(A)\Pr(B|A)$
2. $\Pr\left(\bigcap_{i=1}^{n} A_i\right) = \Pr(A_1)\Pr(A_2|A_1)\Pr(A_3|A_1 \cap A_2)\cdots\Pr(A_n|A_1 \cap A_2 \cap A_3 \cdots \cap A_{n-1})$
3. (Bayes' Rule) $\Pr(A|B) = \dfrac{\Pr(B|A)\Pr(A)}{\Pr(B)}$
4. (Law of total Probability) Given a partition $A_1, A_2, A_3, \ldots, A_n$ of S,

$$\Pr(E) = \sum_{i=1}^{n} \Pr(E \cap A_i) = \sum_{i=1}^{n} \Pr(E|A_i)\Pr(A_i)$$

5. (Bayes' Rule) $\Pr(A|B) = \dfrac{\Pr(B|A)\Pr(A)}{\Pr(B|A)\Pr(A) + \Pr(B|A^c)\Pr(A^c)}$

Definition 2.2: Conditional independence

We say events A and B are conditionally independent given C if

$$\Pr(A \cap B|C) = \Pr(A|C)\Pr(B|C)$$

Solved Problems

Example 2.1: [Click]

A bag contains 10 white and 3 red balls while another bag contains 3 white and 5 red balls. Two balls are drawn at random and put in the second bag. Then a ball is drawn at random from the second bag, what is the probability that it is a white ball?

Solution 2.1

Let E_{WW} be the event that two balls drawn from the first bag are both white. Likewise, E_{WR} be the event that one white ball and one red ball are drawn from the first bag, and E_{RR} be the event that two balls drawn from the first bag are both red. Let W be the event that a white ball is drawn from the second bag. By the Law of total probability,

$$\begin{aligned}
\Pr(W) &= \Pr(E_{WW})\Pr(W|E_{WW}) + \Pr(E_{WR})\Pr(W|E_{WR}) \\
&\quad + \Pr(E_{RR})\Pr(W|E_{RR}) \\
&= \frac{\binom{10}{2}}{\binom{13}{2}} \times \frac{5}{10} + \frac{\binom{10}{1}\binom{3}{1}}{\binom{13}{2}} \times \frac{4}{10} + \frac{\binom{3}{2}}{\binom{13}{2}} \times \frac{3}{10} \\
&= \frac{45}{78} \times \frac{5}{10} + \frac{30}{78} \times \frac{4}{10} + \frac{3}{78} \times \frac{3}{10} \\
&= \frac{354}{780} \\
&= \frac{59}{130} \\
&\approx 0.4538
\end{aligned}$$

Example 2.2: [Click]

Three dice are thrown simultaneously. What is the probability that 4 has appeared on two dice given that 5 has occurred on one dice?

Solution 2.2

Three dice are thrown simultaneously. Let A be the event that 4 appears on two dice, and B be the event that 5 occurs on (exactly) one dice. We want to find the conditional probability of A given B.

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} = \frac{\binom{3}{1}\frac{1}{6^3}}{\binom{3}{1}\frac{5^2}{6^3}} = \frac{1}{25}$$

Example 2.3: [Click]

A company produces light bulbs at three factories A, B and C.

- Factory A produces 40% of the total number of bulbs, of which 2% are defective.
- Factory B produces 35% of the total number of bulbs, of which 4% are defective.
- Factory C produces 25% of the total number of bulbs, of which 3% are defective.

1. A defective bulb is found among the total output. Find the probability that it came from (a) Factory A, (a) Factory B, (c) Factory C.
2. Now suppose a factory is chosen at random, and one of its bulbs is randomly selected. If the bulb is defective, find the probability that it came from (a) Factory A, (a) Factory B, (c) Factory C.

Solution 2.3

1. Let D be the event that the randomly selected bulb is defective. Let A be the event that the randomly selected bulb came from factory A. Likewise, define events B and C. We are given that $\Pr(A) = 0.4$, $\Pr(B) = 0.35$ and $\Pr(C) = 0.25$. Also, $\Pr(D|A) = 0.02$, $\Pr(D|B) = 0.04$, and $\Pr(D|C) = 0.03$. We want to find $\Pr(A|D)$. By Bayes' Rule,

$$\begin{aligned}\Pr(A|D) &= \frac{\Pr(D|A)\Pr(A)}{\Pr(D|A)\Pr(A) + \Pr(D|B)\Pr(B) + \Pr(D|C)\Pr(C)} \\ &= \frac{0.008}{0.008 + 0.014 + 0.0075} \\ &= 0.2712\end{aligned}$$

   Likewise, $\Pr(B|D) = 0.4746$ and $\Pr(C|D) = 0.2542$.

2. In this question we are given that $\Pr(A) = \frac{1}{3}$, $\Pr(B) = \frac{1}{3}$ and $\Pr(C) = \frac{1}{3}$. Also, $\Pr(D|A) = 0.02$, $\Pr(D|B) = 0.04$, and $\Pr(D|C) = 0.03$. To find $\Pr(A|D)$, we use Bayes' Rule (as before), and we get

$$\begin{aligned}\Pr(A|D) &= \frac{\Pr(D|A)\Pr(A)}{\Pr(D|A)\Pr(A) + \Pr(D|B)\Pr(B) + \Pr(D|C)\Pr(C)} \\ &= \frac{2}{9}\end{aligned}$$

   Similarly, $\Pr(B|D) = \frac{4}{9}$ and $\Pr(C|D) = \frac{1}{3}$.

3 | Discrete Random Variables

Definition 3.1: Random Variable

A Random Variable is any real-valued function defined on sample space $X : S \to \mathbb{R}$.

Definition 3.2: Discrete Random Variable

A discrete random variable is the random variable that can take a finite or countably infinite number of values.

Definition 3.3: Probability Mass Function (PMF)

Consider a discrete random variable $X : S \to \mathbb{R}$. Associated with it is a probability mass function (PMF) $p_X : \mathbb{R} \to [0, 1]$ defined as follows:
$p_X(x) = \Pr(\{s \in S | X(s) = x\}) = \Pr(X = x)$

Definition 3.4: Cumulative Distribution Function (CDF)

The cumulative distribution function (CDF) of a random variable X is a function $F_X : \mathbb{R} \to [0, 1]$ defined as follows:
$F_X(x) = \Pr(\{s \in S | X(s) \leq x\}) = \Pr(X \leq x)$

Theorem 3.1: Properties of a CDF

1. F_X is monotonically non-decreasing.
2. $\lim_{x \to -\infty} F_X(x) = 0$ and $\lim_{x \to \infty} F_X(x) = 1$
3. F_X is right-continuous.

Definition 3.5: Bernoulli Random Variable

$X \sim \text{Bern}(p)$ i.e. X is a Bernoulli random variable with parameter p if it indicates whether a trial that results in a success with probability p is a success or not.

$$\begin{aligned} p_X(1) = \Pr(X = 1) = & \quad p \\ p_X(0) = \Pr(X = 0) = & \quad 1 - p \end{aligned}$$

Definition 3.6: Binomial Random Variable

$X \sim \text{Bin}(n, p)$ is a Binomial random variable with parameters n and p if it represents the number of successes in n independent trials when each trial is a success with probability p.

$$p_X(x) = \Pr(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

where $x \in \{0, 1, \ldots, n\}$.

Definition 3.7: Geometric Random Variable

$X \sim \text{Geom}(p)$ is a Geometric random variable with parameter p if it represents the number of failures before the first success where each trial is independently a success with probability p.

$$p_X(x) = \Pr(X = x) = p(1 - p)^x$$

where $x \in \{0, 1, 2, 3 \ldots\}$.

Definition 3.8: Poisson Random Variable.

$X \sim \text{Pois}(\lambda)$ is used to model the number of events that occur when these events are either independent or weakly dependent and each has a small probability of occurrence. It has parameter λ that represents the rate of occurrence.

$$p_X(x) = \Pr(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$$

where $x \in \{0, 1, 2, 3 \ldots\}$.

Definition 3.9: Indicator Random Variable

Let A be any event. Indicator Random Variable, $\mathbb{1}_A : S \to \mathbb{R}$ is a random variable that assigns value 1 to those outcomes when event A occurs, and 0 otherwise.

$$\mathbb{1}_A(s) = \begin{cases} 1 & \text{if } s \in A \\ 0 & \text{if } s \notin A \end{cases}$$

Definition 3.10: Expected Value or Expectation

Expected value of a random variable X, with PMF p_X, is defined by

$$\mathbb{E}(X) = \sum_{x \in X(S)} x p_X(x) = \sum_{s \in S} X(s) \Pr(\{s\})$$

Here $X(S)$ denotes the range of X.

Theorem 3.2

1. For $X \sim \text{Bern}(p)$, $\mathbb{E}(X) = p$
2. For $X \sim \text{Bin}(n, p)$, $\mathbb{E}(X) = np$
3. For $X \sim \text{Geom}(p)$, $\mathbb{E}(X) = \dfrac{1-p}{p}$
4. For $X \sim \text{Pois}(\lambda)$, $\mathbb{E}(X) = \lambda$

Theorem 3.3

Consider $X_n \sim \text{Bin}(n, p_n)$ and $X \sim \text{Pois}(\lambda)$ where $\lambda = np_n$ for all n. If $n \to \infty$ and $p_n \to 0$, then $X_n \to X$ in distribution i.e. $p_{X_n}(x) \to p_X(x)$ for all $x \in \mathbb{Z}_+$.

Theorem 3.4: Linearity of Expectation

Given two random variables $X : S \to \mathbb{R}$ and $Y : S \to \mathbb{R}$, we have

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

and for any $c \in \mathbb{R}$ we also have

$$\mathbb{E}(cX) = c\mathbb{E}(X)$$

Definition 3.11: Transformation of a Discrete Random Variable

Given a random variable $X : S \to \mathbb{R}$, one may generate other random variables by applying various transformations on X. $Y : S \to \mathbb{R}$ is the transformation of a random variable $X : S \to \mathbb{R}$ if $Y(s) = g \circ X(s) = g(X(s))$ for some function $g : \mathbb{R} \to \mathbb{R}$.

Theorem 3.5: PMF of a Transformed Random Variable

Suppose $Y = g(X)$ is a transformation of random variable X using function g. The PMF of Y, p_Y can be calculated using the PMF of X. In particular, to obtain $p_Y(y)$ for any y, we add the probabilities of all values of x such that $g(x) = y$:

$$p_Y(y) = \sum_{\{x:g(x)=y\}} p_X(x)$$

Theorem 3.6: Law of the Unconscious Statistician

Law of the Unconscious Statistician (LOTUS) is

$$\mathbb{E}(g(X)) = \sum_{x \in X(S)} g(x) p_X(x)$$

Example: Find Expected Utility

Definition 3.12: Variance of a Random Variable

Variance of a random variable X, denoted by $\mathbb{V}(X)$, is

$$\mathbb{V}(X) = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$$

Definition 3.13: Standard Deviation of a Random Variable

$\sqrt{\mathbb{V}(X)}$ is known as the standard deviation of X and is denoted by σ_X.

Theorem 3.7

1. For $X \sim \text{Bern}(p)$, $\mathbb{V}(X) = p(1-p)$
2. For $X \sim \text{Bin}(n, p)$, $\mathbb{V}(X) = np(1-p)$
3. For $X \sim \text{Geom}(p)$, $\mathbb{V}(X) = \dfrac{1-p}{p^2}$
4. For $X \sim \text{Pois}(\lambda)$, $\mathbb{V}(X) = \lambda$

Definition 3.14: Negative Binomial Random Variable

$X \sim \text{NBin}(r, p)$ if it is the number of failures before the rth success where each trial is independently a success with probability p.

$$p_X(x) = \Pr(X = x) = \binom{r + x - 1}{r - 1} p^r (1 - p)^x$$

where $x \in \{0, 1, 2, 3 \ldots\}$

Definition 3.15: Hypergeometric Random Variable

$X \sim \text{Hyper}(N, n, m)$ if it is the number of white balls in a random sample of n balls chosen without replacement from an urn of N balls of which m are white.

$$p_X(x) = \Pr(X = x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$$

where $x \in \{\max(0, n + m - N), 1, 2, \ldots, \min\{m, n\}\}$.

Definition 3.16: Discrete Uniform Random Variable

$X \sim \text{DUnif}(a, a + n)$ if

$$p_X(x) = \Pr(X = x) = \frac{1}{n + 1}$$

where $x \in \{a, a + 1, a + 2, \ldots, a + n\}$.

Definition 3.17: Joint Probability Mass Function

Consider two discrete random variables X and Y associated with the same experiment. The probabilities of the values that X and Y can take are captured by the joint PMF of X and Y, denoted $p_{X,Y}$. In particular, if (x, y) is a pair of possible values of X and Y, the probability mass of (x, y) is the probability of the event $X = x, Y = y$:

$$\begin{aligned} p_{X,Y}(x, y) &= \Pr\left(\{s \in S | X(s) = x\} \cap \{s \in S | Y(s) = y\}\right) \\ &= \Pr(X = x \cap Y = y) \\ \text{or simply,} \quad & \Pr(X = x, Y = y) \end{aligned}$$

Definition 3.18: Marginal PMF of X

Given the joint PMFs of X and Y ($p_{X,Y}$), we can calculate the PMF of X by using

$$p_X(x) = \sum_{y \in Y(S)} p_{X,Y}(x, y)$$

We also refer to p_X as the marginal PMF of X.

Definition 3.19: Conditional PMF of X given event A

The conditional PMF of a random variable X, conditioned on a particular event A with $\Pr(A) > 0$, is defined by

$$p_{X|A}(x) = \Pr(X = x|A) = \frac{\Pr(X = x \cap A)}{\Pr(A)}$$

Definition 3.20: Conditional PMF of X given Y

Conditional PMF of X at x given $Y = y$ with $p_Y(y) > 0$, denoted by $p_{X|Y}(x|y)$, is defined as:

$$\begin{aligned} p_{X|Y}(x|y) &= \Pr(X = x|Y = y) \\ &= \frac{\Pr(X = x, Y = y)}{\Pr(Y = y)} \\ \text{or simply} \quad & \frac{p_{X,Y}(x, y)}{p_Y(y)} \end{aligned}$$

Definition 3.21: Independence of Random Variables

We say two random variable X and Y are independent if

$$p_{X,Y}(x, y) = p_X(x) p_Y(y)$$

for all x, y.

Definition 3.22: 2-D LOTUS (Law of the Unconscious Statistician)

For any $g : \mathbb{R}^2 \to \mathbb{R}$, and random variables X and Y,

$$\mathbb{E}(g(X, Y)) = \sum_{x \in X(S)} \sum_{y \in Y(S)} g(x, y) p_{X,Y}(x, y)$$

Definition 3.23: Covariance between X and Y

Covariance between random variables X and Y, denoted by $\mathbb{C}(X,Y)$, is defined as follows:

$$\begin{aligned}\mathbb{C}(X,Y) &= \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - (\mathbb{E}(X)\mathbb{E}(Y))\end{aligned}$$

Theorem 3.8

Given $n + m$ random variables $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$, and $n + m$ real numbers $a_1, \ldots, a_n$ and $b_1, \ldots, b_m$, the following holds

$$\mathbb{C}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_j Y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \mathbb{C}(X_i, Y_j)$$

Theorem 3.9

1. For $X \sim \text{Hyper}(N,n,m)$, $\mathbb{E}(X) = \frac{nm}{N}$, and $\mathbb{V}(X) = \frac{nm(N-m)(N-n)}{N^2(N-1)}$.
2. For $X \sim \text{NBin}(r,p)$, $\mathbb{E}(X) = \frac{r(1-p)}{p}$, and $\mathbb{V}(X) = \frac{r(1-p)}{p^2}$.

Definition 3.24: Correlation between X and Y

Correlation between random variables X and Y, denoted by $\rho_{X,Y}$, is defined as follows:

$$\rho_{X,Y} = \mathbb{C}\left(\frac{X-\mathbb{E}(X)}{\sqrt{\mathbb{V}(X)}}, \frac{Y-\mathbb{E}(Y)}{\sqrt{\mathbb{V}(Y)}}\right) = \frac{\mathbb{C}(X,Y)}{\sigma_X \sigma_Y}$$

Theorem 3.10

For any pair of random variables X and Y, the following holds:

$$-1 \le \rho(X,Y) \le 1$$

Theorem 3.11

If X and Y are independent, they are uncorrelated i.e. $\mathbb{C}(X,Y) = 0$.

Definition 3.25: Conditional Expectation of X given $Y = y$

Conditional Expectation of X given an event $Y = y$, is denoted by $\mathbb{E}(X|Y = y)$, and is defined as follows:

$$\mathbb{E}(X|Y = y) = \sum_{x \in X(S)} x p_{X|Y}(x|y)$$

Note[1]

[1] $\mathbb{E}(X|Y)$ is a random variable

Theorem 3.12

$\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$

Theorem 3.13

$\mathbb{V}(X) = \mathbb{V}(\mathbb{E}(X|Y)) + \mathbb{E}(\mathbb{V}(X|Y))$

Definition 3.26: Median of a Random Variable

We say that m_X is the median of random variable X if

$$\Pr(X \leq m_X) \geq \frac{1}{2}$$
$$\Pr(X \geq m_X) \geq \frac{1}{2}$$

Solved Problems

Example 3.1: [Click]

Assume that X is uniformly distributed on $\{1, 2, \ldots, n\}$ and Y is uniformly distributed on $\{2, 4, \ldots, 2n\}$. Assuming that X and Y are independent random variables, what is the variance of XY?

Solution 3.1

$$\begin{aligned}
\mathbb{V}(XY) &= \mathbb{E}(X^2Y^2) - (\mathbb{E}(XY))^2 \\
&= \mathbb{E}(X^2)\mathbb{E}(Y^2) - (\mathbb{E}(X)\mathbb{E}(Y))^2 \\
&= 4\mathbb{E}(X^2)\mathbb{E}(X^2) - 4(\mathbb{E}(X))^4 \\
&= 4\left[\left(\frac{(n+1)(2n+1)}{6}\right)^2 - \left(\frac{n+1}{2}\right)^4\right]
\end{aligned}$$

Example 3.2: [Click]

Let X and Y be the random variables which respectively are the number of tosses to see your first head and the number of tosses to see your first tail. What is the covariance between X and Y, $\mathrm{Cov}(X,Y)$?

Solution 3.2

X and Y are Geometric (p) Random Variables with parameter $p = 1/2$. So we have $\mathbb{E}(X) = \mathbb{E}(Y) = \frac{1}{p} = 2$. To find the $\mathrm{Cov}(X,Y)$, we'll now find $\mathbb{E}(XY)$. Observe that

$$XY = \begin{cases} Y & \text{if } X = 1 \\ X & \text{if } X > 1 \end{cases}$$

$$\begin{aligned}
\mathbb{E}(XY) &= \mathbb{E}(XY|X=1)\Pr(X=1) + \mathbb{E}(XY|X>1)\Pr(X>1) \\
&= \mathbb{E}(Y|X=1)\Pr(X=1) + \mathbb{E}(X|X>1)\Pr(X>1) \\
&= (1+2)0.5 + (1+2)0.5 \\
&= 3
\end{aligned}$$

Therefore, $\mathrm{Cov}(X,Y) = 3 - 4 = -1$

Example 3.3: [Click]

Can the $\mathrm{Cov}(X, X+Y) = 0$?

Solution 3.3

Yes it is possible. Consider any random variable X and let $Y = -X$.

Example 3.4: [Click]

For n independent Bernoulli trials, each with the probability p of success, let X be the number of successes so that $X \sim \mathrm{Binom}(n,p)$ and let Y be the number of failures i.e. $Y = n - X \sim \mathrm{Binom}(n, 1-p)$. Find the expected value $\mathbb{E}(XY)$ and the covariance $\mathrm{Cov}(X,Y)$?

Solution 3.4

Given that $X \sim \mathrm{B}(n,p)$ and $Y = n - X \sim \mathrm{B}(n, 1-p)$, the correlation coefficient between X and Y (denoted by $\rho_{X,Y}$) equals -1. Consequently, the covariance is
$\mathrm{Cov}(X,Y) = \rho_{X,Y}\sqrt{\mathbb{V}_X\mathbb{V}_Y} = -np(1-p)$ and
$\mathbb{E}(XY) = \mathrm{Cov}(X,Y) + \mathbb{E}(X)\mathbb{E}(Y) = -np(1-p) + n^2p(1-p) = n(n-1)p(1-p)$

Example 3.5: [Click]

Show that the distribution of a random variable X with possible values $0, 1, 2$ is determined by $\mu_1 = \mathbb{E}(X)$ and $\mu_2 = \mathbb{E}(X^2)$?

Solution 3.5

Given that $\mu_1 = \mathbb{E}(X)$ and $\mu_2 = \mathbb{E}(X^2)$, and X take values in the set $\{0, 1, 2\}$, we can write

$$\begin{aligned}
0 \cdot \Pr(X=0) + 1 \cdot \Pr(X=1) + 2 \cdot \Pr(X=2) &= \mu_1 \\
0 \cdot \Pr(X=0) + 1 \cdot \Pr(X=1) + 4 \cdot \Pr(X=2) &= \mu_2 \\
\Pr(X=0) + \Pr(X=1) + \Pr(X=2) &= 1
\end{aligned}$$

Solving the above system of equations, we get
$\Pr(X=2) = \dfrac{\mu_2 - \mu_1}{2}$
$\Pr(X=1) = 2\mu_1 - \mu_2$
$\Pr(X=0) = 1 - \left(\dfrac{3\mu_1 - \mu_2}{2}\right)$

Example 3.6: [Click]

$X \sim \mathrm{Pois}(2)$, $Y = \min(X, 10)$. What is the probability distribution of Y?

Solution 3.6

If F_X and F_Y denote the CDFs of X and Y respectively, and $Y = \min(X, 10)$, we can use CDF of X to get CDF of Y in the following way : $F_Y(t) = \begin{cases} F_X(t) & \text{for } t < 10 \\ 1 & \text{for } t \geq 10 \end{cases}$

Example 3.7: [Click]

Suppose the occurrence of A makes it more likely that B will occur. In that case, show that the occurrence of B makes it more likely that A will occur i.e. show that if $\Pr(B|A) > \Pr(B)$, then it is also true that $\Pr(A|B) > \Pr(A)$.

Solution 3.7

Just rewrite the equality
$\Pr(A \cap B) = \Pr(A)\Pr(B|A) = \Pr(B)\Pr(A|B)$ as
$\dfrac{\Pr(B|A)}{\Pr(B)} = \dfrac{\Pr(A|B)}{\Pr(A)}$
and the result follows.

Example 3.8: [Click]

6 independent fair coins are tossed in a row. What is the expected number of consecutive HH pairs?

Solution 3.8

For $j \in \{2, 3, 4, 5, 6\}$, define

$$I_j = \begin{cases} 1 & \text{if outcomes of } (j-1)\text{th and } j\text{th tosses are heads,} \\ 0 & \text{otherwise} \end{cases}$$

Note that $\mathbb{E}(I_j) = \Pr(I_j = 1) = \frac{1}{4}$.
Let N denotes the number of pairs of heads. We can write N as $N = I_2 + I_3 + I_4 + I_5 + I_6$.
Therefore,
$\mathbb{E}(N) = \mathbb{E}(I_2) + \mathbb{E}(I_3) + \mathbb{E}(I_4) + \mathbb{E}(I_5) + \mathbb{E}(I_6)$
By symmetry,
$\mathbb{E}(N) = 5\mathbb{E}(I_2) = \frac{5}{4} = 1.25$

Example 3.9: [Click]

A coin is weighted so that the probability of obtaining a head in a single toss is 0.3. If the coin is tossed 35 times, then what is the probability of obtaining between 9 and 14 heads exclusively?

Solution 3.9

Here $X \sim \text{Bin}(35, 0.3)$, consequently probability of the event that $a \leq X \leq b$ is
$\Pr(a \leq X \leq b) = \sum_{j=a}^{b} \binom{35}{j} 0.3^j 0.7^{35-j}$
where a and b are integers satisfying $0 \leq a \leq b \leq 35$.

Example 3.10: [Click]

A sample of 4 items are selected randomly from a box containing 12 items of which 5 are defective, find the expected number of defective items.

Solution 3.10

Let I_j be the indicator random variable that takes value 1 if jth item in the sample (of size 4) is defective, and 0 otherwise. Observe that the number of defective items is equal to $I_1 + I_2 + I_3 + I_4$. By linearity of expectation, expected number of defective items is equal to $\mathbb{E}(I_1) + \mathbb{E}(I_2) + \mathbb{E}(I_3) + \mathbb{E}(I_4)$. By symmetry, it is equal to $4\mathbb{E}(I_1)$. Now, expected value of I_1 is equal to the probability that the first item is defective i.e. $\mathbb{E}(I_1) = \Pr(I_1 = 1) = \frac{5}{12}$. Therefore, expected number of defective items is equal to $4 \times \frac{5}{12} = \frac{5}{3}$.

Example 3.11: [Click]

If X and Y are uncorrelated random variables with variance 1, then what is the variance of $X - Y$?

Solution 3.11

$\mathbb{V}(X - Y) = \mathbb{V}(X) + \mathbb{V}(Y) - 2\mathbb{C}(X, Y) = 1 + 1 - 2(0) = 2$

Example 3.12: [Click]

Four identical objects are distributed randomly into 3 distinct boxes. Let X denote the number of objects that end up in the first box. What is the expected value of X?

Solution 3.12

Given that
X = number of objects that end up in the first box
Also let
Y = number of objects that end up in the second box
Z = number of objects that end up in the third box
Since four identical objects are distributed randomly into these 3 distinct boxes, we have
$X + Y + Z = 4$
Therefore,
$\mathbb{E}(X) + \mathbb{E}(Y) + \mathbb{E}(Z) = 4$
By symmetry, $\mathbb{E}(X) = \mathbb{E}(Y) = \mathbb{E}(Z)$
So we get $\mathbb{E}(X) = \frac{4}{3}$.

Example 3.13: [Click]

Let X_1, X_2, X_3 be the numbers on the cards drawn from a pack containing 9 cards numbered $5, 6, 7, 8, 9, 10, 11, 12$ and 13, respectively.

Solution 3.13

Define $X = \max(X_1, X_2, X_3)$, here is the probability mass function of X: $p_X(k) = \frac{\binom{k-5}{2}}{\binom{9}{3}}$ for $k = 7, 8, 9, 10, 11, 12, 13$

Example 3.14: [Click]

There are 10 items in a box, 6 of which are defective. 4 items are selected randomly without replacement. What is the expected number of selected defective items?

Solution 3.14

Let X_i be the indicator random variable that takes value 1 if the ith item is defective and 0 otherwise, where $i \in \{1, 2, 3, 4\}$. Note that X_1, X_2, X_3 and X_4 are not independent, but they have the same distribution:
$\Pr(X_i = 1) = \frac{6}{10} = \frac{3}{5}$
$\Pr(X_i = 0) = 1 - \frac{6}{10} = \frac{2}{5}$
for each $i \in \{1, 2, 3, 4\}$.
This follows from the symmetry of positions. Item on the second position is just as likely to be a defective item as the one on the first position, and same holds for other positions. Therefore, expected value of X_i is equal to the probability that the ith item is defective i.e. $\frac{6}{10}$. If X denotes the number of defective items selected, then we can write X as $X = X_1 + X_2 + X_3 + X_4$.
By linearity of expectation, it follows that
$\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) + \mathbb{E}(X_4) = \frac{12}{5} = 2.4$.

Example 3.15: [Click]

What is the expected number of times that heads will appear when a fair coin is tossed three times?

Solution 3.15

Let X_i be the indicator random variable that takes value 1 if heads appear on the ith toss and 0 otherwise. Therefore, expected value of X_i is equal to the probability that heads appears on the ith toss i.e. $\frac{1}{2}$. If X denotes the number of times that heads will appear when a fair coin is tossed three times, then we can write X as $X = X_1 + X_2 + X_3$. By linearity of expectation, it follows that $\mathbb{E}(X) = \mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3) = \frac{3}{2}$.

Example 3.16: [Click]

If 5 men and 5 women are seated randomly in a single row of chairs. What is the expected number of women sitting next to at least 1 man?

Solution 3.16

Let the seats be numbered $1, 2, 3, \ldots, 10$ in order of arrangement. Define a random variable I_1 in the following way: I_1 takes value 1 if seat number 1 is occupied by a woman and seat number 2 is occupied by a man, and 0 otherwise. Likewise, we define random variables I_j for $2 \leq j \leq 9$ in the similar way: I_j takes value 1 if seat number j is occupied by a woman and at least one of the two seats numbered $(j-1)$ and $(j+1)$ is occupied by a man, and takes value 0 otherwise. I_{10} takes value 1 if seat number 10 is occupied by a woman and seat number 9 is occupied by a man, and 0 otherwise. We define N to be the number of women seating next to at least one man. Therefore,

$$N = I_1 + I_2 + \cdots + I_9 + I_{10}$$

By linearity of expectation,

$$\mathbb{E}(N) = \mathbb{E}(I_1) + \mathbb{E}(I_2) + \cdots + \mathbb{E}(I_9) + \mathbb{E}(I_{10})$$

Notice that while the I_j's are not independent, this is irrelevant for $\mathbb{E}(N)$ By symmetry,

$$\mathbb{E}(I_1) = \mathbb{E}(I_{10}) \quad \text{and} \quad \mathbb{E}(I_2) = \cdots = \mathbb{E}(I_9)$$

So, we just need to find $\mathbb{E}(I_1)$ and $\mathbb{E}(I_2)$.

$$\begin{aligned}
\mathbb{E}(I_1) &= \Pr(I_1 = 1) = \frac{\binom{5}{1}\binom{5}{1}8!}{10!} = \frac{5}{18} \\
\mathbb{E}(I_2) &= \Pr(I_2 = 1) \\
&= 1 - \Pr(I_2 = 0) \\
&= 1 - [\Pr(\text{seat number 2 is occupied by man}) + \\
&\quad \Pr(\text{seat numbers 1, 2 and 3 are occupied by women})] \\
&= 1 - \left[\frac{\binom{5}{1}9!}{10!} + \frac{\binom{5}{3}3!7!}{10!}\right] = 1 - \left[\frac{1}{2} + \frac{1}{12}\right] = \frac{5}{12}
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathbb{E}(N) &= \mathbb{E}(I_1) + \mathbb{E}(I_2) + \cdots + \mathbb{E}(I_9) + \mathbb{E}(I_{10}) \\
&= [2 \times \mathbb{E}(I_1)] + [8 \times \mathbb{E}(I_2)] = \frac{5}{9} + \frac{10}{3} = \frac{35}{9}
\end{aligned}$$

Example 3.17: [Click]

What is the expected number of coin flips until you get 3 heads in a row?

Solution 3.17

Let N_0 denotes the number of coin flips required to get 3 heads in a row starting from an initial state or any other state in which the last coin flip resulted in a tail. Also let N_1 denotes the number of coin flips required to get 3 heads in a row starting from a state where we have observed one head in the only toss so far or any other state in which the last two coin flips resulted in a tail followed by a head. Likewise, let N_2 denotes the number of coin flips required to get 3 heads in a row starting from a state where we have observed one tail followed by two consecutive heads in the last three flips, or simply two consecutive heads if we have just flipped the coin twice so far. Let us use n_i to denote $\mathbb{E}(N_i)$, and we just need to solve the system of equations to get the required quantity.

$$\begin{aligned}
n_0 &= 1 + 0.5n_0 + 0.5n_1 \\
n_1 &= 1 + 0.5n_2 + 0.5n_0 \\
n_2 &= 1 + 0.5n_0
\end{aligned}$$

Solving the above system yields:

$$n_0 = 14, n_1 = 12, n_2 = 8$$

Therefore, the expected number of coin flips until we get three heads in a row is 14.

Example 3.18: [Click]

X_1, X_2, X_3 be a i.i.d random sample of size 3 drawn from a population with $\text{Bern}(p)$ distribution. What is the distribution of $Y = \max(X_1, X_2, X_3)$?

Solution 3.18

Notice that random variable Y takes value 0 when $X_1 = X_2 = X_3 = 0$, and 1 otherwise. Therefore, $\Pr(Y = 0) = (1-p)^3$ and $\Pr(Y = 1) = 1 - (1-p)^3$.

Example 3.19: [Click]

The random variable X takes the values -1 and 1, each with probability 0.5. What is the covariance between X and X^2?

Solution 3.19

Given that X take only two values -1 and 1, $X^3 = X$. Also, X take values -1 and 1 with equal probability so $\mathbb{E}(X) = 0$. It follows that $\mathbb{C}(X, X^2) = \mathbb{E}(X^3) - \mathbb{E}(X^2)\mathbb{E}(X) = \mathbb{E}(X)(1 - \mathbb{E}(X^2)) = 0$.

Example 3.20: [Click]

A random variable X has binomial distribution with mean 4 and variance 2.4. What is the probability that X is positive?

Solution 3.20

Given that $X \sim \text{Bin}(n, p)$, find n and p by solving the following:

$$np = 4, \quad np - np^2 = 2.4$$

and you will get

$$n = 10, \quad p = 0.4$$

So, $\Pr(X > 0) = 1 - \Pr(X = 0) = 1 - 0.6^{10}$

A random variable X takes only two values, 0 and 1, with $\Pr(X = 0) = 0.3$. What is the value of $\mathbb{E}(X^{11})$?

4 Continuous Random Variables

Definition 4.1: Continuous Random Variable

A random variable X is called continuous if there is a non-negative function f_X, called the probability density function of X, or PDF for short, such that

$$\Pr(a < X < b) = \int_a^b f_X(x)dx$$

for all a, b. Note that to qualify as a PDF, a function f_X must be non-negative, i.e., $f_X(x) \geq 0$ for every x, and must also have the normalization property

$$\int_{-\infty}^{\infty} f_X(x)dx = \Pr(-\infty < X < \infty) = 1$$

Definition 4.2: Expectation and Variance of a Continuous Random Variable

$$\begin{aligned} \mathbb{E}(X) &= \int_{-\infty}^{\infty} x f_X(x)dx \\ \mathbb{V}(X) &= \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \end{aligned}$$

[1]

[1] Table of comparison between Discrete and Continuous RVs

Definition 4.3: Uniform Random Variable

$X \sim \mathcal{U}[a, b]$ if

$$f_X(x) = \begin{cases} \dfrac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Theorem 4.1: Transformation of a Continuous Random Variable

Let X be a continuous random variable with PDF f_X, consider a transformation $Y = g(X)$, where g is differentiable and monotonic, then PDF of Y is given by

$$f_Y(y) = f_X(x)\frac{dx}{dy}$$

where $y = g(x)$.

Theorem 4.2: Universality property of Uniform

If $X \sim F_X$ is a continuous random variable then the transformation $F_X(X) \sim \mathcal{U}[0,1]$ follows a uniform distribution.

Definition 4.4: Normal Random Variable

$X \sim \mathcal{N}(\mu, \sigma^2)$ if it has the density function:

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

[2]

[2] 68%-95%-99.7% Rule

Definition 4.5: Moment Generation Function (MGF)

A random variable X has a MGF

$$M_X(t) = \mathbb{E}(e^{tX})$$

, if this is finite for some interval $(-a, a)$, $a > 0$. Not only can a moment-generating function be used to find moments of a random variable, it can also be used to identify which probability mass function a random variable follows.

Theorem 4.3

A moment-generating function uniquely determines the probability distribution of a random variable.

Theorem 4.4

If X and Y are independent then $M_{X+Y}(t) = M_X(t)M_Y(t)$.

Theorem 4.5

MGF of $X \sim \mathcal{N}(\mu, \sigma^2)$ is $M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$.

Definition 4.6: Exponential Random Variable

$X \sim \text{Expo}(\lambda)$ if it has the density function:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Its CDF is

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt = \begin{cases} 1 - e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Theorem 4.6

Consider the transformation $Y = \lambda X$ where $X \sim \text{Expo}(\lambda)$ then $Y \sim \text{Expo}(1)$. $\mathbb{E}(Y) = \mathbb{V}(Y) = 1$. Also MGF of Y is $M_Y(t) = \dfrac{1}{1-t}$.

Theorem 4.7

MGF of $X \sim \text{Expo}(\lambda)$ is $M_X(t) = \dfrac{\lambda}{\lambda - t}$ where $\lambda > t$.

Definition 4.7: Memoryless Property

We say that random variable X satisfies memoryless property if

$$\Pr(X \geq s + t | X \geq s) = \Pr(X \geq t)$$

for all $s, t \in \mathbb{R}_+$.

Theorem 4.8

A continuous positive random variable X satisfies memoryless property if and only if it is distributed $\text{Expo}(\lambda)$ for some $\lambda > 0$.

Definition 4.8: Chi-square Random Variable

We say $X \sim \chi^2(n)$ if $X = \sum_{i=1}^{n} Z_i^2$ where Z_is are i.i.d. $\mathcal{N}(0,1)$. Note that $\chi^2(1) = \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\chi^2(n) = \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\right)$.

Definition 4.9: Student t_n (Gosset)

We say $T \sim t_n$ if

$$T = \frac{Z}{\sqrt{X/n}}$$

where $Z \sim \mathcal{N}(0,1)$ and $X \sim \chi^2(n)$. Also, Z and X are independent.

Definition 4.10: Joint, Marginal and Conditional Distribution

A random vector (X, Y) is called continuous if there is a non-negative function $f_{X,Y}$, called the joint probability density function of X, Y, or joint PDF for short, such that

$$\Pr(a < X < b, c < Y < d) = \int_c^d \int_a^b f_{X,Y}(x,y)dxdy$$

We define Joint CDF of (X, Y) by

$$F_{X,Y}(x,y) = \Pr(X \leq x, Y \leq y)$$

Marginal PDF of X, denoted by $f_X(x)$, can be obtained in the following way:

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$$

For y with $f_Y(y) > 0$, Conditional PDF of X given $Y = y$, is defined by

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

Definition 4.11: Conditional Expectation of X given $Y = y$

Conditional Expectation of X given $Y = y$, denoted by $\mathbb{E}(X|Y = y)$, is defined as follows:

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

Note: $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|Y))$ and $\mathbb{V}(X) = \mathbb{V}(\mathbb{E}(X|Y)) + \mathbb{E}(\mathbb{V}(X|Y))$ holds for the continuous case as well.

Definition 4.12: Beta Random Variable

$X \sim \text{Beta}(a, b)$, where $a, b > 0$ if its density is given by

$$f_X(x) = \begin{cases} cx^{a-1}(1-x)^{b-1}, & \text{if } 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

where c is a normalizing constant whose value depends on a and b.

Definition 4.13: Gamma function

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$$

for all real $a > 0$.

Theorem 4.9

$\Gamma(n) = (n-1)!$ where n is a positive integer.

Definition 4.14: Gamma Random Variable

We say $X \sim \text{Gamma}(a, 1)$ if its density is

$$f_X(x) = \begin{cases} \frac{1}{\Gamma(a)} x^{a-1} e^{-x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

We say $Y \sim \text{Gamma}(a, \lambda)$ if $Y = \frac{X}{\lambda}$ for $X \sim \text{Gamma}(a, 1)$. Density of Y is

$$f_Y(y) = \begin{cases} \frac{\lambda}{\Gamma(a)} (\lambda y)^{a-1} e^{-\lambda y}, & \text{if } y > 0 \\ 0, & \text{otherwise} \end{cases}$$

Theorem 4.10: Transformations of random vectors

Given a random vector $X = (X_1, X_2)$, and a differentiable function $g : \mathbb{R}^2 \to \mathbb{R}^2$ such that a random vector $Y = (Y_1, Y_2) = g(X_1, X_2)$. The joint PDF of Y can be determined by

$$f_Y(y_1, y_2) = f_X(x_1, x_2) \left| \left| \begin{bmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{bmatrix} \right| \right|$$

where $(y_1, y_2) = g(x_1, x_2)$.

Theorem 4.11: Transformations of random vectors

Consider two independent random variables $X \sim \text{Gamma}(a, 1)$ and $Y \sim \text{Gamma}(b, 1)$. Then $X + Y \sim \text{Gamma}(a + b, 1)$. If we let $T = X + Y$ and $W = \dfrac{X}{X+Y}$, then T and W are independent. Also, $W \sim \text{Beta}(a, b)$. We also get the normalizing constant for $\text{Beta}(a, b)$ as $\dfrac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$.

Solved Problems

Example 4.1: [Click]

Find the $\mathbb{E}(\min(2X - Y, X + Y))$ when X and Y are independently and identically distributed uniform random variables on $[0, 1]$.

Solution 4.1

$$\min(2X - Y, X + Y) = \begin{cases} 2X - Y & \text{if } Y > \frac{X}{2} \\ X + Y & \text{if } Y \leq \frac{X}{2} \end{cases}$$

Therefore,

$$\begin{aligned} & \mathbb{E}\left[\min(2X - Y, X + Y)\right] \\ = & \int_0^1 \int_{\frac{x}{2}}^1 (2x - y)dydx + \int_0^1 \int_0^{\frac{x}{2}} (x + y)dydx \\ = & \frac{5}{12} \end{aligned}$$

Example 4.2: [Click]

Let X be a random variable with a uniform distribution over $[0, 1] \quad \cup \quad [3, 4]$. What is the Cumulative distribution function (CDF) of X?

Solution 4.2

Since X is uniformly distributed over $[0,1] \cup [3,4]$, its probability density function (PDF) is

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{for } x \in [0,1] \cup [3,4] \\ 0 & \text{otherwise} \end{cases}$$

Now we can obtain the CDF from the PDF in this way:

$$F_X(t) = \Pr(X \leq t) = \int_{-\infty}^{t} f_x(x)dx = \begin{cases} 0 & \text{for } t \leq 0 \\ \frac{t}{2} & \text{for } 0 < t \leq 1 \\ \frac{1}{2} & \text{for } 1 < t < 3 \\ \frac{1}{2} + \frac{t-3}{2} & \text{for } 3 < t < 4 \\ 1 & \text{for } t \geq 4 \end{cases}$$

Example 4.3: [Click]

A random variable Y has a uniform distribution over the interval $(0, \theta)$. What is the expected value and variance of Y?

Solution 4.3

Given that $Y \sim \text{Unif}(0, \theta)$ we can find its expected value and variance in the following way :

$$\mathbb{E}(Y) = \int_0^\theta \frac{1}{\theta} y dy = \frac{\theta}{2}$$

and

$$\mathbb{E}(Y^2) = \int_0^\theta \frac{1}{\theta} y^2 dy = \frac{\theta^2}{3}$$

Consequently, the variance is

$$\mathbb{V}(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 = \frac{\theta^2}{12}$$

Example 4.4: [Click]

What is $\Pr\left(Y^2 > X > t\right)$, knowing that $t \in (0,1)$ when X and Y are independent $\text{Unif}[0,1]$ random variables?

Solution 4.4

$$\begin{aligned}
\Pr(Y^2 > X > t) &= \int_0^1 \Pr(Y^2 > X > t|Y = y) f_Y(y) dy \\
&= \int_0^1 \Pr(y^2 > X > t|Y = y) dy \\
&= \int_{\sqrt{t}}^1 \Pr(y^2 > X > t) dy \\
&= \int_{\sqrt{t}}^1 (y^2 - t) dy \\
&= \frac{1}{3} - t + \frac{2}{3} t^{\frac{3}{2}}
\end{aligned}$$

Example 4.5: [Click]

Given that $X \sim \text{Unif}(0,1)$ and $Y|X = p \sim \text{Binom}(10, p)$. Find the variance of Y.

Solution 4.5

We are given that $X \sim \text{Unif}(0,1)$ and $Y|X = p \sim \text{Binom}(10, p)$.
We can find the variance of Y using the following law :
$\mathbb{V}(Y) = \mathbb{V}(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbb{V}(Y|X))$
Here
$\mathbb{V}(\mathbb{E}(Y|X)) = \mathbb{V}(10X) = \frac{100}{12}$
$\mathbb{E}(\mathbb{V}(Y|X)) = \mathbb{E}(10X(1-X)) = 10\left(\frac{1}{2} - \frac{1}{3}\right) = \frac{10}{6}$
So, $\mathbb{V}(Y) = \frac{100}{12} + \frac{10}{6} = 10$

Example 4.6: [Click]

Suppose $X \sim \text{Unif}(0,3)$ and $Y \sim \text{Unif}(0,4)$ and they are independent random variables, what is the probability that $X < Y$?

Solution 4.6

Given that $X \sim \text{Unif}(0,3)$ and $Y \sim \text{Unif}(0,4)$ and they are independent random variables, we can find the required probability in the following way:

$$\begin{aligned}
\Pr(X < Y) &= \mathbb{E}(\Pr(X < Y|X)) \\
&= \int_0^3 \Pr(Y > x|X = x) f_X(x) dx \\
&= \int_0^3 \frac{1}{3} \Pr(Y > x) dx \\
&= \int_0^3 \frac{1}{3} \left(\frac{4-x}{4}\right) dx \\
&= \int_0^3 \frac{4-x}{12} dx = \frac{5}{8}
\end{aligned}$$

Example 4.7: [Click]

Can two dependent random variables become independent after you condition them on a third random variable?

Solution 4.7

Suppose $X|Z = z$ and $Y|Z = z$ be independently and identically distributed uniform random variables $\text{Unif}(0,z)$, and let $Z \sim \text{Unif}(0,1)$.
It follows from the definition that X and Y are dependent, but $X|Z = z$ and $Y|Z = z$ are independent.

Example 4.8: [Click]

If $X \sim \mathcal{N}(0,1)$, what is the distribution of $Y = e^X$?

Solution 4.8

If $X \sim \mathcal{N}(0,1)$ and $Y = e^X$, density of Y is
$f_Y(y) = f_X(x)\dfrac{dx}{dy}$, where $y = e^x$.
Therefore, for $y > 0$,
$f_Y(y) = f_X(\ln y)\left(\dfrac{1}{y}\right) = \dfrac{1}{\sqrt{2\pi}y} e^{-\frac{(\ln y)^2}{2}}$
$Y = e^X$, when X is normally distributed, is commonly referred to as a log-normal random variable (because $\ln Y = X$ has a normal distribution)

Example 4.9: [Click]

Given that $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, and $X = ae^Y$ for $a > 0$, what is the density function of X?

Solution 4.9

Given that $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$, and $X = ae^Y$ for $a > 0$.
Density function of X is
$f_X(x) = f_Y(y)\dfrac{dy}{dx}$, where $x = ae^y$
Therefore, for $x > 0$,

$$f_X(x) = f_Y(\ln x - \ln a)\left(\frac{1}{x}\right) = \frac{1}{\sqrt{2\pi}\sigma_Y x}e^{-\frac{(\ln x - \ln a - \mu_Y)^2}{2\sigma_Y^2}}$$

Example 4.10: [Click]

How can the mode of uniform distribution be determined?

Solution 4.10

Let f_X denotes the density function of random variable X.
Mode m of the distribution of X solves the following:

$$\max_x \quad f_X(x) \qquad \ldots(1)$$

For a uniform random variable $X \sim \text{Unif}(0,1)$, density is

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore, every point in the interval $[0,1]$ solves (1). Hence all values in the interval $[0,1]$ are modes of the distribution of X.

Example 4.11: [Click]

Let Y be an exponential random variable with mean $1/\theta$ where θ is positive. The conditional distribution of X given $Y = \lambda$ has Poisson distribution with mean λ. Then the variance of X is?

Solution 4.11

Given that $Y \sim \text{Expo}(\theta)$, where $\theta > 0$ and $X|Y = \lambda \sim \text{Pois}(\lambda)$, variance of X can be computed as follows:

$$\mathbb{V}(X) = \mathbb{E}(\mathbb{V}(X|Y)) + \mathbb{V}(\mathbb{E}(X|Y)) = \mathbb{E}(Y) + \mathbb{V}(Y) = \frac{1}{\theta} + \frac{1}{\theta^2}$$

Example 4.12: [Click]

Let X and Y be i.i.d Unif$(0,1)$ random variables, then what is $\Pr(Y < (X-0.5)^2)$?

Solution 4.12

Given that X and Y are iid Unif$(0,1)$, the joint density of X and Y is

$$f_{X,Y}(x,y) = \begin{cases} 1 & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

and probability of the event $Y < (X-0.5)^2$ is equal to

$$\begin{aligned} \Pr(Y < (X-0.5)^2) &= \int_0^1 \int_0^{(x-0.5)^2} 1 dy dx \\ &= \int_0^1 (x-0.5)^2 dx \\ &= 2\int_{0.5}^1 (x-0.5)^2 dx \\ &= 2\int_0^{0.5} x^2 dx \\ &= \frac{1}{12} \end{aligned}$$

Example 4.13: [Click]

Suppose $A_1 \sim$ Unif$(0,1)$, $A_2|A_1 \sim$ Unif$(0,A_1)$, $A_3|A_2 \sim$ Unif$(0,A_2)$, ..., $A_n|A_{n-1} \sim$ Unif$(0,A_{n-1})$. Given that $S = A_1 + A_2 + \cdots + A_n + \cdots$, find its expected value.

Solution 4.13

$\mathbb{E}(S) = \sum_{i=1}^{\infty} \mathbb{E}(A_i)$
where
$\mathbb{E}(A_1) = \frac{1}{2}$
$\mathbb{E}(A_2) = \mathbb{E}(\mathbb{E}(A_2|A_1)) = \mathbb{E}(\frac{A_1}{2}) = \frac{1}{2}\mathbb{E}(A_1) = \frac{1}{4}$
$\mathbb{E}(A_3) = \mathbb{E}(\mathbb{E}(A_3|A_2)) = \mathbb{E}(\frac{A_2}{2}) = \frac{1}{2}\mathbb{E}(A_2) = \frac{1}{8}$
Likewise by induction step,
if $\mathbb{E}(A_n) = \frac{1}{2^n}$ then $\mathbb{E}(A_{n+1}) = \frac{1}{2^{n+1}}$.
Here is the proof:
$\mathbb{E}(A_{n+1}) = \mathbb{E}(\mathbb{E}(A_{n+1}|A_n)) = \mathbb{E}(\frac{A_n}{2}) = \frac{1}{2}\mathbb{E}(A_n) = \frac{1}{2^{n+1}}$
Therefore,
$\mathbb{E}(S) = \sum_{i=1}^{\infty} \mathbb{E}(A_i) = \sum_{i=1}^{\infty} \frac{1}{2^i} = 1.$

Example 4.14: [Click]

Given that $Y|X = x \sim \text{Unif}(x-1, x+1)$, $\mathbb{E}(X) = 1$ and $\mathbb{V}(X) = \frac{5}{3}$, find $\mathbb{E}(Y)$ and $\mathbb{V}(Y)$.

Solution 4.14

$\mathbb{E}(Y|X = x) = x$
$\mathbb{V}(Y|X = x) = \frac{1}{3}$
So,
$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(X) = 1$
and
$\mathbb{V}(Y) = \mathbb{V}(\mathbb{E}(Y|X)) + \mathbb{E}(\mathbb{V}(Y|X)) = \mathbb{V}(X) + \mathbb{E}\left(\frac{1}{3}\right) = \frac{5}{3} + \frac{1}{3} = 2.$

Example 4.15: [Click]

If a unit length stick is broken into two pieces, what is the probability that the longer portion is thrice as long as the shorter one?

Solution 4.15

Let $X \sim \text{Unif}(0,1)$ denote the point on the stick from where it is broken into two pieces, the probability that the longer portion is at least thrice as long as the shorter one is equal to the probability that $\max(X, 1-X) \geq 3\min(X, 1-X)$.

$$\begin{aligned}
& \Pr(\max(X, 1-X) \geq 3\min(X, 1-X)) \\
= \; & \Pr(X \geq 3(1-X)) + \Pr(1-X \geq 3X) \\
= \; & \Pr\left(X \geq \frac{3}{4}\right) + \Pr\left(X \leq \frac{1}{4}\right) \\
= \; & \frac{1}{4} + \frac{1}{4} \\
= \; & \frac{1}{2}
\end{aligned}$$

Example 4.16: [Click]

Given the joint density of X and Y,

$$f_{X,Y}(x,y) = \begin{cases} 24xy & \text{if } 0 < x < 1, 0 < y < 1-x \\ 0 & \text{elsewhere} \end{cases}$$

What is $\mathbb{E}(X)$?

Solution 4.16

Density of X is

$$f_X(x) = \begin{cases} \int_0^{1-x} 24xy dy = 12x(1-x)^2 & \text{if } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Therefore, the expected value of X is
$\mathbb{E}(X) = \int_0^1 12x^2(1-x)^2 dx = \frac{2}{5}$

Example 4.17: [Click]

Given the joint density of X and Y,

$$f_{X,Y}(x,y) = \begin{cases} 2(1-x) & \text{if } 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Determine the density function of random variable $U = XY$.

Solution 4.17

To find the density of random variable $U = XY$, we can first find its CDF. Observe that range of U is $(0,1)$. For $u \in (0,1)$, CDF of U at u is

$$F_U(u) = \Pr(U \leq u) = 1 - \Pr(U > u)$$

To find $\Pr(U > u)$ we need to integrate the joint density $f_{X,Y}$ over values of (x,y) in the blue region i.e. where $xy > u$. For $u \in (0,1)$,

$$\begin{aligned} \Pr(U > u) &= \int_u^1 \int_{u/x}^1 2(1-x)dydx \\ &= \int_u^1 2(1-x)\left(1 - \frac{u}{x}\right)dx \\ &= 1 - u^2 + 2u \ln u \end{aligned}$$

Therefore,

$$F_U(u) = 1 - \Pr(U > u) = \begin{cases} u^2 - 2u \ln u & \text{if } u \in (0,1) \\ 0 & \text{if } u \leq 0 \\ 1 & \text{if } u \geq 1 \end{cases}$$

To get the density of U, we can differentiate the CDF,

$$f_U(u) = \begin{cases} 2(u - \ln u - 1) & \text{if } u \in (0,1) \\ 0 & \text{otherwise} \end{cases}$$

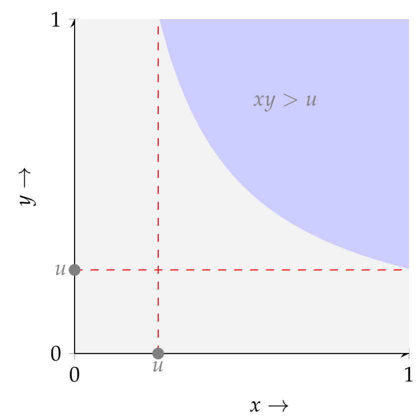

Figure 4.1: $xy > u$

5 Topics in Random Variables

Definition 5.1: Mixture Random variables

Theorem 5.1: Statistical Inequalities

1. Cauchy-Schwarz Inequality. For any two random variables X and Y defined on the same sample space:

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)\mathbb{E}(Y)^2}$$

2. Jensen's Inequality. If g is a convex function, then

$$\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))$$

3. Markov's Inequality. For any $a > 0$,

$$\Pr(|X| \geq a) \leq \frac{\mathbb{E}(|X|)}{a}$$

4. Chebyshev's Inequality. For any $a > 0$,

$$\Pr(|X - \mu| \geq a) \leq \frac{\mathbb{V}(X)}{a^2}$$

   where $\mu = \mathbb{E}(X)$.

Definition 5.2: Convergence in Probability

Let $Y_1, Y_2, \ldots$ be a sequence of random variables (not necessarily independent), and let a be a real number. We say that the sequence Y_n converges to a *in probability*, if for every $\epsilon > 0$, we have

$$\lim_{n \to \infty} \Pr(|Y_n - a| \geq \epsilon) = 0$$

6 Sampling

Definition 6.1: Sample Mean

Consider the sequence $X_1, X_2, \ldots, X_n$ of i.i.d. random variables with mean μ and variance σ^2, we define the sample mean by

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

Theorem 6.1

Let $X_1, X_2, \ldots$ be i.i.d random variables with mean μ and variance σ^2, we have M_n converges to μ *in probability*.

Theorem 6.2: Monte Carlo Integration

Let f be a complicated function whose integral $\int_a^b f(x)dx$ we want to approximate. Assume that $0 \leq f(x) \leq c \; \forall \; x \in [a,b]$, so that we know the integral is finite. The technique of *Monte Carlo Integration* uses random samples to obtain approximations of definite integrals when exact integration methods are unavailable. The procedure is the following: Suppose we pick i.i.d. points $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ uniformly in the rectangle $[a,b] \quad \times \quad [0,c]$. Define indicator r.v.s $I_1, \ldots, I_n$ by letting $I_j = 1$ if $Y_j \leq f(X_j)$ and $I_j = 0$ otherwise. Then I_j are Bernoullin r.v.s with success probability is precisely the ratio of area below f to the area of the rectangle $[a,b] \quad \times \quad [0,c]$. Let $p = \mathbb{E}(I_j)$,

$$p = \mathbb{E}(I_j) = \Pr(I_j = 1) = \frac{\int_a^b f(x)dx}{c(b-a)}$$

We can estimate p using $\frac{1}{n}\sum_{j=1}^n I_j$, and then estimate the desired integral by

$$\int_a^b f(x)dx \approx c(b-a)\frac{1}{n}\sum_{j=1}^n I_j$$

Since I_js are i.i.d with mean p, it follows from the law of large numbers that estimate $\frac{1}{n}\sum_{j=1}^n I_j$ converges to p in probability as the number of points approach infinity.

Theorem 6.3: Convergence of empirical CDF

Let $X_1, X_2, \ldots, X_n$ be i.i.d random variables with CDF F. For each number x, let $R_n(x)$ count how many of $X_1, \ldots, X_n$ are less than or equal to x, that is,

$$R_n(x) = \sum_{j=1}^{n} I(X_j \leq x)$$

Since indicators $I(X_j \leq x)$ are i.i.d with probability of success $F(x)$, we know $R_n(x)$ is binomial with parameters n and $F(x)$. The *empirical* CDF of $X_1, X_2, \ldots, X_n$ is defined as

$$\hat{F}_n(x) = \frac{R_n(x)}{n}$$

considered as a function of x. By law of large numbers, $\hat{F}_n(x)$ converges to $F(x)$ in probability and therefore, can serve as a reasonable estimate of $F(x)$ when n is large.
The empirical CDF is commonly used in nonparametric statistics, a branch of statistics that tries to understand a random sample without making strong assumptions about the family of distributions from which it is originated.

Theorem 6.4: The Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d random variables with mean μ and variance σ^2. Define

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

Then, the CDF of Z_n converges to the standard normal CDF.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2} dt$$

in the sense that

$$\lim_{n \to \infty} \Pr(Z_n \leq z) = \Phi(z)$$

for every z.

Theorem 6.5: Binomial Convergence to Normal

Let $Y \sim \text{Bin}(n, p)$. By Central Limit Theorem, we can consider Y to be a sum of n i.i.d Bern(p) r.v.s. Therefore, for large n,

$$Y \dot{\sim} \mathcal{N}\left(np, np(1-p)\right)$$

If Y is a Binomial random variable with parameters n and p, n is large, and k, l are non negative integers, then

$$\Pr(k \leq Y \leq l) \approx \Phi\left(\frac{l + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{k - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Theorem 6.6: Poisson Convergence to Normal

Let $Y \sim \text{Pois}(n)$. By Central Limit Theorem, we can consider Y to be a sum of n i.i.d Pois(1) r.v.s. Therefore, for large n,

$$Y \dot{\sim} \mathcal{N}(n, n)$$

Theorem 6.7: Gamma Convergence to Normal

Let $Y \sim \text{Gamma}(n, \lambda)$. By Central Limit Theorem, we can consider Y to be a sum of n i.i.d Expo(λ) r.v.s. Therefore, for large n,

$$Y \dot{\sim} \mathcal{N}\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)$$

Theorem 6.8: Volatile Stock

Each day, a very volatile stock rises 70% or drops 50% in price, with equal probabilities and with different days independent. Let Y_n be the stock price after n days, starting from an initial value of $Y_0 = 100$.

1. Explain why $\log Y_n$ is approximately Normal for n large, and state its parameters.
2. What happens to $\mathbb{E}(Y_n)$ as $n \to \infty$?
3. Use the law of large numbers to find out what happens to Y_n as $n \to \infty$.

Theorem 6.9

If $Z_1, Z_2, \ldots, Z_n$ is a random sample from a standard normal distribution, then

1. $\overline{Z}_n$ has a normal distribution with mean 0 and variance $1/n$.
2. $\overline{Z}_n$ and $\sum_{j=1}^{n}(Z_j - \overline{Z}_n)^2$ are independent.
3. $\sum_{j=1}^{n}(Z_j - \overline{Z}_n)^2$ has a chi-square distribution with $n-1$ degrees of freedom.

Theorem 6.10: Distribution of the sample variance

For i.i.d. $X_1, X_2, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$, the sample variance is the r.v.

$$S_n^2 = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X}_n)^2$$

Show that

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi^2_{n-1}$$

Solved Problems

Example 6.1: [*Click*]

Let $X_1, X_2, \ldots$ be the sequence of i.i.d random variables with
$\Pr(X_i = 1) = \frac{1}{4}$
$\Pr(X_i = 2) = \frac{3}{4}$
Define $\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n}$. What is the $\lim_{n\to\infty} \Pr(\overline{X}_n \leq 1.8)$?

Solution 6.1

Given the data above, we can find expectation and variances of X_i and $\overline{X}_n$ as follows:
$\mathbb{E}(X_i) = \frac{7}{4} = 1.75$
$\mathbb{V}(X_i) = \frac{3}{16}$
$\mathbb{E}(\overline{X}_n) = \frac{7}{4} = 1.75$
$\mathbb{V}(\overline{X}_n) = \frac{3}{16n}$
By Chebyshev's Inequality,
$\Pr(|\overline{X}_n - 1.75| > 0.05) \leq \frac{75}{n}$
Therefore, $\lim_{n\to\infty} \Pr(|\overline{X}_n - 1.75| > 0.05) = 0$.
Working on the desired quantity $\Pr(\overline{X}_n \leq 1.8)$, we get

$$\begin{aligned}\Pr(\overline{X}_n \leq 1.8) &= 1 - \Pr(\overline{X}_n - 1.75 > 0.05) \\ &\geq 1 - \Pr(|\overline{X}_n - 1.75| > 0.05)\end{aligned}$$

Since $\lim_{n\to\infty} \Pr(|\overline{X}_n - 1.75| > 0.05) = 0$, the above implies that $\lim_{n\to\infty} \Pr(\overline{X}_n \leq 1.8) = 1$.
Alternatively, we can also apply weak law of large numbers according to which $\overline{X}_n$ converges in probability to 1.75 and consequently, $\lim_{n\to\infty} \Pr(\overline{X}_n \leq 1.8) = 1$ holds.

7 | Estimation

```r
x <- 1:1000
y <- rep(0, length(x))

# for loop
for (i in 1:length(x)){
 if (x[i]%%3 == 0 | x[i] > 50){
   y[i] = x[i]
 } else {
   y[i] = -x[i]
 }
}
sum(y)

x <- 1:1000
y <- x*(x%%3 == 0 | x > 50)-x*(1- (x%%3 == 0 | x > 50))
sum(y)
```

```python
def cost(X, y, theta):
   m = len(y)
   J =
       (1/(2*m))*(((X@theta-y)**2).sum())
   return J

def gradientDescent(X, y,
    theta, alpha, iterations):
   m = len(y)
   J_history =
       np.zeros(iterations)
   for iter in
       range(iterations):
      theta = theta -
          (((alpha/m)*(X@theta
          - y).T@X).T)
      J_history[iter] =
          cost(X, y, theta)
   return theta, J_history
```

Example 7.1: *[Click]*

Suppose that we know that the random variable X is uniform with support $[0, b]$. Suppose that we observed $X = 2.5$, what is an unbiased estimate of b based on this single observation?

Solution 7.1

Given that $X \sim \text{Unif}[0, b]$, $\mathbb{E}(X) = \frac{b}{2}$. Therefore, $\mathbb{E}(2X) = b$. In other words, $2X$ is an unbiased estimator of b, and the corresponding estimate of b when the realization of X equals 2.5 is $2(2.5) = 5$.

Example 7.2: *[Click]*

Given that Y_1, Y_2, Y_3 are iid $\text{Unif}(0, \theta)$, list some unbiased estimators of θ.

Solution 7.2

Given that Y_1, Y_2, Y_3 are iid $\text{Unif}(0, \theta)$, all these are unbiased estimators of θ :

- $\hat{\theta}_1 = \frac{2}{3}(Y_1 + Y_2 + Y_3)$
- $\hat{\theta}_2 = Y_1 + Y_2$
- $\hat{\theta}_3 = 2Y_1$
- $\hat{\theta}_4 = \frac{4}{3}\max(Y_1, Y_2, Y_3)$
- $\hat{\theta}_5 = \max(Y_1, Y_2, Y_3) + \min(Y_1, Y_2, Y_3)$

References

- Introduction to Probability by Blitzstein & Hwang
- Introduction to Probability by Bertsekas & Tsitsiklis