ENGG2760 Probability for Engineers

Ryan Chan

April 9, 2025

Abstract

This is a note for **ENGG2760 - Probability for Engineers**.
Contents are adapted from the lecture notes of ENGG2760, prepared by Farzan Farnia, as well as some online resources.
This note is intended solely as a study aid. While I have done my best to ensure the accuracy of the content, I do not take responsibility for any errors or inaccuracies that may be present. Please use the material thoughtfully and at your own discretion.
If you believe any part of this content infringes on copyright, feel free to contact me, and I will address it promptly.
Mistakes might be found. So please feel free to point out any mistakes.
This note also includes some extra content that isn't included in the syllabus of ENGG2760, but I found it somehow useful for overall understanding. They are mainly adapted from MIT RES.6-012 Introduction to Probability.

Contents

1 Probability and Counting **2**
- 1.1 Introduction . . . 2
- 1.2 Permutation and Combination . . . 3

2 Probability Models and Axioms **6**
- 2.1 Basic Definitions . . . 6
- 2.2 Probability Axioms . . . 7
- 2.3 Rules for Probability Calculation . . . 7

3 Conditional Probability and Bayes' Rule **9**
- 3.1 Conditional Probability . . . 9
- 3.2 Bayes' Rule . . . 12
- 3.3 Independence . . . 13

4 Random Variables **16**
- 4.1 Introduction . . . 16
- 4.2 Bernoulli Random Variable . . . 17
- 4.3 Discrete Uniform Random Variable . . . 17
- 4.4 Binomial Random Variable . . . 17
- 4.5 Geometric Random Variable . . . 19
- 4.6 Cumulative Distribution Function (CDF) . . . 21
- 4.7 Poisson Random Variable . . . 22
- 4.8 Properties of Random Variables . . . 23

5 Expectation, Variance and Conditional PMF **26**
- 5.1 Expectation . . . 26
- 5.2 Variance . . . 28
- 5.3 Conditional PMF . . . 31
- 5.4 Independent Random Variable . . . 33
- 5.5 Appendix . . . 36

6 Continuous Random Variables **37**
- 6.1 Cumulative Distribution Function (CDF) . . . 37
- 6.2 Uniform Random Variable . . . 39
- 6.3 Exponential Random Variable . . . 40
- 6.4 Normal Distribution (Gaussian Random Variable) . . . 42
- 6.5 Multiple Continuous Random Variable . . . 44

7 Chebyshev's Inequality **48**
- 7.1 Chebyshev's Inequality . . . 48
- 7.2 Law of Large Numbers . . . 49
- 7.3 Central Limit Theorem . . . 50

A Z TABLE **52**

Chapter 1

Probability and Counting

1.1 Introduction

We will start with some basic definitions.

Definition 1.1.1 (Sample Space)**.** The sample space Ω is the set of all possible outcomes

For example, when flipping three coins, we have $2^3 = 8$ outcomes:

$$\Omega = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}, \text{THH}, \text{THT}, \text{TTH}, \text{TTT}\}$$

Definition 1.1.2 (Event)**.** An event is a subset of the sample space.

Following the above example, if A is the event that at least two heads occur, we have:

$$A = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}\}$$

Definition 1.1.3. The probability of an event is the sum of the probability of its outcomes.

- Probabilities are non-negative.
- Probabilities add up to one.

Again from the above example, we see that the probability of each event is equal to $\frac{1}{8}$, and they can be summed up to 1.

Definition 1.1.4. The probability of an event is the sum of the probabilities of its outcomes.

For event A, the probability would be

$$\mathbb{P}(A) = \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}$$

Proposition 1.1.1 (Uniform Probability Law)**.** If the outcomes in Ω are equally likely, then the probability of event A will be

$$\mathbb{P}(A) = \frac{\text{Number of outcomes in } A}{\text{Number of outcomes in } \Omega} = \frac{|A|}{|\Omega|}$$

Remark. It can only be used when every outcome is equally likely.

For event A, the probability would be

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}$$

Example. We roll two dice. Which of the following outcome is more likely for the sum of the two dice?

1. 11
2. 12
3. equally likely

Solution: For the sum to be 11, we can have (5, 6) and (6, 5). However, for the sum to be 12, we can only have (6, 6). Therefore, for $|\Omega| = 6^2 = 36$,

$$\mathbb{P}(11) = \frac{2}{36}, \quad \mathbb{P}(12) = \frac{1}{36}$$

Therefore, the sum of 11 would be more likely to occur.

1.2 Permutation and Combination

1.2.1 Counting via Product Rule

Proposition 1.2.1 (Product Rule)**.** Suppose there are n possible outcomes for Experiment 1 and m possible outcomes for Experiment 2, where the two experiments are independent. Then, there are $m \times n$ possible outcomes for the two experiments.

For example, when flipping three coins, each of them has two possible outcomes. Therefore, there are in total $2 \times 2 \times 2 = 8$ possible outcomes.

We can then generalize this rule for cases that the outcomes of experiment 1 may affect the outcomes of experiment 2.

Proposition 1.2.2 (Generalized Product Rule)**.** Suppose that

- There are n possible outcomes for Experiment 1.
- For every outcome of Experiment 1, there are m possible outcomes for Experiment 2.

Then, there are $m \times n$ possible outcomes for the two experiments.

For example, when finding all possible outcomes for rolling two dice with different values, the outcomes of the first experiment, i.e. rolling the first die, would be 6. The outcomes of the second experiment, i.e. rolling the second die, would be 5 (since we need to exclude the outcome of the first die). Then, there are in total $6 \times 5 = 30$ possible outcomes.

Example. We roll two dice. What is the probability that they come out with different values?

Solution: Let A be the desired event. Then we have

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6 \times 5}{6 \times 6} = \frac{5}{6}$$

Example. We roll two dice. What is the probability that the sum of dice equals 7? What is the probability that the sum of dice is an odd number?

Solution: Let A be the event that the sum of dice equals 7. Then we have

$$A = \{(1,6), (2,5), \cdots, (6,1)\}$$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6}{6^2} = \frac{1}{6}$$

Let B be the event that the sum of dice is an odd number. Then we have

$$B = \{(1,2), (1,4), \cdots, (6,5)\}, \quad |B| = 6 \times 3,$$

where for each number in the first die, there will be exactly three numbers in the second die that can be added up to an odd number. Thus,

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} = \frac{6 \times 3}{6^2} = \frac{1}{2}$$

Example. We again roll two dice. What is the probability that the first die is bigger than the second die?

Solution: In this case, we cannot use generalized product rule since for every outcome in the first experiment, there will be a different outcome in the second experiment. Let A be the desired event. Then we have

$$A = \{(2,1), (3,1), \cdots, (6,5)\}$$

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{15}{6^2} = \frac{5}{12}$$

1.2.2 Permutation

Definition 1.2.1 (Permutation)**.** A permutation of n different objects is an arrangement of the objects into an ordered sequence (order matters).

Proposition 1.2.3. For n different objects, there exists $n!$ different permutations:

$$n! = n \times (n-1) \times \cdots \times 2 \times 1$$

Example. We roll six dice. How many ways are there for the six dice to have different values? What is the probability of that event?

Solution: Let A be the desired event. Then we have

$$|A| = 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 6! = 720, \quad \mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{6!}{6^6}$$

Example (Birthday Paradox)**.** Suppose there are n people in a room. We assume that a year only has 365 days, and that every day is equally likely to be the birthday of a person. What is the probability that at least two people have the same birthday? Here we assume that $n < 365$.

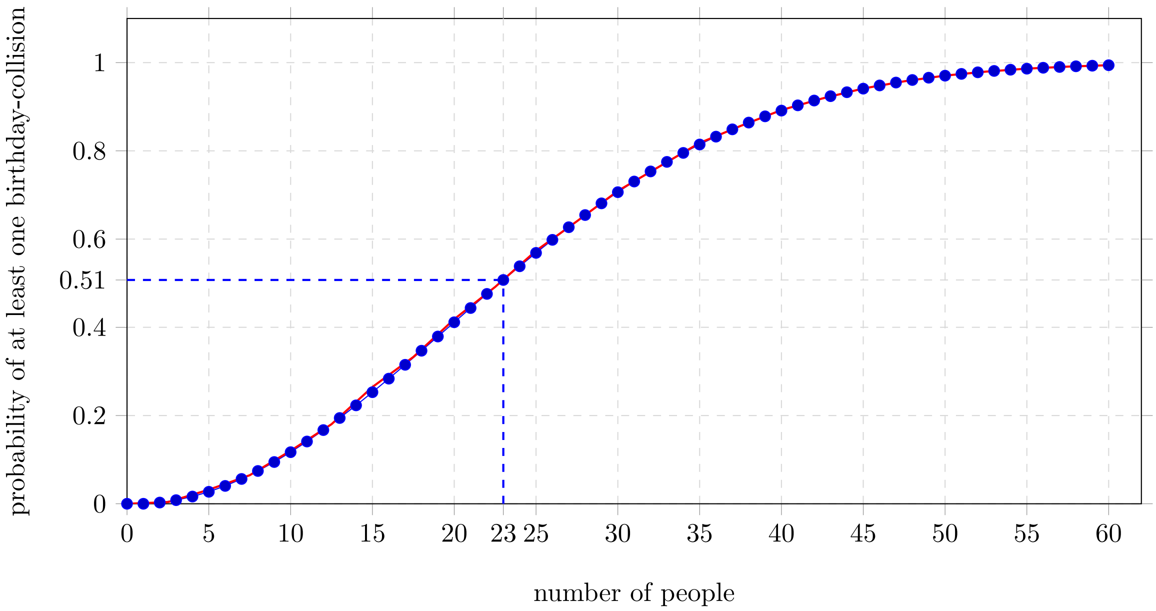
For sample space S we have the set of all possible sequences of n birthday, the $|S| = 365^n$.

Let T be the event in which at least two birthdays are the same. Then we have

$$\mathbb{P}(T) = 1 - \frac{365 \times 364 \times \cdots \times (365 - n + 1)}{365^n},$$

where the term $(365 \times 364 \times \cdots \times (365 - n + 1))$ is to count the possible outcomes for the event that all birthdays are distinct.

Birthday paradox could be visualized as below:



Adapted from MartinThoma

1.2.3 Binomial Coefficient

Proposition 1.2.4 (Binomial Coefficient or "n-Choose-k"). Given a set S of size n, the number of subsets of size k will be

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

It can also be understood as the number of possible arrangements of k objects of Type A and $n-k$ objects of Type B into an ordered sequence.

Example. A box contains 8 red balls and 2 blue balls. You draw 2 balls at random(without replacement). What is the probability that the two balls have different colors?

Solution: Let A be the desired event.

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{\binom{8}{1}\binom{2}{1}}{\binom{10}{2}} = \frac{16}{45}$$

Proposition 1.2.5 (Multinomial Coefficient). For a set S of size n, the number of partitioning of the set to partitions of size $k_1, k_2, \cdots, k_t$ (noted that $n = k_1 + k_2 + \cdots + k_t$) will be

$$\binom{n}{k_1, k_2, \cdots, k_t} = \frac{n!}{k_1!k_2!\cdots k_t!}$$

It can also be understood as the number of possible permutations of k_1 objects of Type 1, k_2 objects of Type 2, ..., and k_t objects of Type t.

Chapter 2

Probability Models and Axioms

2.1 Basic Definitions

We will introduce some definitions here as well.

Definition 2.1.1 (Complement)**.** The complement of event A (denoted by A^c) is the opposite event of A. In other words, A^c happens if and only if A does not happen.

Again, when flipping three coins, we have the following sample space:

$$\Omega = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\}.$$

Let A be the event that at least two heads occur. Then for A^c, we have:

$$A^c = \{\mathrm{TTT}, \mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}\}$$

Definition 2.1.2 (Intersection of Events)**.** The intersection of events happens when all the events occur. We denote this intersection of event A and B with $A \cap B$.

Let B be the event that no consecutive heads occurs. Then, for $A \cap B$, we have the event that at least two heads and no consecutive heads occur.

$$A \cap B = \{\mathrm{HTH}\}$$

Definition 2.1.3 (Union of Events)**.** The union of events happens when at least one of the events occur. We denote the union of events A and B with $A \cup B$.

For example, for $A \cup B$ in the above example, we have

$$A \cup B = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\}$$

Definition 2.1.4 (Disjoint Events)**.** We call event $A_1, A_2, \cdots$ disjoint events (or mutually exclusive events) if the intersection of every two events $A_i, A_j (i \neq j)$ is the null event:

$$\forall i \neq j : \quad A_i \cap A_j = \varnothing$$

Let C be the event that at least three heads occur. Then

$$B \cap C = \varnothing.$$

2.2 Probability Axioms

Definition 2.2.1 (Axioms of Probability). A probability assignment $\mathcal{P}$ to sample space Ω should satisfy the following three axioms:

1. For every event A, $0 \leq \mathbb{P}(A)$;
2. $\mathbb{P}(\Omega) = 1$;
3. If event $A_1, A_2, \cdots$ are disjoint, $\mathbb{P}(A_1 \cup A_2 \cup \cdots) = \mathbb{P}(A_1) + \mathbb{P}(A_2) + \cdots$

Follow these axioms, and we can prove most of the rules for probability calculation.

2.3 Rules for Probability Calculation

Proposition 2.3.1 (Complement Rule). For every event E and its complement E^c:

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E)$$

Proposition 2.3.2 (Difference Rule). If event E, F satisfy $E \subseteq F$, then:

$$\mathbb{P}(F \cap E^c) = \mathbb{P}(F) - \mathbb{P}(E)$$

Remark. As a result, if $E \subseteq F$, then $\mathbb{P}(E) \leq \mathbb{P}(F)$

Proof.

$$\begin{aligned}\mathbb{P}(F \cap E^c) &= \mathbb{P}(F) - \mathbb{P}(E) \\ \mathbb{P}(F) &= \mathbb{P}(F \cap E^c) + \mathbb{P}(E)\end{aligned}$$

Since $(F \cap E^c) \cap E = F \cap (E^c \cap E) = F \cap \varnothing = \varnothing$, $\mathbb{P}(F \cap E^c) + \mathbb{P}(E) \Rightarrow (F \cap E^c) \cup E$

$$\begin{aligned}(F \cap E^c) \cup E &= (F \cup E) \cap (E^c \cup E) \\ (F \cap E^c) \cup E &= (F \cup E) \cap \Omega \\ (F \cap E^c) \cup E &= F \cup E \\ (F \cap E^c) \cup E &= F \text{ (for } E \subseteq F)\end{aligned}$$

■

Proposition 2.3.3 (Inclusion-Exclusion Principle). For events E, F:

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)$$

Remark. We can generalize the principle to more than two events. For example,

$$\mathbb{P}(E \cup F \cup G) = \mathbb{P}(E) + \mathbb{P}(F) + \mathbb{P}(G) - \mathbb{P}(E \cap F) - \mathbb{P}(E \cap G) - \mathbb{P}(F \cap G) + \mathbb{P}(E \cap F \cap G)$$

Example. In a city, 10% of the people are rich, 5% are famous, and 3% are both rich and famous. For a randomly-selected person in the city, find the probability for the following Events.

Here we let R be the event that the person is rich, F be the event that the person is famous,

1. The person is not rich.

$$\mathbb{P}(R^c) = 1 - \mathbb{P}(R) = 1 - 0.1 = 0.9$$

2. The person is not rich but is famous.

$$\mathbb{P}(R^c \cap F) = \mathbb{P}(F) - \mathbb{P}(F \cap R) = 0.05 - 0.03 = 0.02$$

3. The person is neither rich nor famous.

$$\mathbb{P}(F^c \cap R^c) = 1 - \mathbb{P}(F \cup R) = 1 - \mathbb{P}(F) - \mathbb{P}(R) + \mathbb{P}(F \cap R) = 1 - 0.05 - 0.1 + 0.03 = 0.88$$

Chapter 3

Conditional Probability and Bayes' Rule

3.1 Conditional Probability

Let's begin with an example

Example. We toss 3 coins. You win if at least two heads come out. What is the probability of winning?

Solution:

$$\Omega = \{\text{HHH}, \text{HHT}, \cdots, \text{TTT}\} \quad \Rightarrow \quad \mathbb{P}(W) = \frac{|W|}{|\Omega|} = \frac{4}{8} = \frac{1}{2}$$

However, if it is given that the first coin was tossed and came out head. How does this affect your chances of winning? Since the sample space now changes due to the condition that the first toss is heads, the way we find the probability differs.

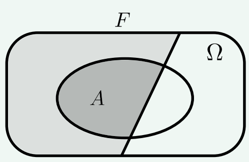
$$\Omega' = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{HTT}\}, \quad W' = \{\text{HHH}, \text{HHT}, \text{HTH}\}$$

This give the probability $\frac{3}{4}$.

3.1.1 Conditional Probability

Definition 3.1.1 (Conditional Probability).

The Conditional Probability $\mathbb{P}(A|F)$ represents the probability of event A assuming (or given) that event F happened.



Remark. All the outcomes of Ω and event A in F should be excluded in the calculation.

The conditional probability of A with respect to reduced sample space F is given by the formula:

$$\mathbb{P}(A|F) = \frac{\mathbb{P}(A \cap F)}{\mathbb{P}(F)}$$

Example. You roll two dice. You win if the sum of the outcomes is 8. If the first die toss is a 4, should you be happy?

Solution: Considering the initial case, we have:

$$\mathbb{P}(W) = \frac{|W|}{|\Omega|} = \frac{5}{36}$$

Let D be the event that the first toss is a 4. Given D, we have:

$$\mathbb{P}(W|D) = \frac{\mathbb{P}(W \cap D)}{\mathbb{P}(D)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

Therefore, the probability of winning increases.

Example. For the same game, you win if the sum of the outcomes is 7. The first toss is a 4. Should you be happy?

Solution: Considering the initial case, we have:

$$\mathbb{P}(W) = \frac{|W|}{|\Omega|} = \frac{6}{36} = \frac{1}{6}$$

Let D be the event that the first toss is a 4. Given D, we have:

$$\mathbb{P}(W|D) = \frac{\mathbb{P}(W \cap D)}{\mathbb{P}(D)} = \frac{\frac{1}{36}}{\frac{1}{6}} = \frac{1}{6}$$

Therefore, the probability does not change.

3.1.2 Properties of Conditional Probability

Proposition 3.1.1 (Properties of Conditional Probability)**.** Conditional Probability $\mathbb{P}(\cdot|D)$ are probabilities over reduced sample space F and satisfy probability axioms:

1. For every A, $\mathbb{P}(A|F) \geq 0$
2. $\mathbb{P}(F|F) = 1$
3. For disjoint events A, B: $\mathbb{P}(A \cup B|F) = \mathbb{P}(A|F) + \mathbb{P}(B|F)$

Remark. There is no assumption about A or B are subsets of F.

We can then generalize the conditional probability using Uniform Probability Law. Under equally-likely outcomes in F:

$$\mathbb{P}(A|F) = \frac{\text{Number of outcomes in } A \cap F}{\text{Number of outcomes in } F} = \frac{|A \cap F|}{|F|}$$

Example.

There are three cards with Black/Black, Red/Red and Red/Black sides. We randomly draw one card and observe that one of its sides is Black.
What is the probability that the card's other side is also Black?

Solution: Let B be the event that the first drawn color is black, and E be the event that the second drawn color is also black. Then we have:

$$\Omega = \{1F, 1B, 2F, 2B, 3F, 3B\}$$

$$\mathbb{P}(E|B) = \frac{|E \cap B|}{|B|} = \frac{2}{3}$$

3.1.3 The Multiplication Rule

Proposition 3.1.2 (The Multiplication Rule). For events E_1, E_2 we can write the probability of their intersection $E_1 \cap E_2$ as

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)$$

In general, for every $E_1, E_2, \cdots, E_n$, the multiplication rule says

$$\mathbb{P}(E_1 \cap E_2 \cap \cdots \cap E_n) = \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\mathbb{P}(E_3|E_1 \cap E_2)\cdots\mathbb{P}(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

Example. A box contains 5 red balls and 15 blue balls. We randomly draw three balls from the box (without replacement). What is the probability that the balls are all red?

Solution: Let R_i be the event that the i-th ball being drawn is red.

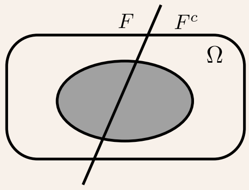
Then we have

$$\mathbb{P}(R_1 \cap R_2 \cap R_3) = \mathbb{P}(R_1)\mathbb{P}(R_1|R_2)\mathbb{P}(R_3|R_1 \cap R_2) = \frac{5}{20} \times \frac{4}{19} \times \frac{3}{18} = \frac{1}{114}$$

Theorem 3.1.1 (Total Probability Theorem).

For every event E and F and its complement F^c,

$$\begin{aligned}\mathbb{P}(E) &= \mathbb{P}(E \cap F) + \mathbb{P}(E \cap F^c) \\ &= \mathbb{P}(E|F)\mathbb{P}(F) + \mathbb{P}(E|F^c)\mathbb{P}(F^c)\end{aligned}$$



More generally, if events $F_1, F_2, \cdots, F_n$ partition Ω (disjoint events and $F_1 \cup F_2 \cup \cdots \cup F_n = \Omega$), then the total probability theorem says

$$\mathbb{P}(E) = \mathbb{P}(E|F_1)\mathbb{P}(F_1) + \mathbb{P}(E|F_2)\mathbb{P}(F_2) + \cdots + \mathbb{P}(E|F_n)\mathbb{P}(F_n)$$

Example. A box contains 5 red balls and 15 blue balls. We randomly draw two balls from the box (without replacement). What is the probability that the balls have different colors?

Solution: Let R be the event that the first ball is red, E be the event that the balls have different colors. Then we have

$$\mathbb{P}(E) = \mathbb{P}(E|R)\mathbb{P}(R) + \mathbb{P}(E|R^c)\mathbb{P}(R^c) = \frac{15}{19} \times \frac{5}{20} + \frac{5}{19} \times \frac{15}{20} = \frac{15}{38}$$

Example. For the situation that a group of students answered one multiple choice question where there are 4 options. What is the probability that a student knows the answer to the question?

Solution: Here we define event K as student knows the answer to the question, and event C as student correctly answers the question. Then, the total probability theorem says

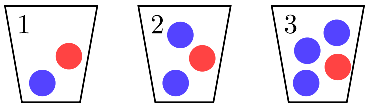
$$\begin{aligned}\mathbb{P}(C) &= \mathbb{P}(C|K)\mathbb{P}(K) + \mathbb{P}(C|K^c)\mathbb{P}(K^c) \\ &= 1 \times \mathbb{P}(K) + \frac{1}{4} \times (1 - \mathbb{P}(K)) \\ &= \frac{3}{4}\mathbb{P}(K) + \frac{1}{4}\end{aligned}$$

This gives

$$\mathbb{P}(K) = \frac{4\mathbb{P}(C) - 1}{3}$$

3.2 Bayes' Rule

We choose a cup at random and then a random ball from that cup. The selected ball is red. Which cup do you guess the ball came from?



This is an example of cause and effect. Here, the cause is the cup number and the effect is the ball's color. Choosing different cups leads to different probabilities of choosing the color of balls.

Theorem 3.2.1 (Bayes' Rule). Consider events C and E. Then,

$$\mathbb{P}(C|E) = \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E)} = \frac{\mathbb{P}(E|C)\mathbb{P}(C)}{\mathbb{P}(E|C)\mathbb{P}(C) + \mathbb{P}(E|C^c)\mathbb{P}(C^c)}$$

More generally, if $C_1, C_2, \cdots, C_n$ partition the set of possible causes S,

$$\mathbb{P}(C_1|E) = \frac{\mathbb{P}(E|C_1)\mathbb{P}(C_1)}{\mathbb{P}(E|C_1)\mathbb{P}(C_1) + \mathbb{P}(E|C_2)\mathbb{P}(C_2) + \cdots + \mathbb{P}(E|C_n)\mathbb{P}(C_n)}$$

Let us revisit the example above. Let R represent the event that a red ball is drawn, and let C_i denote the cup from which the red ball is drawn, where $i = 1, 2, 3$. We can then calculate the probability of each cup being the source of the red ball, given that the ball is red.

$$\mathbb{P}(C_i|R) = \frac{\mathbb{P}(R|C_i)\mathbb{P}(C_i)}{\mathbb{P}(R|C_1)\mathbb{P}(C_1) + \mathbb{P}(R|C_2)\mathbb{P}(C_2) + \mathbb{P}(R|C_3)\mathbb{P}(C_3)}$$

For sample space, we have: $\Omega = \{1B, 1R, 2B_1, 2B_2, 2R, 3B_1, 3B_2, 3B_3, 3R\}$ (not equally likely)

	Cup 1	Cup 2	Cup 3
$\mathbb{P}(C_i)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
$\mathbb{P}(R|C_i)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$

For $\mathbb{P}(R)$, by using total probability theorem, we have

$$\mathbb{P}(R) = \mathbb{P}(R|C_1)\mathbb{P}(C_1) + \mathbb{P}(R|C_2)\mathbb{P}(C_2) + \mathbb{P}(R|C_3)\mathbb{P}(C_3) = \frac{1}{2} \times \frac{1}{3} + \frac{1}{3} \times \frac{1}{3} + \frac{1}{4} \times \frac{1}{3} = \frac{13}{36}$$

Then we have:

$$\mathbb{P}(C_1|R) = \frac{\mathbb{P}(R|C_1)\mathbb{P}(C_1)}{\mathbb{P}(R)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\frac{13}{36}} = \frac{6}{13}, \mathbb{P}(C_2|R) = \frac{\mathbb{P}(R|C_2)\mathbb{P}(C_2)}{\mathbb{P}(R)} = \frac{\frac{1}{3} \times \frac{1}{3}}{\frac{13}{36}} = \frac{4}{13}$$

$$\mathbb{P}(C_3|R) = \frac{\mathbb{P}(R|C_3)\mathbb{P}(C_3)}{\mathbb{P}(R)} = \frac{\frac{1}{4} \times \frac{1}{3}}{\frac{13}{36}} = \frac{3}{13}$$

Example. Two classes take place in the same academic building. Class A has 100 students from whom 20% are female, Class B has 10 students from whom 80% are female.

Now we see a female student in this building. What is the probability that the student is from Class A?

Solution:

$$\mathbb{P}(A|F) = \frac{\mathbb{P}(F|A)\mathbb{P}(A)}{\mathbb{P}(F|A)\mathbb{P}(A) + \mathbb{P}(F|B)\mathbb{P}(B)} = \frac{20\% \times \frac{10}{11}}{20\% \times \frac{10}{11} + 80\% \times \frac{1}{11}} = \frac{5}{7}$$

This example demonstrates a counterintuitive result. Although the proportion of females in Class B is greater than that in Class A, the probability of a female student belonging to Class A is higher. This highlights the importance of using Bayes' rule to update our beliefs, as our initial hypotheses may not always align with the actual probabilities.

To summarize, we see that conditional probability will be a very powerful tool when

1. the studied environment includes causes and effect. This is especially useful for calculating the probability of a cause under an observation of the effect.

2. we want to calculate an ordinary probability and conditioning on the right event can simplify the description of the sample space.

3.3 Independence

3.3.1 Independence of Two Events

Definition 3.3.1 (Independent Events). We call Events A and B independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \quad \text{or equivalently} \quad \mathbb{P}(A|B) = \mathbb{P}(A)$$

Example. We toss a coin three times. Then the following events are independent:

$$\Omega = \{\mathrm{HHH}, \mathrm{HHT}, \mathrm{HTH}, \mathrm{HTT}, \mathrm{THH}, \mathrm{THT}, \mathrm{TTH}, \mathrm{TTT}\}$$

Event E_1: The first toss show heads

Event E_2: The second and third toss both show tails

We have: $\mathbb{P}(E_1) = \frac{4}{8}, \mathbb{P}(E_2) = \frac{2}{8}, \mathbb{P}(E_1 \cap E_2) = \frac{1}{8}, \mathbb{P}(E_1)\mathbb{P}(E_2) = \frac{4}{8} \times \frac{2}{8} = \frac{1}{8} = \mathbb{P}(E_1 \cap E_2)$

This shows that the events are independent.

Example. We roll two dice. Consider events:

E_1: the first die is 4; S_6: the sum of dice is 6; S_7: the sum of dice is 7.

$$\mathbb{P}(E_1) = \frac{1}{6};\ \mathbb{P}(S_6) = \frac{5}{36};\ \mathbb{P}(S_7) = \frac{1}{6};\ \mathbb{P}(E_1 \cap S_6) = \frac{1}{36};\ \mathbb{P}(E_1 \cap S_7) = \frac{1}{36};\ \mathbb{P}(S_6 \cap S_7) = 0.$$

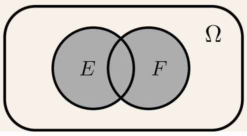
Then we know that (E_1, S_6) and (S_6, S_7) are not independent, (E_1, S_7) are independent.

Remark. Mutually exclusive does not mean independent.

Proposition 3.3.1.

If E, F are independent events, then events E^c, F will also be independent.

$$\mathbb{P}(E^c \cap F) = \mathbb{P}(E^c)\mathbb{P}(F)$$



Proof.

$$\begin{aligned}
\mathbb{P}(E^c \cap F) &= \mathbb{P}(F) - \mathbb{P}(E \cap F) \\
&= \mathbb{P}(F) - \mathbb{P}(E)\mathbb{P}(F) \\
&= \mathbb{P}(F)(1 - \mathbb{P}(E)) \\
&= \mathbb{P}(E^c)\mathbb{P}(F)
\end{aligned}$$

■

3.3.2 Independence of Several Events

We call three events A, B, C independent events if these four conditions are satisfies:

1. A and B are independent: $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
2. B and C are independent: $\mathbb{P}(B \cap C) = \mathbb{P}(B)\mathbb{P}(C)$
3. A and C are independent: $\mathbb{P}(A \cap C) = \mathbb{P}(A)\mathbb{P}(C)$
4. And we require $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$

Example. We roll two dice. Consider events:

E_1: the first die is 4; E_2: the second die is 3; S_7: the sum of dice is 7.

$$\mathbb{P}(E_1) = \frac{1}{6};\ \mathbb{P}(E_2) = \frac{1}{6};\ \mathbb{P}(S_7) = \frac{1}{6};\ \mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1 \cap S_7) = \mathbb{P}(E_2 \cap S_7) = \frac{1}{36}.$$

Then we know that (E_1, E_2), (E_1, S_7) and (E_2, S_7) are independent.

$$\mathbb{P}(E_1 \cap E_2 \cap S_7) = \frac{1}{36} \neq \mathbb{P}(E_1) \times \mathbb{P}(E_2) \times \mathbb{P}(S_7)$$

Then we know that (E_1, E_2, S_7) are not independent.

We call n events $A_1, A_2, \cdots, A_n$ independent events if for every subset $\{j_1, j_2, \cdots, j_t\}$ of $\{A_1, \cdots, A_n\}$, the probability of the intersection is the product of their probabilities:

$$\mathbb{P}(A_{j_1} \cap A_{j_2} \cap \cdots \cap A_{j_t}) = \mathbb{P}(A_{j_1})\mathbb{P}(A_{j_2}) \cdots \mathbb{P}(A_{j_t})$$

If n events $A_1, A_2, \cdots, A_n$ are independent events, then the independence is preserved when we replace some event(s) by their complements, intersection, unions.

For example, if A, B, C, D are independent events, then $A \cup B$ and $C \cap D$ are also independent events.

Example. Alice wins 60% of her ping pong matches against Bob. They meet for a 3 match playoff. What are the chances that Alice will win the playoff?

Solution For Alice, we have

$$\Omega = \{\text{WWW, WWL, WLW, LWW, WLL, LWL, LLW, LLL}\}$$

$$A = \{\text{WWW, WWL, WLW, LWW}\}$$

$$\mathbb{P}(A) = 0.6^3 + 0.6^2 * 0.4 * 3 = \frac{81}{125}$$

Remark. One should also be aware that conditioning can affect independence. For example, consider two coins: coin A with bias $\mathbb{P}(H|\text{coin } A) = 0.9$ and coin B with bias $\mathbb{P}(H|\text{coin } B) = 0.1$. Suppose we randomly select one of the two coins with equal probability. Then, the probability of getting heads on the 11th toss is:

$$\mathbb{P}(\text{toss } 11 = H) = \mathbb{P}(A)\mathbb{P}(H_{11}|A) + \mathbb{P}(B)\mathbb{P}(H_{11}|B) = 0.5 \times 0.9 + 0.5 \times 0.1 = 0.5.$$

However, if we condition on the event that the first 10 tosses resulted in heads, then it is highly likely that the chosen coin is coin A. In fact, under this condition, we can be certain that we are using coin A. Therefore, the probability of getting heads on the 11th toss given that the first 10 tosses were heads is:

$$\mathbb{P}(\text{toss } 11 = H|\text{first 10 tosses are heads}) = 0.9.$$

3.3.3 Conditional Independence

Definition 3.3.2 (Conditional Independence)**.** Events A and B are independent conditioned on event F if

$$\mathbb{P}(A \cap B|F) = \mathbb{P}(A|F)\mathbb{P}(B|F)$$

Note that the above is equivalent to

$$\mathbb{P}(A|B \cap F) = \mathbb{P}(A|F)$$

Example.

Today	Tomorrow
Sunny	80% Sunny, 20% Rainy
Rainy	40% Sunny, 60% Rainy

If Today(Monday) is rainy, what is the probability that Wednesday will also be sunny?

Solution: We suppose weather on Monday and Wednesday are independent conditioned on weather on Tuesday. Let M be the event that Monday is sunny, same for T and W. Then:

$$\begin{aligned}\mathbb{P}(W|M^c) &= \mathbb{P}(W|M^c \cap T)\mathbb{P}(T|M^c) + \mathbb{P}(W|M^c \cap T^c)\mathbb{P}(T^c|M^c) \\ &= 80\% \times 40\% + 40\% \times 60\% \\ &= 0.56\end{aligned}$$

Example (The King's Sibling)**.** The king comes from a family of two children. What is the probability that his sibling is female?

Solution: Given that the conditions are vague, we need to further set up some conditions for this event.

For example, we assume that only boys have precedence. By intuition, it seems the answer is $\frac{1}{2}$. However, there are 4 possibilities for such an event, including (Boy, Boy), (Boy, Girl), (Girl, Boy), and (Girl, Girl), each with probability $\frac{1}{4}$.

We can ignore the (Girl, Girl) combination since it is stated that there is a king. Then, we have:

$$\mathbb{P}(\text{Sibling is female}) = \frac{\frac{1}{4} \times 2}{\frac{3}{4}} = \frac{2}{3}$$

However, one should note that the probability model we use depends on the conditions we set up or are given. For example, if royal families have different rules on having children, then the probability model would surely be different, leading to different results.

Remark. This example comes from MIT RES.6-012 Introduction to Probability, Spring 2018.

Chapter 4

Random Variables

4.1 Introduction

Definition 4.1.1 (Discrete Random Variable)**.** A discrete random variable assigns a discrete value to every outcome in the sample space Ω.

For example, we can assign the number of heads in tossing 3 coins to the random variable X. Then we have:

$$\mathbb{P}(X=0)=\frac{1}{8},\ \mathbb{P}(X=1)=\frac{3}{8},\ \mathbb{P}(X=2)=\frac{3}{8},\ \mathbb{P}(X=3)=\frac{1}{8}.$$

4.1.1 Probability Mass Function (PMF)

Definition 4.1.2 (Probability Mass Function)**.** The Probability Mass Function (PMF) $p : \mathbb{R} \to [0,1]$ of a discrete random variable X is the function

$$p_X(x) = \mathbb{P}(X = x)$$

We can describe the PMF by a table:

x	0	1	2	3
$p_X(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

For every random variable X, its probability mass function satisfies the following (based on the axioms of probability):

1. For every $x \in \mathbb{R}, p(x) \geq 0$ is non-negative
2. If $\mathcal{X}$ is the set of all possible values of X, then the PMF values on $\mathcal{X}$ will add up to 1:

$$\sum_{x \in X} p_X(x) = 1.$$

Example. We roll two 3-sided dice. Let random variable D be the difference between the output of the first and second dice. What is the PMF of random variable D? What is the probability that $D \geq 1$?

Solution: For sample space, we have

$$\Omega = \{(1,1),(1,2),(1,3),\cdots,(3,3)\} \Longrightarrow D = x, x \in [-2,+2]$$

For PMF, we have:

x	-2	-1	0	1	2
$p_X(x)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Then, for the probability that $D \geq 1$, we have

$$\mathbb{P}(D \geq 1) = p_X(1) + p_X(2) = \frac{2}{9} + \frac{1}{9} = \frac{3}{9}$$

4.2 Bernoulli Random Variable

4.2.1 Definition

A Bernoulli(p) random variable X shows the result of a trial where $X = 1$ for the success outcome with probability p and $X = 0$ for the failure outcome with probability $1 - p$.

$$X_i = \begin{cases} 1, & \text{if Experiment } i \text{ is succeeded ;} \\ 0, & \text{otherwise;} \end{cases}$$

with

$$\begin{cases} p_X(0) = 1 - p \\ p_X(1) = p \end{cases}$$

4.2.2 PMF of Bernoulli Random Variable

The probability mass function of a Bernoulli random variable can be easily found and comes in handy for some complicated calculations.

For a Bernoulli random variable, we have an indicator random variable of an event A, where $I_A = 1$ if and only if A occurs.

The PMF of a Bernoulli random variable is as follows:

$$p_{I_A}(1) = \mathbb{P}(I_A = 1) = \mathbb{P}(A)$$

4.3 Discrete Uniform Random Variable

A discrete uniform random variable takes values in a certain range, and each one of the values in that range has the same probability.

Such a variable is determined by two parameters a, b ($a \leq b$), which define the beginning and the end of the range.

Then, for every value of the random variable, the probability is $\frac{1}{b-a+1}$.

We also need to consider one special case where $a = b$. In this case, the random variable is deterministic, i.e., a constant value.

4.4 Binomial Random Variable

4.4.1 Definition

We call X a Binomial (n, p) Random Variable when X represents the number of successes over n independent trials, each with a success probability of p.

For example, if we toss n coins, the number of heads is Binomial$(n, \frac{1}{2})$

Example. We flip a coin 10 times and consider the random variable of the number of consecutive changes (HT or TH) in the 10 coin flips. What is this random variable?

Solution: We have a total of 9 trials, with the possible outcomes being HH, HT, TH, and TT. The probability of a consecutive change is $\frac{1}{2}$. This gives Binomial$(9, \frac{1}{2})$.

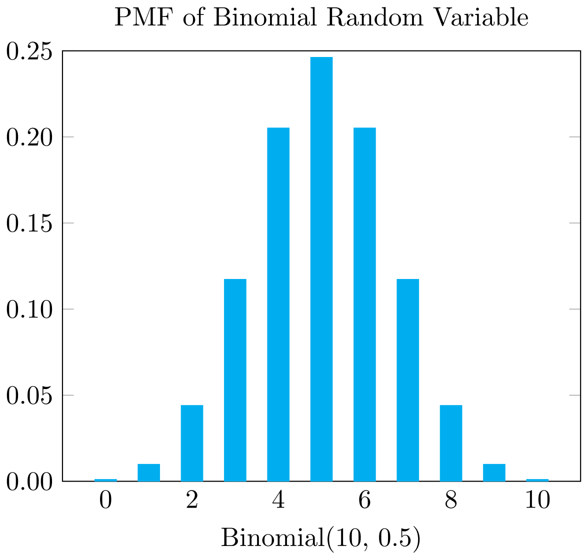
Example. We draw a 10-card hand from a 52-card deck. Let $N =$ Number of Aces among the picked cards. What is the random variable N?

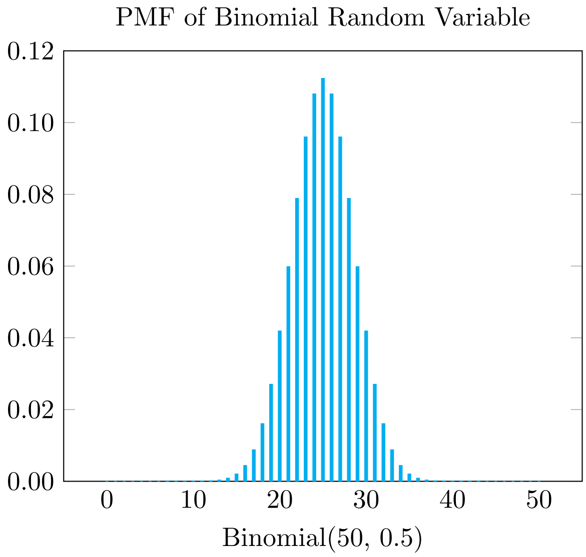
Solution: The definition states that the events must be independent. However, the second draw is influenced by the outcome of the first draw, so the events are not independent.

4.4.2 PMF of Binomial Random Variable

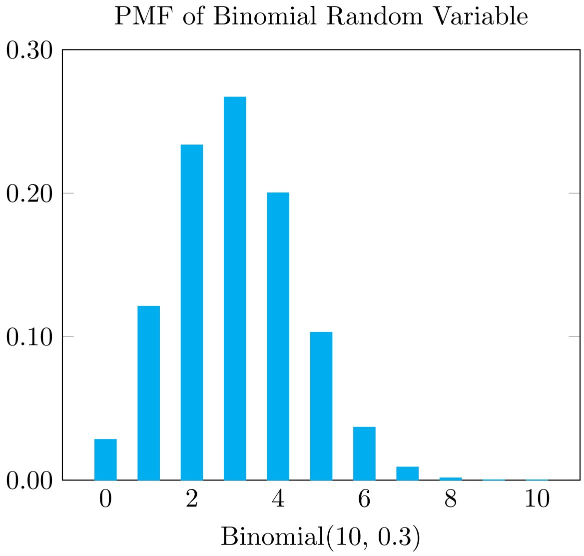
The probability mass function (PMF) of a Binomial(n, p) Random Variable is

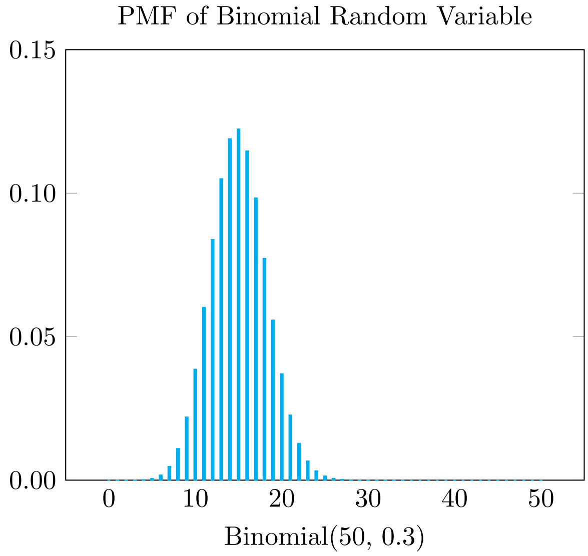
$$p_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$





From the figures above, we can observe that as the value of n increases, the graph is symmetric around the central point. For Binomial$(10, 0.5)$, the probability reaches its maximum at $x = 5$, which can also be calculated as 10×0.5. The same pattern holds for the graph on the right-hand side. This also makes sense because the probability of the event occurring is 0.5. Therefore, we can estimate that half of the events will occur, resulting in the maximum probability at $x = 5$.





For the two graphs above, it also holds true that the maximum value is reached at $10 \times 0.3 = 3$ or $50 \times 0.3 = 15$, due to the same reasoning. However, in this case, the graph is no longer symmetric around the central point.

Additionally, we observe that the graph shifts to the left when $p < 0.5$. Conversely, if $p > 0.5$, the graph shifts to the right.

Example. The Lakers and the Celtics meet for a 7-game playoff.

Lakers win 60% of the time. What is the probability that all 7 games are played? What is the probability that Lakers win the play-off in 6 games?

Solution: For all 7 games to be played, the first 6 games must occur, which means that both the Lakers and the Celtics win 3 games each. So X is a $\text{Binomial}(6, 0.6)$.

$$\mathbb{P}(X = 3) = \binom{6}{3} \times (0.6)^3 \times (1 - 0.6)^3 = \frac{864}{3125}$$

For the Lakers to win in 6 games, we cannot simply use $\mathbb{P}(X = 4)$ because it also includes the probability of the Lakers winning in 4 or 5 games. However, by fixing the Lakers to win the 6th game, we can define a new variable following $\text{Binomial}(5, 0.6)$. We then only need to calculate the probability of the Lakers winning 3 games in the first 5 games.

$$\mathbb{P}(X = 3) \times 0.6 = \binom{5}{3} \times (0.6)^3 \times (1 - 0.6)^2 \times 0.6 = \frac{648}{3125}$$

4.5 Geometric Random Variable

4.5.1 Definition

We call N a Geometric(p) Random Variable when X represents the first time of success over a series of independent trials $X_1, X_2, \cdots$, each with a success probability of p:

$$N = \text{first (smallest) } n \text{ such that } X_n = 1.$$

For example, we toss a coin until we see the first heads. The number of coin tosses to see the first heads is Geometric($\frac{1}{2}$).

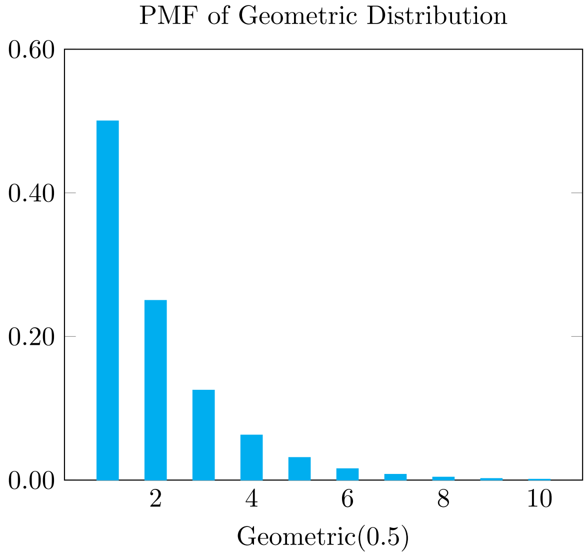
4.5.2 PMF of Geometric Random Variable

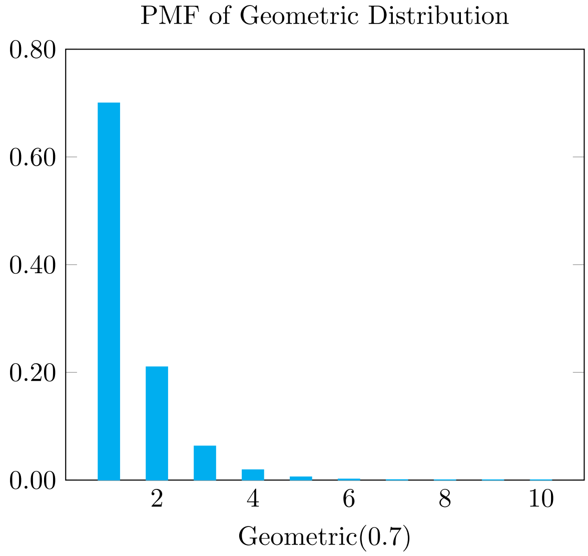
The probability mass function (PMF) of a Geometric(p) Random Variable is

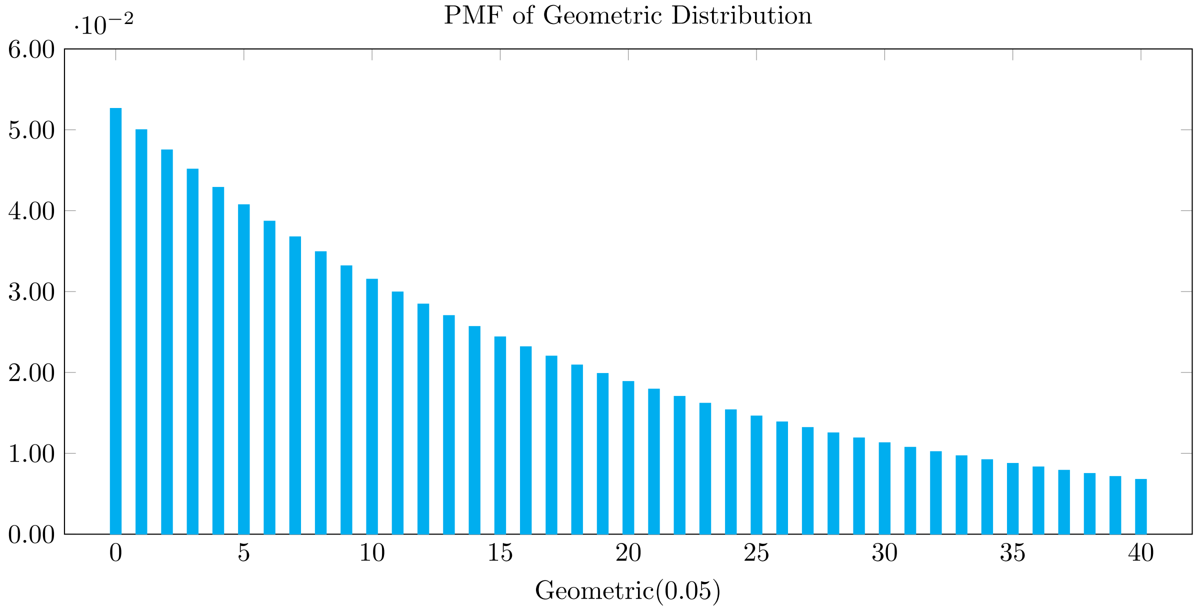
$$p_X(k) = \mathbb{P}(X = k) = p(1-p)^{k-1}$$

Axioms of probability also holds for PMF of geometric random variable. Therefore,

$$\sum_{k=1}^{\infty} p(1-p)^{k-1} = 1$$







Example. You keep rolling dice until you roll a 6. What is the probability that you rolled more than 10 times?

Solution: Define X is a Geometric($\frac{1}{6}$), then we have

$$\mathbb{P}(X \geq 11) = \sum_{k=11}^{\infty} p(1-p)^{k-1} = \sum_{k=11}^{\infty} \frac{1}{6} \times (\frac{5}{6})^{k-1} = \left(\frac{5}{6}\right)^{10}$$

By generalizing the above, we have

$$\sum_{k=m}^{\infty} p(1-p)^{k-1} = (1-p)^{m-1}$$

This formula means that if we want to find the probability of success occurring at or after the m-th trial, we can treat the trials before the m-th one as all failures. This simplifies the calculation.

4.6 Cumulative Distribution Function (CDF)

Definition 4.6.1 (Cumulative Distribution Function (CDF))**.** For a random variable X, its cumulative distribution function (CDF) $F(x)$ is:

$$F(x) = \mathbb{P}(X \leq x)$$

From the definition of PMF, we have

$$F(x) = \sum_{\substack{k\in\mathcal{X} \\ k\leq x}} p(x)$$

For a Geometric(p) random variable X, the CDF will be

$$F(k) = \mathbb{P}(X \leq k) = 1 - \mathbb{P}(X > k) = 1 - (1-p)^k$$

For a Binomial(n, p) random variable X, the CDF will be

$$F(k) = \mathbb{P}(X \leq k) = (1-p)^n + n(1-p)^{n-1}p + \cdots + \binom{n}{k}(1-p)^{n-k}p^k$$

Example. You keep rolling dice until you roll a 6. What is the probability that you roll the dice an even number of times?

Solution 1:

$$\begin{aligned}
\mathbb{P}(X\text{is even}) &= \mathbb{P}(X = 2) + \mathbb{P}(X = 4) + \mathbb{P}(X = 6) + \cdots \\
&= p(2) + p(4) + p(6) + \cdots \\
&= \frac{1}{6} \times \frac{5}{6} + \frac{1}{6} \times \left(\frac{5}{6}\right)^3 + \frac{1}{6} \times \left(\frac{5}{6}\right)^5 + \cdots
\end{aligned}$$

Solution 2: Let's define event A as the event that an even number is rolled. Next, we define another event B, which is independent of event A, such that using conditional probability, we have $\mathbb{P}(A) = \mathbb{P}(A|B)$.

For example, we can define $B = \{X = 1 \text{ or } X = 2\}$. Since A and B are independent, we can then proceed with the calculation.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(X = 2)}{\mathbb{P}(X = 1) + \mathbb{P}(X = 2)} = \frac{p(1-p)}{p + p(1-p)} = \frac{p-1}{p-2} = \frac{\frac{5}{6}}{\frac{11}{6}} = \frac{5}{11}$$

4.7 Poisson Random Variable

Example. Alice randomly sprinkles 25 chocolate chips on 5 cookies.

1. What is the random variable N on how many chips a cookie gets?

$$N \sim \text{ Binomial}(25, \frac{1}{5})$$

2. What is the probability a cookie gets no chips?

$$\mathbb{P}(N = 0) = \binom{25}{0} \times (\frac{1}{5})^0 \times (1 - \frac{1}{5})^{25} = 0.004$$

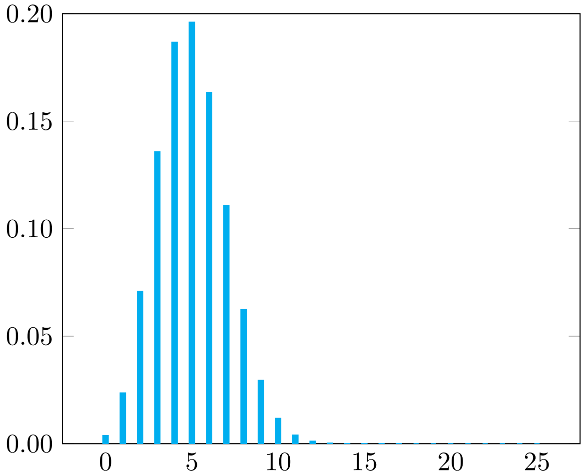
3. What is the probability a cookie gets exactly 5 chips?

$$\mathbb{P}(N = 5) = \binom{25}{5} \times (\frac{1}{5})^5 \times (1 - \frac{1}{5})^{20} = 0.196$$

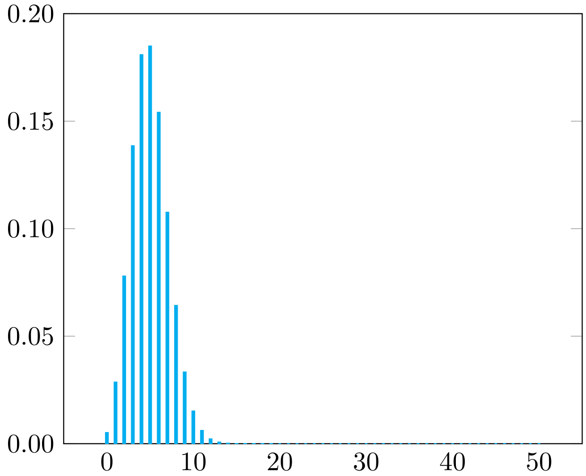
4. What is the probability a cookie gets at most 5 chips?

$$\mathbb{P}(N \leq 5) \approx 0.617$$

We now want to examine how the probability changes if we increase the values, for example, by having 250 chocolate chips and 50 cookies instead. Intuitively, the probability should not change significantly, and through calculation, the results are indeed similar to those obtained in the previous example.



Binomial(25, 0.2)



Binomial(50, 0.1)

From the above, we see that when we double the value, the PMF remains quite similar to the previous one. Additionally, the average rate of chocolate chips per cookie is also the same, i.e., $25 \times 0.2 = 50 \times 0.1 = 5$.

Poisson(5)	$\mathbb{P}(X = 0)$	$\mathbb{P}(X = 5)$	$\mathbb{P}(X \leq 5)$
Binomial(25, 0.2)	0.004	0.196	0.617
Binomial(50, 0.1)	0.005	0.185	0.616
Binomial(500, 0.01)	0.006	0.176	0.615

From the values above, we can see that the probability converges to a certain value. This is what we call a Poisson Random Variable. The values we obtain by using $25 \times 0.2, 50 \times 0.1$ represent the rate of Poisson random variable.

Definition 4.7.1 (Poisson Random Variable)**.** A Poisson(λ) random variable X has the PMF:

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k = 0, 1, 2, 3, \cdots, \text{where } \lambda \text{ is called the rate parameter.}$$

Theorem 4.7.1. Poisson(λ) is the limit approximation of Binomial($n, \frac{\lambda}{n}$) when $n \to \infty$:

$$\mathbb{P}(\text{Poisson}(\lambda) = k) = \lim_{n\to\infty} \mathbb{P}(\text{Binomial}(n, \frac{\lambda}{n}) = k)$$

Poisson random variable can be used when we have a large number of independent trials while the expected number of successes remains small (the rate of successes being constant).

Example. Suppose rain is falling on your head a rate of 3 drops/sec. What is the probability that you get

(1) no hits in the next second?

Solution:

$$\mathbb{P}(\text{Poisson}(3) = 0) = e^{-3}\frac{3^0}{0!} = e^{-3} \approx 0.050$$

(2) at least 3 hits in the next second?

Solution:

$$\mathbb{P}(\text{Poisson}(3) \geq 3) = 1 - \mathbb{P}(\text{Poisson}(3) < 3) = 1 - e^{-3}\frac{3^0}{0!} - e^{-3}\frac{3^1}{1!} - e^{-3}\frac{3^2}{2!} = 1 - \frac{17}{2e^3} \approx 0.576$$

(3) exactly 10 hits in the next 5 seconds?

Solution:

$$\mathbb{P}(\text{Poisson}(3 \times 5) = 10) = e^{-15}\frac{15^{10}}{10!} = 0.049$$

4.8 Properties of Random Variables

4.8.1 Expected Value

Definition 4.8.1 (Expected Value)**.** The expected value (expectation) of a random variable X with PMF $p(x)$ is

$$\mathbb{E}[X] = \sum_x x p_X(x)$$

For example, the expected value of random variable X as the number of heads in tossing 1 coin:

$$\mathbb{E}[X] = \frac{1}{2} \times 0 + \frac{1}{2} \times 1 = \frac{1}{2}$$

Now, instead of flipping one coins, we flip 3 coins. For PMF, we have

$$p(k) = \binom{3}{k}(\frac{1}{2})^k(\frac{1}{2})^{3-k} = \frac{\binom{3}{k}}{8}$$

Then, the expected value of random variable X as the number of heads in tossing 3 coin:

$$\mathbb{E}[X] = \frac{1}{8} \times 0 + \frac{3}{8} \times 1 + \frac{3}{8} \times 2 + \frac{1}{8} \times 3 = \frac{3}{2}$$

The expectation is the average value the random variable takes when the experiment is done many times.

Example. Find the expected value of random variable F as the face value of a six-sided die.

Solution:

$$\mathbb{E}[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} = \frac{7}{2}$$

Example. We play this game that we roll three dice, and you win k dollars if we see $k \geq 1$ "Two" outcomes. You lose 1 dollar if we see no "Two" outcomes.

Should you play this game?

Solution:

$$G = -1, \mathbb{P}(-1) = (\frac{5}{6})^3 = \frac{125}{216}; \quad G = 1, \mathbb{P}(1) = \binom{3}{1}(\frac{1}{6})(\frac{5}{6})^2 = \frac{25}{72};$$

$$G = 2, \mathbb{P}(2) = \binom{3}{2}(\frac{1}{6})^2(\frac{5}{6})^1 = \frac{5}{72}; \quad G = 3, \mathbb{P}(3) = (\frac{1}{6})^3 = \frac{1}{216};$$

$$\mathbb{E}[X] = -1 \times \frac{125}{216} + 1 \times \frac{25}{72} + 2 \times \frac{5}{72} + 3 \times \frac{1}{216} = -\frac{17}{216} \approx -0.079$$

Therefore, it's better not to play this game.

4.8.2 Function of Random Variables

If X is a random variable with PMF p_X, then $Y = f(X)$ will also be a random variable with PMF p_Y:

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

The expected value of $f(X)$ for a function f and random variable X is

$$\mathbb{E}[f(X)] = \sum_x f(x)p_X(x)$$

Example. Consider a random variable X with the following PMF:

x	0	1	2
$p(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$y = X - 1$	-1	0	1
$p(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$z = (X-1)^2$	0	1
$p(z)$	$\frac{1}{3}$	$\frac{2}{3}$

Using $Y = (X-1)^2$ as an example, we have:

$$p_Y(1) = \sum_{x:(x-1)^2=1} p_X(x) = p_X(0) + p_X(2) = \frac{2}{3}$$

Additionally, for the expected values of the above PMF, we have:

$$\mathbb{E}[X] = 0 \times \frac{1}{3} + 1 \times \frac{1}{3} + 2 \times \frac{1}{3} = 1$$
$$\mathbb{E}[X-1] = -1 \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$$
$$\mathbb{E}[(X-1)^2] = 0 \times \frac{1}{3} + 1 \times \frac{2}{3} = \frac{2}{3}$$

Alternatively, we can use the following method to directly calculate the expected values:

$$\mathbb{E}[X-1] = \sum_{x \in \{0,1,2\}} (x-1)p(x) = (0-1) \times \frac{1}{3} + (1-1) \times \frac{1}{3} + (2-1) \times \frac{1}{3} = 0$$
$$\mathbb{E}[(X-1)^2] = \sum_{x \in \{0,1,2\}} (x-1)^2 p(x) = (0-1)^2 \times \frac{1}{3} + (1-1)^2 \times \frac{1}{3} + (2-1)^2 \times \frac{1}{3} = \frac{2}{3}$$

Remark.

$$\mathbb{E}[f(x)] \neq f(\mathbb{E}[X])$$

Example. Suppose the distance between place A and place B is 1 km. There is a 60% chance that the weather will be sunny; in that case, you will walk from A to B at a speed of 5 km/h. Conversely, there is a 40% chance that the weather will be rainy; in that case, you will take a shuttle traveling at a speed of 30 km/h. Find the expected value of time T and speed V.

Solution: Given that $V = \frac{1}{T}$,

$$\mathbb{E}[V] = 0.6 \times 5 + 0.4 \times 30 = 15 \text{ km/h}$$

$$\mathbb{E}[T] = \mathbb{E}[\frac{1}{V}] = 0.6 \times \frac{1}{5} + 0.4 \times \frac{1}{30} = \frac{2}{15} \text{ h} \neq \frac{1}{\mathbb{E}[V]}$$

Chapter 5

Expectation, Variance and Conditional PMF

5.1 Expectation

5.1.1 Joint Probability Mass Function

The joint PMF of random variables X, Y is the bivariate function

$$p(x,y) = \mathbb{P}(X = x, Y = y)$$

Example. There is a bag with 4 cards, with face values 1, 2, 3, and 4, respectively. You draw two cards without replacement. What is the joint PMF of the card values? Let Z be the sum of the card values. Then what is the PMF of Z? What is the expected value of Z?

Solution: Let X, Y represent the values of the first and second cards drawn respectively.

Joint PMF for X, Y:

$$p(x,y) = \mathbb{P}(X = x, Y = y) = \begin{cases} 1/12 & \text{if } x \neq y,\ x, y \in \{1,2,3,4\}; \\ 0 & \text{if } x = y\ . \end{cases}$$

PMF for Z:

Z	3	4	5	6	7
$p(Z)$	$\frac{2}{12}$	$\frac{2}{12}$	$\frac{4}{12}$	$\frac{2}{12}$	$\frac{2}{12}$

$$\mathbb{E}[Z] = 3 \times \frac{2}{12} + 4 \times \frac{2}{12} + 5 \times \frac{4}{12} + 6 \times \frac{2}{12} + 7 \times \frac{2}{12} = 5$$

However, due to the symmetry around 5, we can directly observe that the expected value is 5.

Example. Following the question setup in the previous example, the cards are now drawn with replacement. Let Z be the sum of the card values.

Solution: PMF for Z:

Z	2	3	4	5	6	7	8
$p(Z)$	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$

$$\mathbb{E}[Z] = 2 \times \frac{1}{16} + 3 \times \frac{2}{16} + 4 \times \frac{3}{16} + 5 \times \frac{4}{16} + 6 \times \frac{3}{16} + 7 \times \frac{2}{16} + 8 \times \frac{1}{16} = 5$$

Again, due to the symmetry around 5, we can directly observe that the expected value is 5.

If X, Y are two random variable with Joint PMF p_{XY}, then $Z = f(X, Y)$ will also be a random variable with PMF p_Z:

$$p_Z(z) = \sum_{x,y:f(x,y)=z} p_{XY}(x, y)$$

The Expected Value of $f(X, Y)$ for a function f and random variables X, Y is:

$$\mathbb{E}[f(X, Y)] = \sum_{x,y} f(x, y)p_{XY}(x, y)$$

In the previous example, we can see that:

$$p_Z(6) = \sum_{x,y:X+Y=6} p_{XY}(x, y) = p_{XY}(2, 4) + p_{XY}(3, 3) + p_{XY}(4, 2)$$

Remark. We use p_{XY} to represent Joint PMF, and use p_X or p_Y to represent marginal PMF. We can also convert from joint PMF to marginal(individual) PMF:

$$p_X(x) = \sum_y p_{XY}(x, y) \qquad p_Y(y) = \sum_x p_{XY}(x, y)$$

i.e., summing over all the values of y or x to find the marginal PMF for X or Y

5.1.2 Linearity of Expectation

The Expected Value of $X + Y$, i.e. the sum of random variables X, Y satisfies:

$$\mathbb{E}[\alpha X + \beta Y] = \alpha\mathbb{E}[X] + \beta\mathbb{E}[Y]$$

Example. We again follow the question setup in the previous example.

$$p_X = \frac{1}{4},\ x \in \{1, 2, 3, 4\} \qquad p_Y = \frac{1}{4},\ y \in \{1, 2, 3, 4\}$$

Then we have

$$\mathbb{E}[X] = \mathbb{E}[Y] = 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} = \frac{5}{2} \Longrightarrow \mathbb{E}[X] + \mathbb{E}[Y] = 5 = \mathbb{E}[X + Y]$$

5.1.3 Binomial Random Variable

For the expectation of Bernoulli random variable, we have:

$$\mathbb{E}[X] = 0 \times (1 - p) + 1 \times p = p.$$

By observation, we see that a Binomial(n, p) random variable is the sum of n independent Bernoulli(p) random variables $X_1, X_2, \cdots, X_n$:

$$X = X_1 + X_2 + \cdots + X_n$$

where

$$X_i = \begin{cases} 1, & \text{if Experiment } i \text{ is succeeded ;} \\ 0, & \text{otherwise;} \end{cases}$$

By the linearity of expectation, we see that

$$\mathbb{E}[X] = \mathbb{E}[X_1 + X_2 + \cdots + X_n] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = np$$

We now observe that in the PMF given in Chapter 4.2.2, the PMF of a Binomial random variable attains its maximum value at $n \times p$, which is also the expected value of the PMF. While this is not generally true for all random variables, it holds for Binomial random variables.

5.1.4 Poisson Random Variable

The expected value of a Poisson(λ) random variable X is

$$\mathbb{E}[X] = \lambda$$

This is quite intuitive since in Poisson(λ) random variable, λ is defined as the average rate. Therefore, it also aligns with the definition of the expected value.

Proof. For Poisson(λ) we know that

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{k=0}^{\infty} kp(k) \\
&= \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda \times \lambda^{k-1}}{k!} \\
&= \sum_{k=1}^{\infty} ke^{-\lambda}\frac{\lambda \times \lambda^{k-1}}{k!} \qquad (\text{let } k' = k-1) \\
&= \lambda \times \sum_{k'=0}^{\infty} e^{-\lambda}\frac{\lambda^{k'}}{(k')!} \\
&= \lambda \times 1 \\
&= \lambda
\end{aligned}$$

■

5.1.5 Geometric Random Variable

The expected value of a Geometric(p) random variable X is

$$\mathbb{E}[X] = \frac{1}{p}$$

We will prove this in the later part of this chapter.

5.1.6 Uniform Random Variable

Suppose we have a uniform random variable on $0, 1, \ldots, n$. Then, by definition, the expected value is:

$$\mathbb{E}[X] = 0 \cdot \frac{1}{1+n} + 1 \cdot \frac{1}{1+n} + \cdots + n \cdot \frac{1}{1+n} = \frac{1}{n+1}(0 + 1 + \cdots + n) = \frac{n}{2}$$

5.2 Variance

5.2.1 Definition

In the stock market, for each stock one of the following outcomes will happen:

- Stock doubles in value with probability $\frac{1}{2}$
- Stock loses all its value with probability $\frac{1}{2}$

Also, different stocks perform independently. We want to invest \$25 based on one of these scenarios:

1. Scenario 1 (X): Invest all \$25 on one stock.
2. Scenario 2 (Y): Keep all \$25 without investing
3. Scenario 3 (Z): Invest \$1 on each of 25 different stocks

$$X = \begin{cases} 50, & w.p.\ \frac{1}{2}, \\ 0, & w.p.\ \frac{1}{2}. \end{cases}; \qquad Y = 25\ w.p.\ 1; \qquad Z = 2 \times \text{ Binomial}(25, \frac{1}{2})$$

Then we have $\mathbb{E}[X] = 25, \mathbb{E}[Y] = 25, \mathbb{E}[Z] = 25$. Therefore, we cannot determine which investment strategy we should choose.

Definition 5.2.1 (Variance and Standard Deviation)**.** Consider random variable X with expected value $\mu = \mathbb{E}[X]$. Then, we define the variance of X as

$$\text{Var}[X] := \mathbb{E}\left[(X-\mu)^2\right]$$

Furthermore, we define the standard deviation of X to be

$$\sigma_X := \sqrt{\text{Var}[X]} = \sqrt{\mathbb{E}\left[(X-\mu)^2\right]}$$

Note that variance measures how close X and $\mathbb{E}[X]$ are for a typical outcome of X.

So for the example above, we have

$\text{Var}[X] = \mathbb{E}[(X-25)^2] = \frac{1}{2}(50-25)^2 + \frac{1}{2}(0-25)^2 = 625 \Longrightarrow \sigma = \sqrt{625} = 25$

$\text{Var}[Y] = \mathbb{E}[(Y-25)^2] = 1 \times (25-25)^2 = 0 \Longrightarrow \sigma = \sqrt{0} = 0$

$\text{Var}[Z] = 2^2 \times 25 \times \frac{1}{2} \times (1-\frac{1}{2}) = 25 \Longrightarrow \sigma = \sqrt{25} = 5$

We can see that scenario 1 has the highest risk.

We have another formula for variance:

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Proof.

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}\left[(X-\mu)^2\right] \\
&= \mathbb{E}\left[X^2 - 2X\mu + \mu^2\right] \\
&= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 \\
&= \mathbb{E}[X^2] - 2\mathbb{E}[X] \times \mathbb{E}[X] + \mathbb{E}[X]^2 \\
&= \mathbb{E}[X^2] - \mathbb{E}[X]^2
\end{aligned}$$

■

Using this formula, we can find the variance from the previous example.

$$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{2} \times 0^2 + \frac{1}{2} \times 50^2 - (25^2) = 625$$

Example. We roll a die. What are the expected value and variance?

Solution:

$\mathbb{E}[X] = \frac{1}{6}(1+2+3+4+5+6) = \frac{7}{2}$

$\text{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{6}(1+4+9+16+25+36) - (\frac{7}{2})^2 = \frac{35}{12}$

5.2.2 Variance of Bernoulli Random Variable

From previous we have

$$\mathbb{E}[X] = p$$

By definition, we have

$$\operatorname{Var}[X] = \sum_x (x - \mathbb{E}[X])^2 p_X(x) = (1-p)^2 \cdot p + (0-p)^2 \cdot (1-p) = p(1-p)$$

5.2.3 Variance of Binomial Random Variable

Suppose X is a Binomial(n, p) random variable. Then,

$$\operatorname{Var}[X] = np(1-p)$$

Proof. We know that $X = X_1 + X_2 + \cdots + X_n$ is the sum of n independent Bernoulli(p) random variables. Therefore,

$$\begin{aligned}
\operatorname{Var}[X] &= \operatorname{Var}[X_1] + \operatorname{Var}[X_2] + \cdots + \operatorname{Var}[X_n] \\
&= n \operatorname{Var}[X_1] \\
&= n(p - p^2) \\
&= np(1-p)
\end{aligned}$$

■

To summarize, if you know $X \sim$ Binomial(n, p), you can find

$$\begin{aligned}
\mathbb{E}[X] &= np \\
\operatorname{Var}[X] &= np(1-p) \\
p &= 1 - \frac{\operatorname{Var}[X]}{\mathbb{E}[X]} \\
n &= \frac{\mathbb{E}[X]}{1 - \frac{\operatorname{Var}[X]}{\mathbb{E}[X]}}
\end{aligned}$$

5.2.4 Variance of Uniform Random Variable

For the case where the uniform random variable takes values from 1 to n, we can deduce the variance by:

$$\operatorname{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{n+1}(0^2 + 1^2 + \cdots + n^2) - \frac{n^2}{4} = \frac{1}{12}n(n+2)$$

However, for general cases where the random variable takes values from a to b, we can take $n = b - a$, then we have:

$$\operatorname{Var}[X] = \frac{1}{12}(b-a)(b-a+2)$$

5.2.5 Variance of Poisson Random Variable

Suppose Y is a Poisson(λ) random variable. Then,

$$\operatorname{Var}[Y] = \lambda$$

Informal Proof:

$$\begin{aligned}
\text{Poisson}(\lambda) &= \lim_{n \to \infty} \text{Binomial}(n, \frac{\lambda}{n}) \\
\operatorname{Var}[Y] &= \lim_{n \to \infty} n \cdot \frac{\lambda}{n} \cdot (1 - \frac{\lambda}{n}) \\
\operatorname{Var}[Y] &= \lambda
\end{aligned}$$

5.3 Conditional PMF

5.3.1 Definition

Definition 5.3.1 (Conditional PMF). The conditional PMF $p_{X|Y}(\cdot|\cdot)$ of X given Y is defined as

$$p_{X|Y}(x|y) = \frac{p_{X,Y}(x,y)}{P_Y(y)} = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} = \mathbb{P}(X=x|Y=y)$$

Example. We roll two 3-sided dice.

1. What is the PMF of the sum given the first roll?

Solution: Let $X, Y \in \{1,2,3\}$, $S = X+Y \in \{2,3,4,5,6\}$. For the joint PMF of $p_{X,S}(x,s)$, we have:

$X\backslash S$	2	3	4	5	6
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0	0
2	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0
3	0	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

;

X	1	2	3
$p(X)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

where we can use $p_{X,S}(x,s) = \mathbb{P}(X=x, S=s) = \mathbb{P}(X=x, Y=y)$ to find the probability.

For example, for $p_{X,S}(1,2) = \mathbb{P}(X=1, S=2) = \mathbb{P}(X=1, Y=1) = (\frac{1}{3})^2 = \frac{1}{9}$

Then, for $p_{S|X}(s|x)$, we have

$X\backslash S$	2	3	4	5	6
1	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0	0
2	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	0
3	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

For the table above, if we sum the values in each row, we see that the sum equals 1. This is because we conditioned on a specific value of x in each row.

2. What is the PMF of the first roll given the sum?

Solution: For the joint PMF of $p_{X,S}(x,s)$, we have:

$X\backslash S$	2	3	4	5	6
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0	0
2	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$	0
3	0	0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

;

S	2	3	4	5	6
$p(X)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{1}{9}$

Then, for $p_{X|S}(x|s)$, we have

$X\backslash S$	2	3	4	5	6
1	1	$\frac{1}{2}$	$\frac{1}{3}$	0	0
2	0	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{2}$	0
3	0	0	$\frac{1}{3}$	$\frac{1}{2}$	1

For the table above, if we sum the values in each column, we see that the sum equals 1. This is because we conditioned on a specific value of x in each row.

Definition 5.3.2 (Conditional Expectation). The conditional expectation $\mathbb{E}[X|Y=y]$ of X given $Y=y$ is defined as

$$\mathbb{E}[X|Y=y] = \sum_x x \cdot p_{X|Y}(x|y).$$

Remark. For a fixed y, $p_{X|Y}(\cdot|y)$ is a PMF as a function of X.

Theorem 5.3.1 (Total Expectation Theorem)**.** For random variables X, Y, the following holds:

$$\mathbb{E}[X] = \sum_y \mathbb{P}(Y = y)\mathbb{E}[X|Y = y] \quad (\text{or} \quad \mathbb{E}[X] = \mathbb{E}_Y\left[\mathbb{E}[X|Y]\right])$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \sum_x x \cdot p_X(x) \\
&= \sum_x x \cdot \mathbb{P}(X = x) \\
&= \sum_x x \cdot \left(\sum_y \mathbb{P}(Y = y)\mathbb{P}(X = x|Y = y) \right) \quad (\text{partition } X \text{ to } Y) \\
&= \sum_y \mathbb{P}(Y = y) \sum_x x \cdot p_{X|Y}(x|y). \\
&= \sum_y \mathbb{P}(Y = y)\mathbb{E}[X|Y = y]
\end{aligned}$$

■

Total expectation theorem can be equivalently shown for disjoint events $A_1, A_2, \cdots, A_k$ partitioning the sample space $A_1 \cup \cdots \cup A_k = \Omega$ as

$$\mathbb{E}[X] = \sum_{i=1}^{k} \mathbb{P}(A_i)\mathbb{E}[X|A_i].$$

Example. You flip 10 coins. What is the expected number of heads given that there is at least one heads?

Solution: Let X be the number of heads, $A = \{X \geq 1\}, A^c = \{X = 0\}$. Then we have

$$\begin{aligned}
\mathbb{E}[X] &= p(A)\mathbb{E}[X|A] + p(A^c)\mathbb{E}[X|A^c] \\
10 \times \frac{1}{2} &= \left(1 - (\frac{1}{2})^{10}\right) \times \mathbb{E}[X|A] + (\frac{1}{2})^{10} \times 0 \\
\mathbb{E}[X|A] &= \frac{5}{1 - (\frac{1}{2})^{10}}
\end{aligned}$$

5.3.2 Geometric Random Variable

In a geometric random variable, the number of failed trials before a success does not affect the number of future trials needed for success. This is called the memoryless property. With conditional probability, we say that the number of remaining coin tosses, conditioned on tails in the first n tosses, is geometric with parameter p, i.e.

$$\mathbb{P}_{X-1|X>1}k = \mathbb{P}_X k = \mathbb{P}_{X-n|X>n}k$$

Now we can prove the expected value of Geometric(p) random variable.

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{P}(X > 1)\mathbb{E}[X|X > 1] + \mathbb{P}(X = 1)\mathbb{E}[X|X = 1] \\
\mathbb{E}[X] &= (1 - \mathbb{P}(X = 1))\mathbb{E}[X|X > 1] + \mathbb{P}(X = 1) \times 1 \\
\mathbb{E}[X] &= (1 - \mathbb{P}(X = 1))(1 + \mathbb{E}[X]) + \mathbb{P}(X = 1) \\
\mathbb{E}[X] &= \frac{1}{\mathbb{P}(X = 1)} = \frac{1}{p}
\end{aligned}$$

■

Consider a Geometric(p) random variable X. To find the variance of Geometric(p) random variable, we can use the above example by letting $A = \{X > 1\}, A^c = \{X = 1\}$:

$$\begin{aligned}
\mathbb{E}[X^2] &= p(A^c)\mathbb{E}[X^2|A^c] + p(A)\mathbb{E}[X^2|A] \\
\mathbb{E}[X^2] &= p \times 1^2 + (1-p) \times \mathbb{E}[X^2|X > 1] \\
\mathbb{E}[X^2] &= p \times 1^2 + (1-p) \times \mathbb{E}[(X+1)^2] \\
\mathbb{E}[X^2] &= p \times 1^2 + (1-p) \times (\mathbb{E}[X^2] + \frac{2}{p} + 1) \\
\mathbb{E}[X^2] &= \frac{2-p}{p^2}
\end{aligned}$$

Then, we have

$$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$$

5.4 Independent Random Variable

5.4.1 Introduction

Definition 5.4.1 (Independent Random Variable)**.** X and Y are called independent random variables if every outcome pair $X = x$ and $Y = y$ are independent events for all x, y values:

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \text{or} \quad p_{X,Y}(x, y) = p_X(x)p_Y(y)$$

Random Variables X, Y are independent if and only if for every outcome y the conditional PMF $p_{X|Y}(\cdot|y)$ is the same as X's PMF $p_X(\cdot)$:

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

Example.

(a) Let X, Y be the face values of two 4-sided dice. Are X and Y independent?

Solution:

$$p_{X,Y}(x, y) = \frac{1}{16} = \frac{1}{4} \times \frac{1}{4} = p_X(x)p_Y(y) \quad (x, y \in \{1, 2, 3, 4\})$$

(b) How about $Z = \max(X, Y)$ and $W = \min(X, Y)$?

Solution:

$$p_{Z,W}(2, 3) = 0; \quad p_Z(z) = \frac{3}{16}; \quad p_W(w) = \frac{3}{16}; \quad p_Z(z)p_W(w) = \frac{3}{16} \times \frac{3}{16} \neq 0$$

Theorem 5.4.1. X and Y are independent if and only if for every function f, g we have

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)]\mathbb{E}[g(Y)]$$

In particular, if X and Y are independent, then

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

To show this, we can start from the expected value of joint PMF,

$$\begin{aligned}
\mathbb{E}[f(X)g(Y)] &= \sum_{x,y} p_{X,Y}(x, y)f(x)g(y) \\
&= \sum_{x,y} p_X(x)p_Y(y)f(x)g(y) \\
&= \sum_{x} p_X(x)f(x) \sum_{y} p_Y(y)g(y) \\
&= \mathbb{E}[f(X)]\mathbb{E}[g(Y)]
\end{aligned}$$

Note that $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ is not enough to guarantee X and Y are independent. For example, for random variable X with $\mathbb{P}(X = -1) = \mathbb{P}(X = 0) = \mathbb{P}(X = 1) = \frac{1}{3}$, X and $Y = X^2$ satisfy

$$\mathbb{E}[XY] = \mathbb{E}[X^3] = 0 = \mathbb{E}[X]\mathbb{E}[Y]$$

However, we can see that

$$p(X = 1, Y = 0) = 0,\ p(X = 1) = \frac{1}{3},\ p(Y = 0) = \frac{1}{3},\ p(X = 1) \times p(Y = 0) \neq p(X = 1, Y = 0).$$

5.4.2 Covariance

Definition 5.4.2 (Covariance)**.** The covariance of random variables X, Y with expected values $\mu_X = \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$ is defined as

$$\mathrm{Cov}[X, Y] := \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

The covariance can also be found using the formula

$$\mathrm{Cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

Remark. We call random variable X, Y uncorrelated if their covariance is zero: $\mathrm{Cov}[X, Y] = 0$ or equivalently $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$.

Therefore, every independent X, Y will be uncorrelated, but the converse of this statement is not always true.

To show this, we proceed as follows:

$$\begin{aligned}
\mathrm{Cov}[X, Y] &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] \\
&= \mathbb{E}[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\
&= \mathbb{E}[XY] - \mathbb{E}[X\mu_Y] - \mathbb{E}[Y\mu_X] + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_Y\mathbb{E}[X] - \mu_X\mathbb{E}[Y] + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mu_X\mu_Y \\
&= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]
\end{aligned}$$

If $X = Y$, we see that $\mathrm{Cov}[X, X] := \mathbb{E}[(X - \mu_X)(X - \mu_X)] = \mathbb{E}[(X - \mu_X)^2] = \mathrm{Var}[X]$.

Example. There is a bag with 3 cards, with face values 1, 2, and 3.

1. You draw two cards with replacement. X, Y are the face values of the first and second cards. What is $\mathbb{E}[XY]$?

Solution 1:

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] = \frac{1}{3}(1 + 2 + 3) \times \frac{1}{3}(1 + 2 + 3) = 4$$

Solution 2:

$$\mathbb{E}[XY] = \frac{1}{9}((1 + 2 + 3)(1 + 2 + 3)) = 4$$

2. You draw two cards without replacement. What is $\mathbb{E}[XY]$?

Solution :

$$\mathbb{E}[XY] = \frac{1}{6}(1 \times 2 + 1 \times 3 + 2 \times 1 + 2 \times 3 + 3 \times 1 + 3 \times 2) = \frac{11}{3}$$

5.4.3 Variance of Sum of Independent Random Variables

Theorem 5.4.2 (Variance of Sum of Independent Random Variables)**.** If X, Y are independent. Then,

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

To show this, we proceed as follows:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathbb{E}[(X+Y-(\mu_X+\mu_Y))^2] \\
&= \mathbb{E}[((X-\mu_X)+(Y-\mu_Y))^2] \\
&= \mathbb{E}[(X-\mu_X)^2+(Y-\mu_Y)^2+2(X-\mu_X)(Y-\mu_Y)] \\
&= \mathbb{E}[(X-\mu_X)^2]+\mathbb{E}[(Y-\mu_Y)^2]+\mathbb{E}[2(X-\mu_X)(Y-\mu_Y)] \\
&= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\,\mathrm{Cov}[X,Y]
\end{aligned}$$

Since X and Y are independent (thus uncorrelated), $\mathrm{Cov}[X,Y] = 0 \implies \mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$.

Remark. X and Y are uncorrelated if and only if

$$\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y]$$

Example. There is a bag with 3 cards, with face values 1, 2, and 3.

1. You draw two cards with replacement. X, Y are the face values of the first and second cards. What is $\mathrm{Var}[X+Y]$?

Solution:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathrm{Var}[X]+\mathrm{Var}[Y] \\
&= 2\times\mathrm{Var}[X] \\
&= 2\times(\mathbb{E}[X^2]-\mathbb{E}[X]^2) \\
&= 2\times(\frac{1}{3}(1^2+2^2+3^2)-(\frac{1}{3}(1+2+3))^2) \\
&= \frac{4}{3}
\end{aligned}$$

2. You draw two cards without replacement. What is $\mathrm{Var}[X+Y]$?

Solution:

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \mathrm{Var}[X]+\mathrm{Var}[Y]+2\,\mathrm{Cov}[X,Y] \\
&= \frac{4}{3}+2(\mathbb{E}[XY]-\mathbb{E}[X]\mathbb{E}[Y]) \\
&= \frac{4}{3}+2(\frac{11}{3}-2\times 2) \\
&= \frac{2}{3}
\end{aligned}$$

Remark. Covariance can have negative values while variance cannot.

Suppose $X_1, X_2, \cdots, X_n$ are pairwise independent random variables, meaning for every $i \neq j$, X_i and X_j are independent. Then,

$$\mathrm{Var}[X_1+X_2+\cdots+X_n] = \mathrm{Var}[X_1]+\mathrm{Var}[X_2]+\cdots+\mathrm{Var}[X_n]$$

Compared to the statement that for independent random variables, we require all of them to be independent from each other, i.e., $\mathbb{P}(X_1=x_1, X_2=x_2, \cdots, X_n=x_n) = \mathbb{P}(X_1=x_1)\mathbb{P}(X_2=x_2)\cdots\mathbb{P}(X_n=x_n)$, pairwise independence is a weaker statement, which only requires pairs of random variables to be independent. Also, we have:

$$\mathrm{Var}[X_1+X_2+\cdots+X_n] = \sum_{i=1}^{n}\mathrm{Var}[X_i]+\sum_{\substack{i\neq j\\ 1\leq i-j\leq n}}\mathrm{Cov}[X_i,X_j] \quad \text{(if pairwise uncorrelated, the later part} = 0\text{)}$$

5.5 Appendix

Example (The Hat Problem)**.** Three are n people throwing their hats in a box and then picking one at random. Suppose all permutations are equally likely, and everyone is equivalent to picking one hat at a time. Let X be the number of people who get their own hat. Find $\mathbb{E}[X]$ and $\mathrm{Var}[X]$.

Solution: Suppose all permutations are equally likely, we have $X = X_1 + X_2 + \cdots + X_n$, where

$$X_i = \begin{cases} 1, & \text{if } i \text{ selects own hat,} \\ 0, & \text{otherwise.} \end{cases}$$

Then we have

$$\mathbb{E}[X_i] = \mathbb{E}[X_1] = \mathbb{P}(X_1 = 1) = \frac{1}{n}$$

Thus, $\mathbb{E}[X] = \mathbb{E}[X_1] + \mathbb{E}[X_2] + \cdots + \mathbb{E}[X_n] = n \times \frac{1}{n} = 1$

However, by intuition, the random variables $X_1, X_2, \cdots, X_n$ are not independent. We thus cannot simply add up all the variances together.

From previous, we have $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$. For $X = X_1 + X_2 + \cdots + X_n$,

$$X^2 = \underbrace{\sum_i X_i^2}_{n \text{ terms}} + \underbrace{\sum_{i,j:i\neq j} X_i X_j}_{n(n-1) \text{ terms}}$$

$$\begin{aligned}
\mathrm{Var}[X] &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 \\
&= \mathbb{E}\left(\sum_i X_i^2 + \sum_{i,j:i\neq j} X_i X_j\right) - 1^2 \\
&= \sum_i \mathbb{E}[X_1^2] + \sum_{i,j:i\neq j} \mathbb{E}[X_i X_j] - 1 \\
&= n \cdot \frac{1}{n} + n \cdot (n-1) \cdot \mathbb{E}[X_1 X_1] - 1 \\
&= 1 + n \cdot (n-1) \cdot \frac{1}{n} \cdot \frac{1}{n-1} - 1 \quad (*) \\
&= 1
\end{aligned}$$

$$(*) = \mathbb{E}[X_1 X_1] = \mathbb{P}(X_1 X_2 = 1) = \mathbb{P}(X_1 = 1, X_2 = 1) = \mathbb{P}(X_1 = 1)\mathbb{P}(X_2 = 1 | X_1 = 1) = \frac{1}{n} \cdot \frac{1}{n-1}$$

Remark. This example comes from MIT RES.6-012 Introduction to Probability, Spring 2018.

Chapter 6

Continuous Random Variables

The probability models we used earlier were discrete. Starting from this chapter, we will discuss some non-discrete probability models.

6.1 Cumulative Distribution Function (CDF)

The Cumulative Distribution Function works for both discrete and continuous probabilities.

Definition 6.1.1 (Cumulative Distribution Function (CDF)). The cumulative distribution function (CDF) F of a random variable X is:

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(t)dt = \sum_{k \leq x} p_X(k)$$

Every CDF $F : \mathbb{R} \to [0, 1]$ must satisfy the following properties:

1. F is monotonically increasing: $x \leq y \implies F_X(x) \leq F(y)$
2. $\lim_{x \to -\infty} F_X(x) = 0$
3. $\lim_{x \to +\infty} F_X(x) = 1$

Example. Find the CDF of a Geometric($p = \frac{1}{2}$) random variable.

Solution:

$$F(k) = \mathbb{P}[X \leq k] = 1 - \mathbb{P}[X > k] = 1 - (1 - p)^k$$

6.1.1 Probability Density Function (PDF)

Definition 6.1.2 (Probability Density Function (PDF)). For a continuous random variable X, we define its probability density function (PDF) f as the derivative of its CDF:

$$f_X(x) := \lim_{\delta \to 0} \frac{\mathbb{P}(x \leq X \leq x + \delta)}{\delta} = \lim_{\delta \to 0} \frac{F_X(x + \delta) - F_X(x)}{\delta} = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x}$$

Remark. PDF is not probability, it is the density of probability.

For example,

$$F_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ x, & \text{if } 0 \leq x \leq 1, \\ 1, & \text{if } 1 < x. \end{cases} \implies f_X(x) = \begin{cases} 0, & \text{if } x < 0, \\ 1, & \text{if } 0 \leq x \leq 1, \\ 0, & \text{if } 1 < x. \end{cases}$$

Example. The CDF of random variable X is $F_X(x) = \begin{cases} 1 - \dfrac{1}{x^2}, & \text{if } x > 1 \\ 0, & \text{if } x \le 1 \end{cases}$.

Find the PDF of X.

Solution: First, we need to verify whether it is a valid CDF. Since it satisfies the three properties, we can then determine the PDF of X.

$$f_X(x) = \frac{\mathrm{d}F_X(x)}{\mathrm{d}x} = \begin{cases} \dfrac{2}{x^3}, & \text{if } x > 1 \\ 0, & \text{if } x \le 1 \end{cases}$$

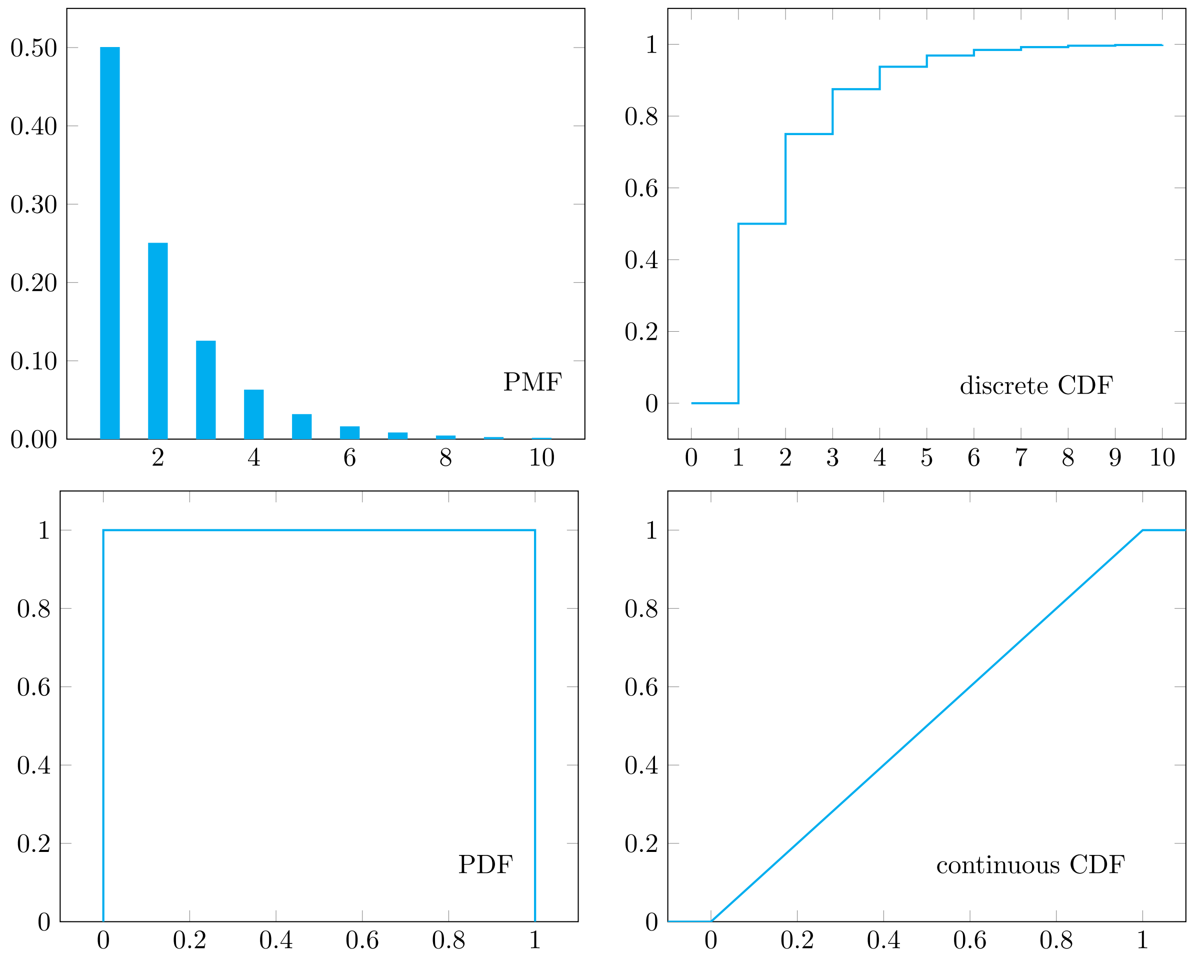
Remark. A PDF can take any non-negative value $f_X(x) \ge 0$ and its value can be greater than 1.

For a small $\delta > 0$, we know $F(x+\delta) - F_X(x) \approx f_X(x)\delta$ and so for small δ,

$$\mathbb{P}(x \le X \le x + \delta) \approx f_X(x)\delta.$$

However, when $\sigma = 0$, $\mathbb{P}(x \le X \le x + \delta) = 0$, which shows that in a continuous distribution, the probability for any specific point in the PDF is 0, i.e.

$$\mathbb{P}(X = x) = 0$$



6.1.2 Integral of PDF and Probability Calculation

For a continuous random variable X, the probability of an event E can be calculated using the integral of PDF f. Therefore,

- $\mathbb{P}(E) = \int_E f_X(x)dx$ (i.e. $\mathbb{P}(a \le X \le b) = \int_a^b f_X(x)dx$)
- $\mathbb{P}(X \le t) = \int_{-\infty}^t f_X(x)dx$
- $\int_{-\infty}^{\infty} f_X(x)dx = 1$

To summarize, we have

	PMF $p_X(x)$	PDF $f_X(x)$
$\mathbb{P}(X \le a)$	$\sum_{x \le a} p_X(x)$	$\int_{-\infty}^a f_X(x)dx$
$\mathbb{E}[X]$	$\sum_x x p_X(x)$	$\int_{-\infty}^{\infty} x f_X(x)dx$
$\mathbb{E}[X^2]$	$\sum_x x^2 p_X(x)$	$\int_{-\infty}^{\infty} x^2 f_X(x)dx$
$\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$	$\sum_x (x-\mu)^2 p_X(x)$	$\int_{-\infty}^{\infty} (x-\mu)^2 f_X(x)dx$

6.2 Uniform Random Variable

Suppose a postal package is to be delivered to you between noon and 1:00 pm. You have to leave home between 12:30 to 12:45 pm. What is the probability you miss the delivery?

This question is quite simple, and we can immediately see that the answer is $\frac{15}{60}$. However, how can we define a probability model here?

Definition 6.2.1 (Uniform Random Variable). A uniform random variable X over interval $[0,1]$ satisfies

$$\mathbb{P}(X \le x) = \begin{cases} 0, & \text{if } x \le 0, \\ x, & \text{if } 0 < x \le 1, \\ 1, & \text{if } 1 < x. \end{cases}$$

Based on the above definition, we can calculate the following:

$$\begin{aligned} \mathbb{P}(T > \frac{1}{2}) &= 1 - \mathbb{P}(T \le \frac{1}{2}) = 1 - \frac{1}{2} = \frac{1}{2} \\ \mathbb{P}(T > \frac{3}{4}) &= 1 - \mathbb{P}(T \le \frac{3}{4}) = 1 - \frac{3}{4} = \frac{1}{4} \\ \mathbb{P}(\frac{1}{2} < T \le \frac{3}{4}) &= \mathbb{P}(T > \frac{1}{2}) - \mathbb{P}(T > \frac{3}{4}) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4} \end{aligned}$$

Example (Continuous Random Variables and Zero Probability for Points). Consider a Uniform($[0,1]$) random variable T.

1. $\mathbb{P}(T \le 0.3) = 0.3$
2. $\mathbb{P}(T = 0.3) = 0$

Assume that $\mathbb{P}(T = 0.3) = \varepsilon > 0$, $\mathbb{P}(T \le 0.3 - \frac{\varepsilon}{4}) = 0.3 - \frac{\varepsilon}{4}$, $\mathbb{P}(T \le 0.3 + \frac{\varepsilon}{4}) = 0.3 + \frac{\varepsilon}{4}$

Above gives $\mathbb{P}(0.3 - \frac{\varepsilon}{4} < T < 0.3 + \frac{\varepsilon}{4}) = \mathbb{P}(T \le 0.3 + \frac{\varepsilon}{4}) - \mathbb{P}(T \le 0.3 - \frac{\varepsilon}{4}) = \frac{\varepsilon}{2}$

However, for $\mathbb{P}(T = 0.3) \subseteq \mathbb{P}(0.3 - \frac{\varepsilon}{4} < T < 0.3 + \frac{\varepsilon}{4})$, we have $\varepsilon < \frac{\varepsilon}{2}$, which by contradiction shows that $\mathbb{P}(T = 0.3) = 0$

3. $\mathbb{P}(T < 0.3) = \mathbb{P}(T \le 0.3) - \mathbb{P}(T = 0.3) = 0.3$

Example. Find the expected value and variance of a Uniform($[0,1]$) random variable.

Solution:

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

$$\mathbb{E}(x) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \cdot 1 dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}$$

$$\mathbb{E}(x^2) = \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^1 x^2 \cdot 1 dx = \frac{x^3}{3}\Big|_0^1 = \frac{1}{3}$$

$$\mathrm{Var}[x] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{3} - (\frac{1}{2})^2 = \frac{1}{12}$$

Consider the general case where a uniform random variable X is defined over the interval $[a,b]$, i.e.,

$$F_X(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{x-a}{b-a}, & \text{if } a \le x \le b \\ 1, & \text{if } b < x \end{cases}$$

Then,

1. The PDF of X is $f_X(x) = \begin{cases} 0, & \text{if } x < a \\ \frac{1}{b-a}, & \text{if } a \le x \le b \\ 0, & \text{if } b < x \end{cases}$.

If $f_X(x) = c$ for $a \le x \le b$, since $\int_{-\infty}^{\infty} f_X(x) dx = 1 \Rightarrow c(b-a) = 1 \Rightarrow c = \frac{1}{b-a}$

2. The expected value of X is $\mathbb{E}[X] = \frac{a+b}{2}$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a}\int_a^b x dx = \frac{x^2}{2(b-a)}\Big|_a^b = \frac{a+b}{2}$$

3. The variance of X is $\mathrm{Var}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{b^3-a^3}{3(b-a)} - (\frac{a+b}{2})^2 = \frac{(b-a)^2}{12}$

Let Y be Uniform(0, 1) $\rightarrow \mathbb{E}[Y] = \frac{1}{2}, \mathrm{Var}[Y] = \frac{1}{12}$

Let $Z = (b-a)Y + a$, which gives Uniform(a,b)

$$\mathbb{E}[Z] = (b-a)\mathbb{E}[Y] + a = \frac{b+a}{2}$$

$$\mathrm{Var}[Z] = (b-a)^2\,\mathrm{Var}[Y] = \frac{(b-a)^2}{12}$$

6.3 Exponential Random Variable

Consider rain is falling on your head at a rate of λ drops/sec. How long do we wait until the next drop? We can first divide 1 second to n sub-intervals, then the probability of rain drop would be $\frac{\lambda}{n}$.

Then we have

$$\frac{\mathbb{P}(T = \frac{t}{n})}{\frac{1}{n}} = \frac{\frac{\lambda}{n}\left(1 - \frac{\lambda}{n}\right)^{t-1}}{\frac{1}{n}} \quad (\text{Let } t = ns) = \lambda\left(1 - \frac{\lambda}{n}\right)^{ns-1}$$

$$\lim_{n\to\infty} \frac{\mathbb{P}(T = \frac{t}{n})}{\frac{1}{n}} = \lim_{n\to\infty} \lambda\left(1 - \frac{\lambda}{n}\right)^{ns-1} = \lim_{n\to\infty} \lambda\left(1 - \frac{\lambda}{n}\right)^{ns}\left(1 - \frac{\lambda}{n}\right)^{-1} = \lambda\left(\lim_{n\to\infty}\left(1 - \frac{\lambda}{n}\right)^n\right)^s = \lambda e^{-\lambda s}$$

Definition 6.3.1 (Exponential Random Variable)**.** An Exponential(λ) random variable has the following PDF:

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

To check the validity of the above, we proceed as follows:

$$\int_{-\infty}^{\infty} f_X(x)dx = \int_0^{\infty} \lambda e^{-\lambda x}dx = -e^{-\lambda x}\Big|_0^{\infty} = 1$$

The Exponential(λ) random variable X satisfies:

1. The CDF of X is $F_X(x) = 1 - e^{-\lambda x}$ $(x \geq 0)$

$$F_X(x) = \int_{-\infty}^{x} f_X(x)dx = \int_0^{x} \lambda e^{-\lambda x}dx = -e^{-\lambda x}\Big|_0^{x} = 1 - e^{-\lambda x}$$

2. The expected value of X is $\mathbb{E}[X] = \frac{1}{\lambda}$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} xf(x)dx = \int_0^{\infty} \lambda x e^{-\lambda x}dx = \left[-xe^{-\lambda x} - \frac{1}{\lambda}e^{-\lambda x}\right]_0^{\infty} = 0 - (0 - \frac{1}{\lambda}) = \frac{1}{\lambda}$$

3. The variance of X is $\mathrm{Var}[X] = \frac{1}{\lambda^2}$

Exponential and Geometric Random Variables can be related as the following:

$$\lim_{n\to\infty} \frac{1}{n}\text{Geometric}(\frac{\lambda}{n}) = \text{Exponential}(\lambda)$$

The exponential distribution has the memoryless property, which says that future probabilities do not depend on any past information

Example (Memoryless Property of Exponential Distribution)**.**

Buses arrive at a rate of one in 5 minutes.

(a) A bus arrives now. How likely are you to wait at least 5 minutes for the next bus?

Solution: Let X = time to wait until the next bus arrives. Given that $\mathbb{E}[X] = 5, \frac{1}{\lambda} = 5, \lambda = 0.2$

$$\mathbb{P}(X \geq 5) = 1 - F_X(5) = 1 - (1 - e^{-\lambda x}) = e^{-0.2\times 5} = e^{-1}$$

(b) The last bus arrived 5 minutes ago. How likely are you to wait at least 5 minutes from now for the next bus to come?

Solution:

$$\mathbb{P}(X \geq 10 | X \geq 5) = \frac{\mathbb{P}(X \geq 10 \cap X \geq 5)}{\mathbb{P}(X \geq 5)} = \frac{\mathbb{P}(X \geq 10)}{\mathbb{P}(X \geq 5)} = \frac{e^{-0.2\times 10}}{e^{-0.2\times 5}} = e^{-1}$$

The Exponential distribution is memoryless (the past has no effects on its future), which means for every exponential random variable X and constants $s, t \geq 0$:

$$\mathbb{P}(X \geq s + t | X \geq t) = \mathbb{P}(X \geq s)$$

Proof.

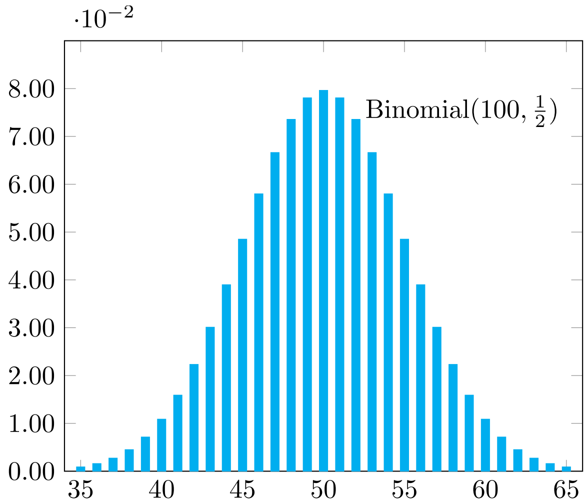
$$\mathbb{P}(X \geq s + t | X \geq t) = \frac{\mathbb{P}(X \geq s + t \cap X \geq t)}{\mathbb{P}(X \geq t)} = \frac{\mathbb{P}(X \geq s + t)}{\mathbb{P}(X \geq t)} = \frac{e^{-(s+t)\lambda}}{e^{-t\lambda}} = e^{-s\lambda} = \mathbb{P}(X \geq s)$$

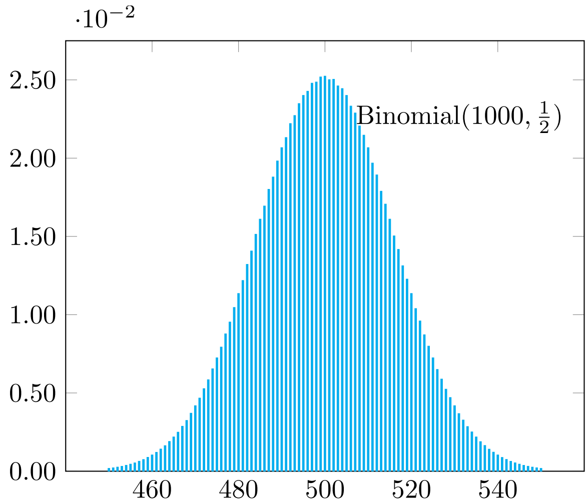
■

6.4 Normal Distribution (Gaussian Random Variable)

6.4.1 Normal Random Variable

We see from the following that as n increases, the shape converges to a bell shape curve.





Definition 6.4.1 (Normal Distribution)**.** We define normal (Gaussian) probability density function (PDF) $\mathcal{N}(\mu, \sigma^2)$ with the parameters (mean) μ and (variance) σ^2 as

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

Remark. From calculus we know that

$$\int_{x=-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{x=-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = 1 \Longrightarrow \int_{x=-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \sqrt{2\pi\sigma^2}$$

which makes it easier for the calculation of some integral. For example, when $\mu = 0, \sigma = 1$,

$$\int_{x=-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}.$$

Note that the parameters μ and σ^2 in the above definition are equal to the mean and variance of the normal random variable X:

$$\mathbb{E}[X] = \mu, \qquad \mathrm{Var}[X] = \sigma^2$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{\infty} \frac{x}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\
&= \underbrace{\int_{-\infty}^{\infty} \frac{x-\mu}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}_{\text{let } \tilde{x}=x-\mu} + \underbrace{\mu \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx}_{\mu \int_{-\infty}^{\infty} f_X(x) = \mu \times 1} \\
&= \underbrace{\int_{-\infty}^{\infty} \frac{\tilde{x}}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\tilde{x})^2}{2\sigma^2}} dx}_{\text{odd function gives } 0} + \mu \\
&= \mu
\end{aligned}$$

■

6.4.2 Standard Normal Distribution

Standard normal $\mathcal{N}(0,1)$is the distribution of a normal random variable Z with parameters (mean) $\mu = 0$ and (variance) $\sigma^2 = 1$, i.e. with zero mean and unit variance, as

$$f_X(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

We don't have an easy-to-use CDF for a normal random variable. Therefore, every time we try to find the CDF, we have to use

$$F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{t \le x} e^{-\frac{t^2}{2}} dt.$$

However, if we know the CDF for the standard normal distribution, we don't need to derive the CDF for other normal distributions.

Suppose a random variable X has expected value $\mathbb{E}[X] = \mu$ and variance $\mathrm{Var}[X] = \sigma^2$. Then, random variable $Z = \frac{X-\mu}{\sigma}$ has zero expected value and unit variance, i.e.,

$$\mathbb{E}[Z] = 0, \qquad \mathrm{Var}[Z] = 1.$$

To show this, we proceed as follows:

$$\mathbb{E}[Z] = \mathbb{E}[\frac{X-\mu}{\sigma}] = \frac{\mathbb{E}[X]-\mu}{\sigma} = 0; \quad \mathrm{Var}[Z] = \mathbb{E}[(Z-0)^2] = \mathbb{E}\left[(\frac{X-\mu}{\sigma})^2\right] = \frac{\mathbb{E}[(X-\mu)^2]}{\sigma^2} = 1$$

If random variable X has a normal distribution with parameters μ and σ^2, then random variable $Z = \frac{X-\mu}{\sigma}$ has a standard normal distribution. We can then find $X = \mu + \sigma Z$.

It simply means that if we shift and scale any normal random variable using $Z = \frac{X-\mu}{\sigma}$, the calculation becomes simpler.

Example. Suppose X is a normal random variable with parameters $\mu = 3$ and $\sigma^2 = 9$. Find $\mathbb{P}(|X-3| < 6)$.

Solution: Since $Z = \frac{X-\mu}{\sigma} = \frac{X-3}{3} \sim \mathcal{N}(0,1)$

$$\begin{aligned}
\mathbb{P}(|X-3| < 6) &= \mathbb{P}(-6 < X-3 < 6) \\
&= \mathbb{P}(-2 < \frac{X-3}{3} < 2) \\
&= \mathbb{P}(-2 < Z < 2) \\
&= F_Z(2) - F_Z(-2) \\
&= \Phi(2) - (1 - \Phi(2)) \\
&= 2\Phi(2) - 1 \\
&= 2 \times 0.9772 - 1 \\
&= 0.9544
\end{aligned}$$

Remark.

$$\mathbb{P}(Z \le -\alpha) = \mathbb{P}(-Z \ge \alpha) = \mathbb{P}(Z \ge \alpha) = 1 - \mathbb{P}(Z \le \alpha) = 1 - F_Z(\alpha)$$

6.4.3 Linear Function of a Normal Random Variable

Suppose we have $Y = aX + b$, where $X \sim \mathcal{N}(\mu, \sigma^2)$. Then we have $\mathbb{E}[Y] = a\mu + b$ and $\mathrm{Var}[Y] = a^2\sigma^2$, where $Y \sim \mathcal{N}(a\mu + b, a^2\sigma^2)$.

Since Y is a function of a normal random variable, it is evident that Y is also a normal random variable.

However, consider a special case where $a = 0$. In this case, $Y = b$. Although it is a constant (discrete) random variable, we can represent it as $Y \sim \mathcal{N}(b, 0)$, where the variance is 0. This shows that it is always true that a linear function of a normal random variable results in a normal random variable.

6.4.4 Normal Approximation of Binomial Distribution

Theorem 6.4.1. If S_n is a Binomial(n,p) random variable with mean np and variance $np(1-p)$, then for every $a < b$:

$$\lim_{n\to\infty} \mathbb{P}\left(a \le \frac{S_n - np}{\sqrt{np(1-p)}} \le b\right) = \Phi(b) - \Phi(a)$$

where Φ is the CDF of standard normal distribution. One may use the Z-Table to look up the value.

Remark. Note that $Z_n = \frac{S_n - np}{\sqrt{np(1-p)}}$ has zero mean and unit variance for every n, and the above theorem shows the CDF of Z_n converges to the CDF of standard normal.

Example. X is the number of heads when we flip a fair coin 100 times. Use the normal approximation to approximate the probability $45 \le X \le 55$.

Solution: Since $X \sim \text{Binomial}(100, \frac{1}{2})$, we have

$$\mathbb{E}[X] = np = 100 \times \frac{1}{2} = 50; \quad \text{Var}[X] = np(1-p) = 100 \times \frac{1}{2} \times \frac{1}{2} = 25; \quad Z = \frac{X-50}{5}$$

$$\begin{aligned}
\mathbb{P}(45 \le X \le 55) &= \underbrace{\mathbb{P}(44.5 \le X \le 55.5)}_{\text{we here use 44.5 and 55.5 for better approximation}} \\
&= \mathbb{P}(\frac{44.5-50}{5} \le \frac{X-50}{5} \le \frac{55.5-50}{5}) \\
&= \mathbb{P}(-1.1 \le \frac{X-50}{5} \le 1.1) \\
&\approx 2\Phi(1.1) - 1 \\
&\approx 2 \times 0.8643 - 1 \\
&\approx 0.7286
\end{aligned}$$

6.5 Multiple Continuous Random Variable

6.5.1 Joint CDF and PDF of Random Variable

Definition 6.5.1 (Joint CDF and PDF of Random Variable)**.** A pair of random variables X, Y can be described by their joint CDF F_{XY} which is defined as:

$$F_{XY}(x,y) = \mathbb{P}(X \le x, Y \le y)$$

Also, we define the joint PDF f_{XY} using the derivative of joint CDF F_{XY}:

$$f_{XY}(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y} F_{XY}(x,y) = \lim_{\epsilon,\delta\to 0} \frac{\mathbb{P}(x \le X \le x+\epsilon, y \le Y \le y+\delta)}{\epsilon \cdot \delta}$$

Suppose F_{XY} is the joint CDF of X, Y. Then, the marginal CDFs F_X, F_Y can be found as

$$F_X(x) = \mathbb{P}(X \le x) = F_{XY}(x, +\infty) = \lim_{y\to+\infty} F_{XY}(x,y)$$

$$F_Y(y) = \mathbb{P}(Y \le y) = F_{XY}(+\infty, y) = \lim_{x\to+\infty} F_{XY}(x,y)$$

Suppose f_{XY} is the joint PDF of X, Y. Then, the marginal PDFs f_X, f_Y can be found by the integration of joint PDF

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{XY}(x,y)dydx = 1$$

$$f_X(x) = \int_{y=-\infty}^{+\infty} f_{XY}(x,y)dy, \quad f_Y(y) = \int_{x=-\infty}^{+\infty} f_{XY}(x,y)dx$$

For probability of an event A, we have

$$\mathbb{P}(A) = \int\int_A f_{XY}(x,y)dxdy$$

For expected value, we have

$$\mathbb{E}[g(X,Y)] = \int\int g(x,y)f_{XY}(x,y)dxdy$$

Example. Rain drops at a rate of 1 drop/sec. Let X and Y be the arrival times of the first and second raindrops. Find F_{XY} and f_{XY}.

Solution: Since rate = 1 drop/sec, $X \sim$ Exponential(1). For Joint CDF F_{XY},

$$\begin{aligned}
F_{XY}(x,y) &= \mathbb{P}(X \le x, Y \le y) \\
&= \int_{z=-\infty}^{x} f_X(z)\mathbb{P}(Y \le y | X = z)dz \quad (\mathbb{P}(Y - X \le y - z | X = z), \text{ independent}) \\
&= \int_{z=0}^{x} (e^{-z})\mathbb{P}(Y - X \le y - z)dz \\
&= \int_{z=0}^{x} (e^{-z})(1 - e^{-1\cdot(y-z)})dz \\
&= \int_{z=0}^{x} (e^{-z} - e^{-y})dz \\
&= -e^{-z} - ze^{-y}\Big|_0^x \\
&= 1 - e^{-x} - xe^{-y}
\end{aligned}$$

Then we have:

$$F_{XY}(x,y) = \begin{cases} 1 - e^{-y} - ye^{-y}, & \text{if } x \ge y \ge 0 \\ 1 - e^{-x} - xe^{-y}, & \text{if } y \ge x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

For Joint PDF f_{XY}, we have:

$$f_{XY}(x,y) = \begin{cases} 0, & \text{if } x \ge y \ge 0 \\ e^{-y}, & \text{if } y \ge x \ge 0 \\ 0, & \text{otherwise} \end{cases} \implies f_{XY}(x,y) = \begin{cases} e^{-y}, & \text{if } y \ge x \ge 0 \\ 0, & \text{otherwise} \end{cases}$$

6.5.2 Independent Continuous Random Variables

Definition 6.5.2 (Independent Continuous Random Variables)**.** Random variables X, Y are called independent if their joint CDF F_{XY} is the product of marginal CDFs:

$$F_{XY}(x,y) = F_X(x)F_Y(y)$$

The above definition is equivalent to the joint PDF f_{XY} being the product of marginal PDFs:

$$f_{XY}(x,y) = f_X(x)f_Y(y)$$

To prove this, we see that

$$\frac{\partial}{\partial x}\frac{\partial}{\partial y}F_{XY}(x,y) = (\frac{\partial}{\partial x}F_X(x))(\frac{\partial}{\partial y}F_Y(y)) = f_X(x)f_Y(y)$$

The Joint PDF of independent normal random variables X, Y with mean μ and variance σ^2:

$$f_{XY}(x,y) = f_X(x)f_Y(y) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}\right)\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\right) = \frac{1}{2\pi\sigma^2}e^{-\frac{(x-\mu)^2+(y-\mu)^2}{2\sigma^2}}$$

Example. Suppose X, Y are independent uniform random variables over $[0,1]$. Then,

$$f_{XY}(x,y) = \begin{cases} 1, & \text{if } 0 \le x, y \le 1 \\ 0, & \text{otherwise} \end{cases}$$

For $A \subseteq [0,1]^2$, we can calculate A's area $\mathbb{P}(A) = \int_A f_{XY}(x,y)dxdy = \text{area}(A)$.

Definition 6.5.3 (Uncorrelated Continuous Random Variables)**.** Random variables X, Y are called uncorrelated if they satisfy

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

If the above does not hold, we call X and Y correlated.

Example.

(a) Find joint PDF of X, Y is uniform over the unit circle $\{[x,y] \in \mathbb{R}^2 : x^2 + y^2 \le 1\}$.

Solution:

$$f_{XY}(x,y) = \begin{cases} c, & \text{if } x^2 + y^2 \le 1 \\ 0, & \text{otherwise} \end{cases}$$

$$\begin{aligned} \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} f_{XY}(x,y)dydx = 1 \\ c \cdot \text{Area}(\{[x,y] \in \mathbb{R}^2 : x^2 + y^2 \le 1\}) = 1 \\ c = \frac{1}{\pi} \end{aligned}$$

(b) Are X and Y independent?

Solution:

$$\begin{aligned} f_{XY}(\frac{4}{5}, \frac{4}{5}) &= 0 \text{ for } (\frac{4}{5})^2 + (\frac{4}{5})^2 > 1 \\ f_X(\frac{4}{5}) &= \int_{y=-\infty}^{+\infty} \underbrace{f_{XY}(\frac{4}{5}, y)}_{\frac{1}{\pi} \text{ only if } (\frac{4}{5})^2 + y^2 \le 1} dy = \int_{y=-\frac{3}{5}}^{\frac{3}{5}} \frac{1}{\pi} dy = \frac{6}{5\pi} \neq 0 \\ f_Y(\frac{4}{5}) &\neq 0 \\ &\implies f_{XY}(\frac{4}{5}, \frac{4}{5}) \neq f_X(\frac{4}{5}) f_Y(\frac{4}{5}) \end{aligned}$$

Therefore, they are not independent.

(c) Are X and Y uncorrelated?

Solution:

$$\begin{aligned} \mathbb{E}[XY] &= \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} xy f_{XY}(x,y)dydx \\ &= \int_{x=-\infty}^{+\infty} x \underbrace{\int_{y=-\infty}^{+\infty} y f_{XY}(x,y)dy}_{\frac{y}{\pi} \text{ if } x^2+y^2 \le 1 \to \text{odd function}} dx \\ &= 0 \\ \mathbb{E}[Y] &= \int_{x=-\infty}^{+\infty} \int_{y=-\infty}^{+\infty} y f_{XY}(x,y)dydx = 0 \end{aligned}$$

Since $\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$, they are uncorrelated.

Same as for a discrete random variable, given an event, we can condition on the probability density function.

Suppose we have a continuous random variable X, given that $X \in A$,

$$\begin{aligned}
\mathbb{P}(x \leq X \leq x+\sigma | X \in A) &\approx f_{X|X\in A}(x) \cdot \sigma \\
\frac{\mathbb{P}(x \leq X \leq x+\sigma, X \in A)}{\mathbb{P}(A)} &\approx f_{X|X\in A}(x) \cdot \sigma \\
\frac{\mathbb{P}(x \leq X \leq x+\sigma)}{\mathbb{P}(A)} &\approx f_{X|X\in A}(x) \cdot \sigma \\
\frac{f_X(x)\sigma}{\mathbb{P}(A)} &\approx f_{X|X\in A}(x) \cdot \sigma \\
\frac{f_X(x)}{\mathbb{P}(A)} &\approx f_{X|X\in A}(x)
\end{aligned}$$

Thus, we have

$$f_{X|X\in A}(x) = \begin{cases} 0, & \text{if } x \notin A \\ \dfrac{f_X(x)}{\mathbb{P}(A)}, & \text{if } x \in A \end{cases}$$

Definition 6.5.4. For random variables X, Y, we define the conditional PDF $f_{X|Y}$ as

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

Then we have the following properties:

1. Independent Random Variable: $f_{X|Y}(x|y) \forall x, y \in \mathbb{R}$.
2. Total Probability Theorem: $f_X(x) = \displaystyle\int_{-\infty}^{+\infty} f_{X|Y}(x|y) f_Y(y) dy$.
3. Total Expectation Theorem: $\mathbb{E}[X] = \underbrace{\displaystyle\int_{-\infty}^{+\infty} \mathbb{E}[X|Y=y] f_Y(y) dy}_{\mathbb{E}[\mathbb{E}[X|Y]]}$, $\mathbb{E}[X|Y=y] = \int_{-\infty}^{+\infty} x f_{X|Y}(x|y) dx$.

Theorem 6.5.1 (Continuous Bayes Rule)**.** For random variables X, Y, the conditional PDF $f_{X|Y}$ and $f_{Y|X}$ satisfy

$$f_{X|Y}(x|y) = \frac{f_{Y|X}(y|x) f_X(x)}{f_Y(y)} = \frac{f_{Y|X}(y|x) f_X(x)}{\displaystyle\int_{x=-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx}$$

Theorem 6.5.2. The PDF of the $Z = X + Y$, the summation of independent random variables X, Y, is the convolution of the marginal PDFs:

$$f_Z(z) = \int_{x=-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

Corollary 6.5.1. The sum $Z = X + Y$ of independent normal random variables $X \sim \mathcal{N}(\mu_x, \sigma_x^2), Y \sim \mathcal{N}(\mu_y, \sigma_y^2)$ is a normal random variable with mean $\mu_x + \mu_y$ and variance $\sigma_x^2 + \sigma_y^2$:

$$X + Y \sim \mathcal{N}(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Chapter 7

Chebyshev's Inequality

7.1 Chebyshev's Inequality

Theorem 7.1.1 (Markov's Inequality)**.** Suppose that random variable $X \geq 0$ only takes non-negative values. Then, for every $t > 0$

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}[X]}{t}$$

Proof.

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{P}(X \geq t)\mathbb{E}[X|X \geq t] + \mathbb{P}(X < t)\mathbb{E}[X|X < t] \\
\mathbb{E}[X] &\geq \mathbb{P}(X \geq t)\mathbb{E}[X|X \geq t] \\
\mathbb{E}[X] &\geq t \times \mathbb{P}(X \geq t) \\
\mathbb{P}(X \geq t) &\leq \frac{\mathbb{E}[X]}{t}
\end{aligned}$$

■

Above could be understood as: if $X \geq 0$ and $\mathbb{E}[X]$ is small, then X is unlikely to be very large.

However, what if we take negative values for the random variable we use?

Theorem 7.1.2 (Chebyshev's Inequality)**.** Suppose that random variable X has a mean $\mathbb{E}[X] = \mu$ and variance $\mathrm{Var}[X] = \sigma^2$. Then, for every $t > 0$,

$$\mathbb{P}(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}$$

Proof. The proof follows from Markov's inequality. Let $Z = (X - \mu)^2 \geq 0$

$$\begin{aligned}
\mathbb{P}(Z \geq \varepsilon^2) &\leq \frac{\mathbb{E}[Z]}{\varepsilon^2} \\
&\leq \frac{\mathbb{E}[(X - \mu)^2]}{\varepsilon^2} \\
&\leq \frac{\mathrm{Var}[X]}{\varepsilon^2} \\
&\leq \frac{\sigma_x^2}{\varepsilon^2} \\
\implies \mathbb{P}(|X - \mu| \geq \varepsilon) &\leq \frac{\sigma^2}{\varepsilon^2}
\end{aligned}$$

■

Again, by intuition, the above says that if the variance is small, then X is unlikely to be too far from the mean.

Example.

(a) Find a bound on the probability of ≤ 24 heads in 64 coin flips?

Solution:

Let $X \sim \text{Binomial}(64, \frac{1}{2}) \rightarrow \mathbb{E}[X] = 64 \times \frac{1}{2} = 32$, $\text{Var}[X] = np(1-p) = 64 \times \frac{1}{2} \times \frac{1}{2} = 16, \sigma_x = \sqrt{16} = 4$

$$\begin{aligned} \mathbb{P}(X \leq 24) &= \mathbb{P}(X - \mu \leq 24 - 32) \\ &= \mathbb{P}(X - \mu \leq -2\sigma_x) \\ &\leq \mathbb{P}(|X - \mu| \geq 2\sigma_x) \\ &\leq \frac{\sigma_x^2}{(2\sigma_x)^2} \\ &= \frac{1}{4} \end{aligned}$$

(b) Find a bound on the probability of ≤ 24 or ≥ 40 heads in 64 coin flips?

Solution:

$$\begin{aligned} \mathbb{P}(|X - \mu| \geq 2\sigma_x) &= \mathbb{P}(|X - 32| \geq 8) \\ &= \mathbb{P}(X \leq 24 \text{ or } X \geq 40) \\ &= \frac{\sigma_x^2}{(2\sigma_x)^2} \\ &= \frac{1}{4} \end{aligned}$$

7.2 Law of Large Numbers

Imagine you observe Independent and Identically Distributed (IID) random variables $X_1, X_2, \cdots, X_n$ distributed according to PMF p with expected value $\mathbb{E}[X_i] = \mu$. We have the sample mean M_n, where

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

Notice that M_n is a random variable since it is a function of random variables. It is, therefore, different from μ.

The goal of the Law of Large Numbers is to show:

$$M_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \longrightarrow \mu \quad (n \rightarrow \infty)$$

Proof.

$$\mathbb{E}[\frac{X_1 + X_2 + \cdots + X_n}{n}] = \mathbb{E}[\frac{X_1}{n}] + \mathbb{E}[\frac{X_2}{n}] + \cdots + \mathbb{E}[\frac{X_n}{n}] = \frac{\mu}{n} + \frac{\mu}{n} + \cdots + \frac{\mu}{n} = \mu$$

■

For variance, we have

$$\begin{aligned} \text{Var}\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) &= \text{Var}\left(\frac{X_1}{n}\right) + \text{Var}\left(\frac{X_2}{n}\right) + \cdots + \text{Var}\left(\frac{X_n}{n}\right) \\ &= \frac{\sigma^2}{n^2} + \frac{\sigma^2}{n^2} + \cdots + \frac{\sigma^2}{n^2} \\ &= \frac{\sigma^2}{n} \end{aligned}$$

Suppose $X_1, X_2, \cdots, X_n$ are IID with mean $\mathbb{E}[X_i] = \mu$ and variance $\text{Var}[X_i] = \sigma^2$. Then, Chebyshev's inequality shows

$$\mathbb{P}\left(|M_n - \mu| \geq \epsilon\right) \leq \frac{\text{Var}[M_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}$$

Theorem 7.2.1 (Weak Law of Large Numbers (WLLN)). Suppose $X_1, X_2, \cdots$ is a sequence of IID random variables with expected value $E[X_i] = \mu$ and finite variance$\mathrm{Var}[X_i] < \infty$. Define $\overline{X_n} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then, for every $\epsilon > 0$:

$$\lim_{n\to\infty} \mathbb{P}(|\overline{X_n} - \mu| \geq \epsilon) = 0$$

Explanation. As the number of samples increases, i.e., $n \to \infty$, the probability of the sample mean being greater than the actual mean will decrease and converge to 0 for a sufficiently large number of samples. ⊛

Example (The Pollster's Problem). Suppose in a referendum we want to determine the fraction of the population that will vote "yes," and we use p to describe this fraction. Now you take n samples of people to know their poll, and then you want to come up with an M_n that approximates the true p since you would not be able to know the actual p. Then, what is the n that you should take to minimize the error?

Solution: Suppose the i-th randomly selected person polled X_i, where

$$X_i = \begin{cases} 1, & \text{if yes} \\ 0, & \text{if no} \end{cases},$$

and we have the fraction of "yes" in our sample described as M_n, where

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

We would like a "small error," e.g., $|M_n - p| < 0.01$. However, since we have no prior knowledge of p, we can only ensure a small error by bounding it with a small probability, i.e., the probability of the error exceeding a certain value should not exceed a specified percentage. Then, we can try $n = 10,000$. By Chebyshev's inequality, we have

$$\mathbb{P}(|M_{10,000} - p| \geq 0.01) \leq \frac{\sigma^2}{n\epsilon^2} = \frac{p(1-p)}{10^4 \cdot 10^{-4}} = p(1-p)$$

Also, p remains unknown. Since it is uniformly distributed from 0 to 1, we see that $p(1-p)$ forms a parabola, with the maximum value at $\frac{1}{4}$. Then we have

$$\mathbb{P}(|M_{10,000} - p| \geq 0.01) \leq \frac{1}{4} = 25\%$$

The above bound is, however, too large, and we want a smaller percentage, for example, 5%. We can then determine the n we need by

$$\frac{\frac{1}{4}}{n \cdot 10^{-4}} \leq \frac{5}{10^2} \iff n \geq \frac{10^6}{20} = 50,000,$$

which gives

$$\mathbb{P}(|M_{50,000} - p| \geq 0.01) \leq 0.05$$

7.3 Central Limit Theorem

Suppose a random variable X has expected value $\mathbb{E}[X] = \mu$ and variance $\mathrm{Var}[X] = \sigma^2$. Then, random variable $Z = \dfrac{1}{\sigma}(X - \mu)$ has zero expected value and unit variance: $\mathbb{E}[Z] = 0$, $\mathrm{Var}[Z] = 1$.

Consider IID random variables $X_1, X_2, \cdots, X_n$ has expected value $\mathbb{E}[X_i] = \mu$ and variance $\mathrm{Var}[X_i] = \sigma^2$ and define

$$S_n = X_1 + \cdots + X_n$$

Then, random variable $Z_n = \dfrac{S_n - n\mu}{\sigma\sqrt{n}}$ has zero expected value and unit variance.

Theorem 7.3.1 (Central Limit Theorem (CLT)). Consider IID random variables $X_1, X_2, \cdots$ with expected value $\mathbb{E}[X_i] = \mu$ and variance $\mathrm{Var}[X_i] = \sigma^2$. Then, if we define

$$Z_n = \frac{(X_1 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}},$$

we have

$$\mathbb{E}[Z_n] = \mathbb{E}\left[\frac{(X_1 + \cdots + X_n) - n\mu}{\sigma\sqrt{n}}\right] = \frac{n\mu - n\mu}{\sigma\sqrt{n}} = 0, \quad \mathrm{Var}[X_i] = \frac{\sigma\sqrt{n}}{\sigma\sqrt{n}} = 1.$$

The CDF of random variable Z_n will converge to the CDF of a standard normal distribution $\Phi(z)$, i.e. for every $z \in \mathbb{R}$

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

Therefore, according to the central limit theorem, the sum of many independent random numbers will approximately have a normal distribution.

Example. Suppose a package weighs X_i, i.i.d. with $X_i \sim \text{Exponential}\left(\frac{1}{2}\right)$. We now load the container with $n = 100$ packages. Find

(1) $\mathbb{P}(S_n \geq 210)$;

(2) a such that $\mathbb{P}(S_n \geq a) \approx 0.05$;

(3) largest n such that $\mathbb{P}(S_n \geq 210) \approx 0.05$;

(4) $\mathbb{P}(N \geq 100)$, where N is the number of packages loaded. We can load the container until weight exceed 210.

Solution:

(1)

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}, \mu = \sigma = 2$$

$$\mathbb{P}(S_n \geq 210) = \mathbb{P}(\frac{S_n - 200}{20} \geq \frac{210 - 200}{20}) = \mathbb{P}(Z \geq 0.5) = 1 - \mathbb{P}(Z < 0.5) = 1 - \Phi(0.5) = 0.3085$$

(2)

$$0.05 \approx \mathbb{P}(\frac{S_n - 200}{20} \geq \frac{a - 200}{20}) = \mathbb{P}(Z \geq \frac{a - 200}{20}) = 1 - \underbrace{\Phi(\frac{a - 200}{20})}_{0.95}$$

$$\frac{a - 200}{20} = 1.645 \implies a = 232.9$$

(3)

$$\mathbb{P}(\frac{S_n - 2n}{2\sqrt{n}} \geq \frac{210 - 2n}{2\sqrt{n}}) \approx 1 - \Phi(\frac{210 - 2n}{2\sqrt{n}}) \approx 0.05$$
$$\frac{210 - 2n}{2\sqrt{n}} = 1.645$$
$$n = 89$$

(4)

$$\mathbb{P}(N \geq 100) = \mathbb{P}\left(\sum_{i=1}^{100} X_i \leq 210\right) \approx \Phi\left(\frac{210 - 200}{20}\right) = \Phi(0.5) = 0.6915$$

Appendix A

Z TABLE

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990