

# Probability Theory

Ji Yong-hyeon

December 3, 2020

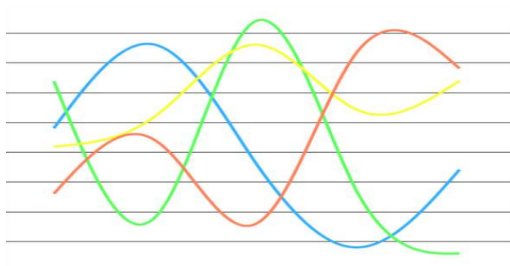
## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Introduction . . . . .	3
1.2	Gambling . . . . .	3
1.3	History of Probability . . . . .	5
1.4	Equally Likely Outcomes . . . . .	5
1.5	Basic Concepts . . . . .	5
1.6	Problems . . . . .	5
<b>2</b>	<b>Combinatorial Analysis</b>	<b>6</b>
2.1	Introduction . . . . .	6
2.2	The Basic Principle of Counting . . . . .	7
2.3	Permutations . . . . .	8
2.4	Combinations . . . . .	9
2.5	Multinomial Coefficients . . . . .	10
2.6	Problems . . . . .	12
<b>3</b>	<b>Axioms of Probability</b>	<b>13</b>
3.1	Sample Space and Events and Axioms of Probability . . . . .	13
3.2	Some Simple Properties . . . . .	15
3.3	Problems . . . . .	20
<b>4</b>	<b>Conditional Probability and Independence</b>	<b>21</b>
4.1	Introduction . . . . .	21
4.2	Conditional Probabilities . . . . .	21
4.3	Bayes' Formula . . . . .	23
4.4	Independent Events . . . . .	27
4.5	Bernoulli trials . . . . .	30
4.6	Problems . . . . .	33
<b>5</b>	<b>Discrete Random Variables</b>	<b>34</b>
5.1	Random Variables . . . . .	34
5.2	Discrete Random Variables . . . . .	34
5.3	Expectations and Uniform Discrete Random Variable . . . . .	35
5.4	Bernoulli Random Variable and Binomial Random Variable . . . . .	37
5.5	Poisson Random Variable . . . . .	40
5.6	Geometric Random Variables . . . . .	44
5.7	Problems . . . . .	44

<b>6</b>	<b>Continuous Random Variables</b>	<b>45</b>
6.1	Continuous Random Variables . . . . .	45
6.2	Uniform Random Variable . . . . .	47
6.3	Exponential Random Variable . . . . .	49
6.4	Normal Random Variable . . . . .	51
6.5	Exercises for Chapter 6 . . . . .	57
<b>7</b>	<b>Joint Distributions and Independence</b>	<b>60</b>
7.1	Discrete Case . . . . .	60
7.2	Continuous Case . . . . .	61
7.3	Conditional Distributions . . . . .	66
<b>8</b>	<b>Properties of Expectation and Variance</b>	<b>68</b>
8.1	Expectations . . . . .	68
8.2	Linearity and Monotonicity of Expectation . . . . .	69

# 1 Introduction

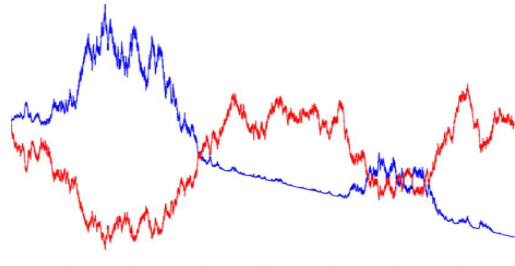
## 1.1 Introduction



- Continuous and **Differentiable**
- Predictable: Taylor series,

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x - a)^n.$$

- **Deterministic** world
- Mathematics of **Strong sense**
  - Definition of **Function**
- Classical mechanics
  - Newtonian mechanics
  - Law of universal gravitation

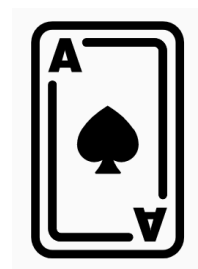


- Continuous but **Nowhere-Differentiable**
- **Unpredictable**
- **Non-Deterministic** world
- Mathematics of **Weak sense**
  - Definition of **Random variable**
  - **Probability Theory**
- Quantum mechanics
  - Particle physics
  - Brownian motion

## 1.2 Gambling

The theory of probability has always been associated with **gambling**.

- A **pair** coin gives you Heads( $H$ ) or Tails( $T$ ) with *equal* probability.
- A **pair** die gives you 1, 2, 3, 4, 5, or 6 with *equal* probability.
- A **shuffled deck of cards** means that any ordering of cards is *equally likely*.



**Example 1.1.**

1. Start with a *shuffled deck of cards* and distribute all 52 cards to 4 player, 13 cards to each. What is the probability that each player gets an Ace?
2. Next, assume that you are a player and you get a single Ace. What is the probability now that each player gets an Ace?

**Sol.**

1. If any ordering of cards is equally likely, then any position of the four Aces in the deck is also equally likely. There are

$$\binom{52}{4} = \frac{52!}{4!48!}$$

possibilities for the positions (slots) for the 4 aces. Out of these, the number of positions that give each player an Ace  $13^4$  pick the first slot among the cards that the first player gets, then the second slot among the second player's card, then the third and the fourth slot. Therefore, the answer is

$$\frac{13^4}{\binom{52}{4}} \approx 0.1055.$$

After you see that you have a single Ace, the probability goes up the previous answer need to be divided by the probability that you get a single Ace, which is

$$\frac{13 \cdot \binom{39}{3}}{\binom{52}{4}} \approx 0.4388.$$

Note that

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}.$$

The answer then becomes

$$\frac{13^4}{13 \cdot \binom{39}{3}} \approx 0.2404.$$

### 1.3 History of Probability

- **Blaise Pascal** (1623~1662)
- **Pierre de Fermat** (1607~1665)

1654 letter from the Flemish aristocrat and notorious gambler Chevalier de Méré to the mathematician and philosopher Blaise Pascal.

### 1.4 Equally Likely Outcomes

**Example 1.2.** In a family with 4 children, what is the probability of a 2:2 boy-girl split?

- One common wrong answer:  $1/5$ , as the 5 possibilities for the number of boys are not equally likely. 0:4, 1:3, 2:2, 3:1, 4:0 are not equally likely outcomes.
- Another common guess: close to 1, as this is the *most balanced* possibility.
- This represents the **mistaken belief** that symmetry in probabilities should very likely result in symmetry in the outcome.

Let  $X$  = the number of boys, then  $X \sim$  Binomial distribution with  $n = 4$ ,  $p = 1/2$ . Then

$$\Pr[2 : 2] = P(X = 2) = \binom{4}{2} \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2 = \frac{3}{8}.$$

Outcome	0 : 4	1 : 3	2 : 2	3 : 1	4 : 0
Probability	$\frac{1}{16}$	$\frac{4}{16}$	$\frac{6}{16}$	$\frac{4}{16}$	$\frac{1}{16}$

### 1.5 Basic Concepts

Suppose an experiment is performed, with  $n$  possible outcomes comprising a set  $S$ . Assume also that all outcomes are equally likely. Whether this assumption is *realistic* depends on the context. An *event*  $E$  is a set of outcomes, i.e.,  $E \subset S$ . If an event  $E$  consists of  $m$  different outcomes, then the probability of  $E$  is given by

$$P(E) = \frac{m}{n}.$$

The definition of probability only in a **finite sample space**  $S$  with equally likely outcomes (Combinatorics).

**Example 1.3.** A fair die has 6 outcomes; take  $E = \{2, 4, 6\}$ . Then  $P(E) = 1/2$ .

- The most straightforward interpretation is that for a *very large number* of rolls about half of the outcomes will be even.
- Note that this requires at least **the concept of a limit**.
- This *relative frequency* interpretation of probability will be explained in detail much later.

### 1.6 Problems

The problem of points, also called the problem of division of the stakes, is a classical problem in probability theory.

**Independent trials** resulting in a *success* with probability  $p$  and a *failure* with probability  $1 - p$  are performed. What is the probability that  $n$  successes occur before  $m$  failures? If we think of  $A$  and  $B$  as playing a game such that  $A$  gains 1 point when a success occurs and  $B$  gains 1 point when a failure occurs, then the desired probability is the probability that  $A$  would win if the game were to be continued in a position where  $A$  needed  $n$  and  $B$  needed  $m$  more points to win.

## 2 Combinatorial Analysis

### 2.1 Introduction

**Example 2.1.** Toss three fair coins. What is the probability of exactly one Heads (H)?

**Sol.** There are 8 equally likely outcomes:

$$S = \{HHH, HHT, HTH, HTT, TTH, THT, TTH, TTT\}.$$

Out of these, 3 have exactly one H, that is,

$$E = \{HTT, THT, TTH\} \subset S \quad \text{and} \quad P[E] = \frac{3}{8}.$$

□

**Example 2.2.** Let us now compute the probability of a 2:2 boy-girl split in a four-children family. We have 16 outcomes which we will *assume* are equally likely although this is not quite true in reality. We list the outcomes below, although we will soon see that there is a better way.

BBBB	BBBG	BBGB	BBGG
BGBB	BGBG	BGGB	BGGG
GBBB	GBBG	GBGB	GBGG
GGBB	GGBG	GGGB	GGGG

We conclude that

$$P[2:2 \text{ split}] = \frac{6}{16} = \frac{3}{8},$$

$$P[1:3 \text{ split or } 3:1 \text{ split}] = \frac{8}{16} = \frac{1}{2},$$

$$P[4:0 \text{ split or } 0:4 \text{ split}] = \frac{2}{16} = \frac{1}{8}.$$

**Example 2.3.** Roll two dice. What is the most likely sum?

**Sol.** Outcomes are ordered pairs  $(i, j)$ ,  $1 \leq i \leq 6, 1 \leq j \leq 6$ .

Sum	No. of outcomes
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Let  $S = \{(i, j) : 1 \leq i, j \leq 6\}$  then  $\#(S) = 36$ .  
And

$$E_7 = \{(i, j) : i + j = 7\}$$

$$= \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} \subset S.$$

Thus,

$$P[E_7] = \frac{\#(E_7)}{\#(S)} = \frac{6}{36} = \frac{1}{6}.$$

Therefore, our answer is 7, and  $P[\text{sum} = 7] = 6/36 = 1/6$ .

□

## 2.2 The Basic Principle of Counting

*How to count?*

Listing all outcomes is very inefficient, especially if their number is *large*. We will, therefore, learn a trivial, but conceptually important fact.

### The basic principle of counting

If an experiment consists of two stages and the first stage has  $m$  outcomes, while the second stage has  $n$  outcomes *regardless of the outcome at the first stage*, then the experiment as a whole has  $mn$  outcomes.

**Example 2.4.** Roll a die 4 times. What is the probability that you get different numbers?

**Sol.** First, we set the sample space  $S$  as follows:

$$S = \{(a, b, c, d) : a, b, c, d \in \{1, 2, 3, 4, 5, 6\}\}.$$

Let  $E$  be the event of different numbers then

$$E = \{(a, b, c, d) : a \neq b \neq c \neq d\} \subset S.$$

, so

$$P(E) = \frac{|E|}{|S|} = \frac{6 \cdot 5 \cdot 4 \cdot 3}{6 \cdot 6 \cdot 6 \cdot 6} = \frac{5}{18} (\approx 0.2778).$$

□

**Example 2.5.** Let us now compute probabilities for de Méré's games. In Game 1, there are 4 rolls and he wins with at least one 6. The number of good events is  $6^4 - 5^4$ , as the number of *bad* events is  $5^4$ . Therefore

$$P(\text{win}) = 1 - \left(\frac{5}{6}\right)^4 \approx 0.5177 > \frac{1}{2}.$$

In Game 2, there are 24 rolls of two dice and he wins by at least one pair of 6's rolled. The number of outcomes is  $36^{24}$ , the number of bad one is  $35^{24}$ , thus number of good outcomes equals  $36^{24} - 35^{24}$ . Therefore

$$P(\text{win}) = 1 - \left(\frac{35}{36}\right)^{24} \approx 0.4914 < \frac{1}{2}.$$

Chevalier de Méré overcounted the good outcomes in both cases. His count  $4 \cdot 6^3$  in Game 1 selects a die with a 6 and arbitrary numbers for other dice. However, many outcomes have more than one six and are hence counted more than once.

## 2.3 Permutations

Assume you have  $n$  object. The number of ways to fill  $n$  ordered slots with them is

$$n \cdot (n - 1) \cdots 2 \cdot 1 = n!,$$

while the number of ways to fill  $k \leq n$  ordered slot is

$$n \cdot (n - 1) \cdots (n - k + 1) = \frac{n!}{(n - k)!}.$$

**Example 2.6.** Shuffle a deck of cards.

- $P[\text{top card is an Ace}] = \frac{4 \cdot 51!}{52!} = \frac{1}{13}.$
- $P[\text{all cards of the same suit end up next to each other}] = \frac{4! \cdot (13!)^4}{52!} \approx 4.5 \times 10^{-28}.$
- $P[\text{heart are together}] = \frac{(39 + 1)! \cdot 13!}{52!} = \frac{40! \cdot 13!}{52!} = 6 \cdot 10^{-11}.$

- Random choice is made, this mean that *all available choices are equally likely*.
- **Statistics (sampling)**
  - *Sampling with replacement* refers to choosing, at random, an object from a population, noting its properties, putting the object back into the population, and then repeating.
  - Sampling without replacement omits the putting back part.

**Example 2.7.** A bag has 6 pieces of paper, each with one of the letters,  $E, E, P, P, P$ , and  $R$ , on it. Pull 6 pieces at random out of the bag (1) without, and (2) with replacement. What is the probability that these pieces, in order, spell  $PEPPER$ ?

**Sol.** For sampling without replacement:

1. An outcome is an ordering of the pieces of paper  $E_1 E_2 P_1 P_2 P_3 R$ .
2. The number of outcomes thus is  $6!$ .
3. The number of good outcomes is  $3!2!$ .

Thus the probability is  $\frac{3!2!}{6!}$ . Another solution is

$$\frac{3}{6} \times \frac{2}{5} \times \frac{2}{4} \times \frac{1}{3} \times \frac{1}{2} \times 1 = \frac{3!2!}{6!}.$$

For sampling with replacement, the answer is  $\frac{3^3 \cdot 2^2}{6^6} = \frac{1}{2 \cdot 6^3}$  quite a lot smaller. □



**Example 2.8.** Sit 3 men and 3 women at random (1) in a row of chairs and (2) around a table. Compute  $P[\text{all women sit together}]$ , In case (2), also compute  $P[\text{men and women alternate}]$ .

**Sol.** In case (1), the answer is  $\frac{4!3!}{6!} = \frac{1}{5}$ . For case (2), pick a man, say John Smith, and sit him first. Then, we reduce to a row problem with  $5!$  outcomes; the number of good outcomes is  $3! \cdot 3!$ . The answer is  $\frac{3}{10}$ .

For last question, the seats for the men and women are fixed after John Smith takes his seat and so the answer is  $\frac{3!2!}{5!} = \frac{1}{10}$ .  $\square$

**Example 2.9.** A group consists of 3 Norwegians, 4 Swedes, and 5 Finns, and they sit at random around a table. What is the probability that all groups end up sitting together?

**Sol.** The answer is  $\frac{3!4!5!2!}{11!}$ .  $\square$

## 2.4 Combinations

Let  $\binom{n}{k}$  be the number of different subsets with  $k$  elements of a set with  $n$  elements. Then

$$\begin{aligned}\binom{n}{k} &= \frac{n(n-1) \cdots (n-k+1)}{k!} \\ &= \frac{n!}{k!(n-k)!}.\end{aligned}$$

To understand why the above is true, first choose a subset, then order its elements in a row to fill  $k$  ordered slots with elements from the set with  $n$  objects. Then distinct choices of a subset and its ordering will end up as distinct orderings. Therefore,

$$\binom{n}{k} k! = n(n-1) \cdots (n-k+1).$$

$\binom{n}{k}$  is called  $n$  choose  $k$  or binomial coefficient as it appears in the binomial theorem:

$$(x+y)^n = \sum_{k=0}^n \binom{n}{k} x^k y^{n-k}$$

Also, note that

$$\binom{n}{0} = \binom{n}{n} = 0 \quad \text{and} \quad \binom{n}{k} = \binom{n}{n-k}.$$

## 2.5 Multinomial Coefficients

**Notation.** The number of ways to divide a set of  $n$  elements into  $r$  (distinguishable) subsets of  $n_1, n_2, \dots, n_r$  elements, where  $n_1 + \dots + n_r = n$ , is denoted by  $\binom{n}{n_1, \dots, n_r}$  and

$$\begin{aligned} \binom{n}{n_1, \dots, n_r} &= \binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \dots \binom{n-n_1-\dots-n_{r-1}}{n_r} \\ &= \frac{n!}{n_1! n_2! \dots n_r!}. \end{aligned}$$

To understand the slightly confusing word *distinguishable*, just think of painting  $n_1$  elements red, then  $n_2$  different elements blue, etc. These colors distinguish among the different subsets.

**Example 2.10.** A fair coin is tossed 10 times. What is the probability that we get exactly 5 Head?

**Sol.**  $P[\text{exactly 5 Heads}] = \binom{10}{5}/2^{10} (\approx 0.2461)$ , as one needs to choose the position of the five heads among 10 slots to fix a good outcome.  $\square$

**Example 2.11.** We have a bag that contains 100 balls, 50 of them red and 50 blue. Select 5 balls at random. What is the probability that 3 are blue and 2 are red?

**Sol.**

$$P[3 \text{ are blue and 2 are red}] = \frac{\binom{50}{3} \binom{50}{2}}{\binom{100}{5}} (\approx 0.3189).$$

$\square$

**Example 2.12.** Here we return Example 1.1 and solve it more slowly. Shuffle a standard deck of 52 cards and deal 13 cards to each of the 4 players. What is the probability that each player gets an Ace?

**Sol.** The first way uses an outcomes to be an ordering of 52 cards:

1. There are  $52!$  equally likely outcomes.
2. Let us first 13 cards go to the first player, the second 13 cards to the second player, etc. Pick a slot within each of the our segments of 13 slots for an Ace. There are  $13^4$  possibilities to choose these four slots for the Aces.
3. The number of choices to fill these four positions with (four different) Aces is  $4!$
4. Order the rest of the cards in  $48!$  ways.

The probability, then, is  $\frac{13^4 4! 48!}{52!} (\approx 0.1055)$ .

The second way, via a small leap of faith, assumes that each set of the our *positions of the four Aces* among the 52 shuffled cards is equally likely. You may choose to believe this intuitive fact or try to write down a formal proof: the number of permutations that result in a given set  $F$  of four positions is independent of  $F$ . Then:

1. The outcomes are the positions of the 4 Aces among the 52 slots for the shuffled cards of the deck.
2. The number of outcomes is  $\binom{52}{4}$ .
3. The number of good outcomes is  $13^4$ , as we need to choose one slot among 13 cards that go to the first player, etc.

The probability, then, is  $\frac{13^4}{\binom{52}{4}}$  which agrees with the number we obtained the first way.

Let us also compute  $P[\text{one person has all four Aces}]$ . Doing the problem the second way, we get

1. The number of outcomes is  $\binom{52}{4}$ .
2. To fix a good outcome, pick one player ( $\binom{4}{1}$  choices) and pick four slots for the Aces for that player ( $\binom{13}{4}$  choices).

The answer, then, is

$$\frac{\binom{4}{1} \binom{13}{4}}{\binom{52}{4}} (\approx 0.0106)$$

a lot smaller than the probability of each player getting. □

**Example 2.13.** Roll a die 12 times. What is the  $P[\text{each number appears exactly twice}]$ ?

*Sol.*

1. An outcome consists of filling each of the 12 slots (for the 12 rolls) with an integer between 1 and 6 (the outcomes of the roll).
2. The number of outcomes, therefore, is  $6^{12}$ .
3. To fix a good outcome, pick two slots for 1, then pick two slots for 2, with  $\binom{12}{2} \binom{10}{2} \cdots \binom{2}{2}$  choices.

The probability, then, is  $\frac{\binom{12}{2} \binom{10}{2} \cdots \binom{2}{2}}{6^{12}}$ .

Another question: what is the  $P[1 \text{ appears exactly 3 times, } 2 \text{ appears exactly 2 times}]$ ?

To fix a good outcome now, pick three slots for 1, two slots for 2, and fill the remaining 7 slots by numbers 3,  $\dots$ , 6. The number of choices is  $\binom{12}{3} \binom{9}{2} 4^7$ . □

**Example 2.14.** We have 14 rooms and 4 colors, white, blue, green, and yellow. Each room is painted at random with one of the four colors. There are  $4^{14}$  equally likely outcomes, so, for example,

$$P[5 \text{ white, 4 blue, 3 green, 2 yellow rooms}] = \frac{\binom{14}{5} \binom{9}{4} \binom{5}{3} \binom{2}{2}}{4^{14}}.$$

**Example 2.15.** A middle row on a plane seat 7 people. Three of them order chicken(C) and the remaining four pasta(P). The flight attendant returns with the meals, but has forgotten who ordered what and discovers that they are all asleep, so she puts the meals in front of them at random. what is the probability that all receive correct meals?

**Sol.** A reformulation makes the problem clearer: we are interested in  $P[3 \text{ people who orderd C get C}]$ . Let us label the people  $1, \dots, 7$  and assume that 1, 2 and 3 ordered  $C$ . The outcome is a selection of 3 people from the 7 who receive  $C$ , the number of them is  $\binom{7}{3}$ , and there is a *single* good outcomes. So, the answer is  $\frac{1}{\binom{7}{3}} = \frac{1}{35}$ . Similarly,

$$P[\text{no one who ordered C get C}] = \frac{\binom{4}{3}}{\binom{7}{3}} = \frac{4}{35},$$

$$P[\text{a single person who ordered C get C}] = \frac{3 \cdot \binom{4}{2}}{\binom{7}{3}} = \frac{18}{35},$$

$$P[\text{two persons who ordered C get C}] = \frac{\binom{3}{2} \cdot 4}{\binom{7}{3}} = \frac{12}{35}.$$

□

## 2.6 Problems

1. A California licence plate consists of a sequence of seven symbols: number, letter, letter, letter, number, number, number, where a letter is any one of 26 letters and a number is one among  $0, 1, \dots, 9$ . Assume that all licence plates are equally likely. (a) What is the probability that all symbols are different? (b) What is the probability that all symbols are different and the first number is the largest among the numbers?

**Sol.**

(a) content...

□

### 3 Axioms of Probability

How can we *mathematically* define a *random experiment*?

What we have are **outcomes** (which tell you exactly what happens), **events** (sets containing certain outcomes), and **probability** (which attaches to every event the likelihood that it happens).

**Fact 3.1.** *It is not possible to choose a uniform positive integer at random.*

*Proof.* Suppose that we have a uniform distribution over the positive integers and that

$$0 < \varepsilon < 1$$

is a candidate for the probability of a specific positive integer occurring. Then

$$\begin{aligned} P\left(\left\{1, 2, \dots, \left\lceil \frac{1}{\varepsilon} \right\rceil + 1\right\}\right) &= \varepsilon \left(\left\lceil \frac{1}{\varepsilon} \right\rceil + 1\right) \\ &\geq \varepsilon \left(\frac{1}{\varepsilon} + 1\right) = (1 + \varepsilon) \\ &> 1 \quad (\because \varepsilon \in (0, 1)). \end{aligned}$$

This is a contradiction as *no* probability may be strictly greater than 1. □

#### 3.1 Sample Space and Events and Axioms of Probability

A *probability space* is a triple  $(\Omega, \mathcal{F}, P)$ .

- $\Omega$  is an arbitrary set, representing *the set of outcomes*, sometimes called the **sample space**.
- $\mathcal{F}$  ( $\sigma$ -algebra) is a collection of events, i.e., a set of subsets of  $\Omega$ .
  - An **event**  $A \in \mathcal{F}$  is necessarily a subset of  $\Omega$ .
  - $\mathcal{F}$  needs to be a  $\sigma$ -algebra (measure theory):
    - (1)  $\emptyset \in \mathcal{F}$
    - (2)  $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
    - (3)  $A_1, A_2, \dots \in \mathcal{F} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .
  - We call events  $A_1, A_2, \dots$  are **pairwise disjoint** if  $A_i \cap A_j = \emptyset$  for  $i \neq j$ .
- The probability  $P$  is a number attached to every event  $A$  and satisfies the following three axioms:

**Axiom 1.** For every event  $A$ ,  $P(A) \geq 0$ .

**Axiom 2.**  $P(\Omega) = 1$ .

**Axiom 3.** If  $A_1, A_2, \dots$  is a sequence of pairwise disjoint events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) = \sum_{i=1}^{\infty} P(A_i) \quad (\text{Countable additivity}).$$

**Example 3.1.** For  $\Omega = \{1, 2, 3, 4, 5, 6\}$ ,

$$P(A) = \frac{(\text{number of elements in } A)}{6}.$$

The random experiment here is rolling a fair die. Clearly, this can be generalized to any finite set with equally likely outcomes.

**Example 3.2.** For  $\Omega = \{1, 2, \dots\} = \mathbb{N}$ , you have numbers  $p_1, p_2, \dots \geq 0$  with  $p_1 + p_2 + \dots = 1$ . For any  $A \subset \Omega$ ,

$$P(A) = \sum_{i \in A} p_i.$$

For example, toss a fair coin repeatedly until the first Heads. Your outcome is the number of tosses. Here,  $p_i = 1/2^i$ , so

$$\sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^i = \frac{1/2}{1 - 1/2} = 1.$$

Note that  $p_i$  cannot be chosen to be equal, as you cannot make the sum of infinitely many numbers to be 1:  $\forall p_i = p \geq 0$ ,

$$\sum_{i=1}^{\infty} p = \lim_{n \rightarrow \infty} \sum_{i=1}^n p = \lim_{n \rightarrow \infty} np = \infty \neq 1.$$

**Example 3.3.** Pick a point from inside the unit circle centered at the origin. Here  $\Omega = \{x, y\} : x^2 + y^2 < 1$  and

$$P(A) = \frac{\text{Area}(A)}{\text{Area}(\Omega)} = \frac{\text{Area}(A)}{\pi}.$$

It is important to observe that if  $A$  is singleton (a set whose element is a single point) then

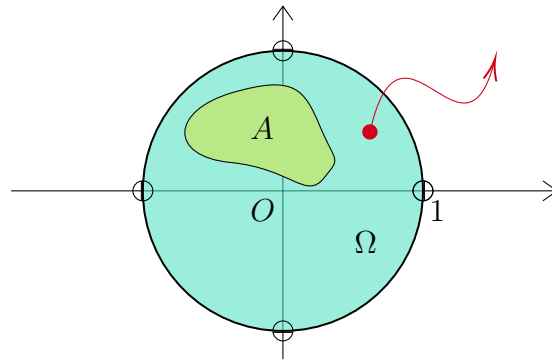


Figure 2

$P(A) = 0$ . This means that we cannot attach the probability to outcomes—you hit a single point (or even a line) with probability 0, but a “fatter” set with positive area you hit with positive probability.

### 3.2 Some Simple Properties

#### Consequences of the axioms:

(C0)  $P(\emptyset) = 0$ .

*Proof.* For  $A_1 = A_2 = \dots = \emptyset$ , by **Axiom 3**,

$$P(\emptyset) = P(\emptyset \cup \emptyset \cup \dots) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(\emptyset) = \lim_{n \rightarrow \infty} nP(\emptyset).$$

Suppose that  $P(\emptyset) \neq 0$ . Then  $P(\emptyset) = \lim_{n \rightarrow \infty} nP(\emptyset) = \infty$ . It is contradiction. Hence  $P(\emptyset) = 0$ .  $\square$

(C1) If  $A_1 \cap A_2 = \emptyset$ , then  $P(A_1 \cup A_2) = P(A_1) + P(A_2)$ .

*Proof.* Let  $A_i = \emptyset$  for  $i \geq 3$ . Suppose  $A_1 \cap A_2 = \emptyset$ . Then, by **Axiom 3**,

$$\begin{aligned} P(A_1 \cup A_2) &= P(A_1 \cup A_2 \cup \emptyset \cup \emptyset \cup \dots) \\ &= P(A_1 \cup A_2 \cup A_3 \cup A_4 \cup \dots) \\ &= P\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(A_i) \\ &= \lim_{n \rightarrow \infty} \left( \sum_{i=1}^2 P(A_i) + \sum_{i=3}^n P(A_i) \right) \\ &= \sum_{i=1}^2 P(A_i) + \lim_{n \rightarrow \infty} \sum_{i=3}^n P(A_i) \\ &= P(A_1) + P(A_2). \end{aligned}$$

$\square$

(C2)  $P(A^c) = 1 - P(A)$ .

*Proof.* Apply to (C1) to  $A_1 = A, A_2 = A^c$ . Then

$$P(\Omega) = P(A \cup A^c) = P(A) + P(A^c) = 1.$$

$\square$

(C3)  $0 \leq P(A) \leq 1$ .

*Proof.* By **Axiom 1**,  $0 \leq P(A)$ . Using that  $P(A^c) \geq 0$  in (C2),

$$P(A^c) = 1 - P(A) \geq 0 \implies P(A) \leq 1.$$

$\square$

(C4) If  $A \subset B$ ,  $P(B) = P(A) + P(B \setminus A) \geq P(A)$ .

*Proof.* Using (C1) for  $A_1 = A$  and  $A_2 = B \setminus A$ ,

$$P(B) = P(A) + P(B \setminus A)$$

and since  $P(B \setminus A) \geq 0$  by **Axiom 1**,  $P(B) \geq P(A)$ . □

$$(C5) \quad P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

*Proof.* Let  $P(A \setminus B) = p_1$ ,  $P(A \cap B) = p_{12}$ , and  $P(B \setminus A) = p_2$ . Note that  $A \setminus B$ ,  $A \cap B$ , and  $B \setminus A$  are pairwise disjoint. Then  $P(A) = p_1 + p_{12}$ ,  $P(B) = p_2 + p_{12}$ , and  $P(A \cup B) = p_1 + p_2 + p_{12}$ . □

(C6)

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= P(A_1) + P(A_2) + P(A_3) \\ &\quad - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) \\ &\quad + P(A_1 \cap A_2 \cap A_3) \end{aligned}$$

and more generally

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= \sum_{i=1}^n P(A_i) \\ &\quad - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &\quad + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) \\ &\quad + \dots + (-1)^{n-1} P(A_1 \cap \dots \cap A_n) \end{aligned}$$

This is called the **inclusion-exclusion formula** and is commonly used when it is easier to compute probabilities of intersection than of unions.

*Proof.* We prove this only for  $n = 3$ . Note that Fig 3.

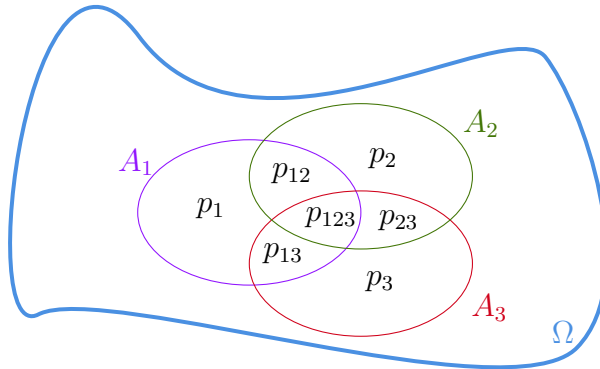


Figure 3



Let

$$p_1 = P(A_1 \cap A_2^c \cap A_3^c), \quad p_2 = P(A_1^c \cap A_2 \cap A_3^c), \quad p_3 = P(A_1^c \cap A_2^c \cap A_3),$$

$$p_{12} = P(A_1 \cap A_2 \cap A_3^c), \quad p_{13} = P(A_1 \cap A_2^c \cap A_3), \quad p_{23} = P(A_1^c \cap A_2 \cap A_3),$$

and  $p_{123} = P(A_1 \cap A_2 \cap A_3)$ . Then

$$\begin{aligned} & P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3) \\ &= (p_1 + p_{12} + p_{13} + p_{123}) + (p_2 + p_{12} + p_{23} + p_{123}) + (p_3 + p_{13} + p_{23} + p_{123}) \\ &- (p_{12} + p_{123}) - (p_{13} + p_{123}) - (p_{23} + p_{123}) + p_{123} \\ &= p_1 + p_2 + p_3 + p_{12} + p_{13} + p_{23} + p_{123} = P(A_1 \cup A_2 \cup A_3). \end{aligned}$$

□

**Example 3.4.** Pick an integer in  $[1, 1000]$  at random. Compute the probability that it is divisible neither by 12 nor by 15.

**Sol.** The sample space  $S$  consists of the 1000 integers between 1 and 1000, so  $|S| = 1000$ . Let  $A_r$  be the subset consisting of integers divisible by  $r$ . The cardinality of  $A_r$  is  $\left\lfloor \frac{1000}{r} \right\rfloor$ . Note that  $A_r \cap A_s = A_{\text{lcm}(r,s)}$ . Thus,

$$\begin{aligned} 1 - P(A_{12} \cap A_{15}) &= 1 - P(A_{12}) - P(A_{15}) + P(A_{12} \cap A_{15}) \\ &= 1 - P(A_{12}) - P(A_{15}) + P(A_{60}) \\ &= 1 - \frac{83}{1000} - \frac{66}{1000} + \frac{16}{1000} = 0.867. \end{aligned}$$

□

**Example 3.5.** Sit 3 men and 4 women at random in a row. What is the probability that either all the men or all women end up sitting together?

**Sol.** Let  $A_1 = \{\text{all women sit together}\}$ ,  $A_2 = \{\text{all men sit together}\}$ ,  $A_1 \cap A_2 = \{\text{both women and men sit together}\}$ , and so the answer is

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) = \frac{4!4!}{7!} + \frac{5!3!}{7!} - \frac{2!3!4!}{7!}.$$

□

**Example 3.6.** A group of 3 Norwegians, 4 Swedes, and 5 Finns is seated at random around a table. Compute the probability that at least one of the three groups end up sitting together.

**Sol.** Define  $A_N = \{\text{Norwegians sit together}\}$  and similarly  $A_S, A_F$ . Then we have

$$\begin{aligned} P(A_N) &= \frac{3!9!}{11!}, \quad P(A_S) = \frac{4!8!}{11!}, \quad P(A_F) = \frac{5!7!}{11!} \\ P(A_N \cap A_S) &= \frac{3!4!6!}{11!}, \quad P(A_N \cap A_F) = \frac{3!5!5!}{11!}, \quad P(A_S \cap A_F) = \frac{4!5!4!}{11!} \\ P(A_N \cap A_S \cap A_F) &= \frac{3!4!5!2!}{11!}. \end{aligned}$$

Therefore

$$P(A_N \cup A_S \cup A_F) = \frac{3!9! + 4!8! + 5!7! - 3!4!6! - 3!5!5! - 4!5!4! + 3!4!5!2!}{11!}.$$

□

**Example 3.7 (Matching Problem).** A large company with  $n$  employees has a scheme according to which each employee buys a Christmas gift and the gifts are then distributed at random to the employees. What is the probability that someone gets his or her own gift?

**Sol.** Note that this is different from asking, assuming that you are one of the employees, for the probability that *you* get your own gift, which is  $1/n$ .

Let  $A_i = \{i\text{-th employee gets his or her own gift}\}$ . Then, what we are looking for is

$$P\left(\bigcup_{i=1}^n A_i\right).$$

Note that

$$\begin{aligned} P(A_i) &= \frac{1}{n} \text{ (for all } i), \\ P(A_i \cap A_j) &= \frac{(n-2)!}{n!} = \frac{1}{n(n-1)} \text{ (for all } i < j), \\ P(A_i \cap A_j \cap A_k) &= \frac{(n-3)!}{n!} = \frac{1}{n(n-1)(n-2)} \text{ (for all } i < j < k), \\ P(A_1 \cap \cdots \cap A_n) &= \frac{1}{n!}. \end{aligned}$$

Recall that

$$\begin{aligned} P(A_1 \cup \cdots \cup A_n) &= \sum_{i=1}^n P(A_i) \\ &\quad - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) \\ &\quad + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) \\ &\quad + \cdots + (-1)^{n-1} P(A_1 \cap \cdots \cap A_n) \end{aligned}$$

and that

$$\begin{aligned} \exp(x) &= \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots, \\ \exp(x) &= \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n. \end{aligned}$$

Therefore, our answer is

$$\begin{aligned} n \cdot \frac{1}{n} - \binom{n}{2} \cdot \frac{1}{n(n-1)} + \binom{n}{3} \cdot \frac{1}{n(n-1)(n-2)} - \cdots + (-1)^{n-1} \frac{1}{n!} &= 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + (-1)^{n-1} \frac{1}{n!} \\ &= 1 - \frac{1}{e} \approx 0.6321 \text{ (as } n \rightarrow \infty), \end{aligned}$$

□

**Example 3.8 (Birthday Problem).** Assume that there are  $k$  people in the room. What is the probability that there are two who share a birthday? We will ignore leap years, assume all birthdays are equally likely, and generalize the problem a little: from  $n$  possible birthdays, sample  $k$  times with replacement.

$$P(\text{a shared birthday}) = 1 - P(\text{no shared birthdays}) = 1 - \frac{n(n-1) \cdots (n-k+1)}{n^k}.$$

When  $n = 365$ , the lowest  $k$  for which the above exceeds 0.5 is, famously,  $k = 23$ . Some values are given in the following table:

$k$	prob. for $n = 365$
10	0.1169
23	0.5073
41	0.9032
57	0.9901
70	0.9992

**Example 3.9 (Coupon Collector Problem).** Within the context of the previous problem, assume that  $k \geq n$  and compute  $P(\text{all } n \text{ birthdays are represented})$ .

**Sol.** This is described in terms of cereal boxes, each of which contains one of  $n$  different cards (coupons), chosen at random. If you buy  $k$  boxes, what is the probability that you have a complete collection?

When  $k = n$ , our probability is  $n!/n^n$ . More generally, let

$$A_i = \{i\text{-th birthday is missing}\}.$$

Then, we need to compute

$$1 - P\left(\bigcup_{i=1}^n A_i\right).$$

Now,

$$\begin{aligned} P(A_i) &= \frac{(n-1)^k}{n^k} \text{ (for all } i), \\ P(A_i \cap A_j) &= \frac{(n-2)^k}{n^k} \text{ for all } i < j, \\ P(A_1 \cap \cdots \cap A_n) &= 0 \end{aligned}$$

and our answer is

$$\begin{aligned} 1 - P\left(\bigcup_{i=1}^n A_i\right) &= 1 - n \left(\frac{n-1}{n}\right)^k + \binom{n}{2} \left(\frac{n-2}{n}\right)^k - \cdots + (-1)^{n-1} \binom{n}{n-1} \left(\frac{1}{n}\right)^k \\ &= \sum_{i=0}^n \binom{n}{i} (-1)^i \left(1 - \frac{i}{n}\right)^k. \end{aligned}$$

□

**Example 3.10.** Roll a die 12 times. Compute the probability that a number occurs 6 times and two other numbers occur three times each.

*Sol.* The number of outcomes is  $6^{12}$ . To count the number of good outcomes:

1. Pick the number that occurs 6 times:  $\binom{6}{1} = 6$  choices.
2. Pick the two numbers that occurs 3 times each:  $\binom{5}{2}$  choices.
3. Pick slots (rolls) for the number that occurs 6 times:  $\binom{12}{6}$  choices.
4. Pick slots (rolls) for one of the numbers that occurs 3 times:  $\binom{6}{3}$  choices.

Therefore, out probability is  $\frac{6\binom{5}{2}\binom{12}{6}\binom{6}{3}}{6^{12}}$ .

□

### 3.3 Problems

## 4 Conditional Probability and Independence

### 4.1 Introduction

**Example 4.1.** Assume that you have a bag with 11 cubes, 7 of which have a fuzzy surface and 4 are smooth. Out of the 7 fuzzy ones, 3 are red and 4 are blue; out of 4 smooth ones, 2 are red and 2 are blue. So, there are 5 red and 6 blue cubes. Other than color and fuzziness, the cubes have no other distinguishing characteristic.

Your experiment clearly has 11 outcomes. Consider the events  $R, B, F, S$  that selected cube is red, blue, fuzzy, and smooth, respectively. We observed  $P(R) = 5/11$ . For the probability of a red cube, *conditioned on it being fuzzy*, we do not have notation, so we introduce it here:  $P(R|F) = 3/7$ . Note that this also equals

$$\frac{P(R \cap F)}{P(F)} = \frac{3/11}{7/11} = \frac{3}{7}.$$

**Example 4.2.** Here is a question asked on *Wall Street job interviews*.

“Let’s play a game of Russian roulette. You are tied to your chair. Here’s a gun, a revolver. Here’s the barrel of the gun, six chambers, all empty. Now watch me as I put *two* bullets into the barrel, into *two adjacent chambers*. I close the barrel and spin it. I put a gun to your head and pull the trigger. Click. Lucky you! Now I’m going to pull the trigger one more time. Which would you prefer: that I *spin the barrel first* or that I *just pull the trigger*?”

Assume that the barrel rotates clockwise after the hammer hits and is pulled back. You are given the choice between an *unconditional* and a *conditional* probability of death.

- The former, if the barrel is spin again, remains  $1/3$ .
- The latter, if the trigger is pulled without the extra spin, equals the probability that the hammer clicked on an empty slot, which is next to a bullet in the counterclockwise direction, and equals  $1/4$ .

### 4.2 Conditional Probabilities

**Definition.** Assume that  $P(B) > 0$ . The **conditional probability of  $A$  given  $B$**  equals

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

For a fixed condition  $B$ , and acting on events  $A$ , the conditional probability  $Q(A) = P(A|B)$  satisfies the three axioms of probability. Thus,  $Q$  is another probability assignment all consequences of the axioms are valid for it.

**Example 4.3.** Toss two fair coins, blindfolded. Somebody tells you that you tossed at least one Heads. What is the probability that both tosses are Heads?

**Sol.** Here  $A = \{\text{both } H\}$ ,  $B = \{\text{at least one } H\}$ , and

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(\text{both } H)}{P(\text{at least one } H)} = \frac{1/4}{3/4} = \frac{1}{3}.$$

□

**Advanced.** If  $P(A) = 0$ , then the event  $A$  cannot be occurred. (**False**)

*Proof.*

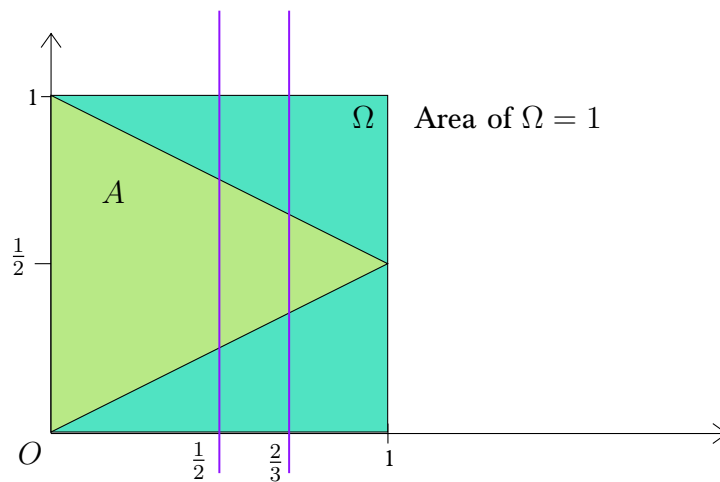


Figure 4

From Fig 4.,

$$P(A) = P[(x, y) \in A] = \frac{1}{2}$$

and

$$P(\{x = 1/2\}) = P(\{x = 2/3\}) = 0.$$

Then

$$P(A | \{x = 1/2\}) = \frac{1}{2},$$

$$P(A | \{x = 2/3\}) = \frac{1}{3}.$$

□

**Example 4.4.** Toss a coin 10 times. If you know (a) that exactly 7 Heads are tossed, (b) that at least 7 Heads are tossed, what is the probability that your first toss is Heads?

**Sol.** For (a),

$$P(\text{first toss } H | \text{exactly 7 H's}) = \frac{\binom{9}{6} \cdot \frac{1}{2^{10}}}{\binom{10}{7} \cdot \frac{1}{2^{10}}} = \frac{7}{10} = 0.7.$$

If you choose 7 tosses out of 10 at random, the first toss is included in your choice with probability 7/10.

For (b), the answer is, after canceling  $1/2^{10}$ ,

$$\frac{\binom{9}{6} + \binom{9}{7} + \binom{9}{8} + \binom{9}{9}}{\binom{10}{7} + \binom{10}{8} + \binom{10}{9} + \binom{10}{10}} = \frac{65}{88} \approx 0.7388.$$

Clearly, the answer should be a little larger than before, because this condition is more advantageous for Heads. □

### 4.3 Bayes' Formula

Conditional probabilities are sometimes *given*, or can be easily determined, especially in sequential random experiments. Then, we can use

$$P(A_1 \cap A_2) = P(A_1)P(A_2|A_1),$$

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2),$$

etc.

**Example 4.5.** An urn contains 10 black and 10 white balls. Draw 3 (a) without replacement, and (b) with replacement. What is the probability that all three are white?

**Sol.** Part (a):

1. Number of outcomes:  $\binom{20}{3}$ .
2. Number of ways to select 3 balls out of 10 white ones:  $\binom{10}{3}$ .

Our probability is then  $\frac{\binom{10}{3}}{\binom{20}{3}}$ .

To do this problem another way, imagine drawing the balls sequentially. Then, we are computing the probability of the intersection of the three events:  $P(\text{1st ball is white, 2nd ball is white, 3rd ball is white})$ . The relevant probabilities are:

1.  $P(\text{1st ball is white}) = 1/2$ .
2.  $P(\text{2nd ball is white} | \text{1st ball is white}) = 9/19$ .
3.  $P(\text{1st ball is white} | \text{1st two picked are white}) = 8/18$ .

Our probability is, then, the product  $1/2 \cdot 9/19 \cdot 8/18$ , which equals, as it must, what we obtained before.

This approach is particularly easy in case (b), where the previous colors of the selected balls do not affect the probabilities at subsequent stages. The answer, therefore, is  $(1/2)^3$  by

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2).$$

□

**Theorem 4.1 (First Bayes' Formula).** Assume that  $F_1, \dots, F_n$  are pairwise disjoint and that  $\bigcup_{i=1}^n F_i = \Omega$ , that is, exactly one of them always happens. In other words,  $F_i$  is partition of  $\Omega$ . Then for an event  $A$ ,

$$P(A) = P(F_1)P(A|F_1) + P(F_2)P(A|F_2) + \dots + P(F_n)P(A|F_n).$$

It is also called **Law of Total Probability**.

*Proof.*

$$\begin{aligned} P(A) &= P(A \cap \Omega) \\ &= P \left[ A \cap \left( \bigcup_{i=1}^n F_i \right) \right] \\ &= P \left[ (A \cap F_1) \cup (A \cap F_2) \cup \dots \cup (A \cap F_n) \right] \\ &= P(A \cap F_1) + P(A \cap F_2) + \dots + P(A \cap F_n) \\ &= P(F_1)P(A|F_1) + P(F_2)P(A|F_2) + \dots + P(F_n)P(A|F_n). \end{aligned}$$

□

We call an instance of using this formula “computing the probability by conditioning on which of the events  $F_i$  happens.” The formula is useful in sequential experiments, when you face different experimental conditions at the second stage depending on what happens at the first stage. Quite often, there are just two events  $F_i$ , that is, an event  $F$  and its complement  $F^c$ , and we are thus conditioning on whether  $F$  happens or not.

**Example 4.6.** Flip a fair coin. If you toss Heads, roll 1 die. If you toss Tails, roll 2 dice. Compute the probability that you roll exactly one 6.

**Sol.** Here you condition on the outcome of the coin toss, which could be Head (event  $F$ ) or Tail (event  $F^c$ ). If  $A = \{\text{exactly one 6}\}$ , then  $P(A|F) = 1/6$ ,  $P(A|F^c) = 2 \cdot 5/36$ ,  $P(F) = P(F^c) = 1/2$ , and so

$$P(A) = P(F)P(A|F) + P(F^c)P(A|F^c) = \frac{2}{9}.$$

□

**Example 4.7.** Roll a die, then select at random, without replacement, as many cards from the deck as the number shown on the die. What is the probability that you get at least one Ace?

**Sol.** Let  $F_i = \{\text{number shown on the die is } i\}$ , for  $i = 1, \dots, 6$ . Clearly,  $P(F_i) = 1/6$ . If  $A$  is the event that you get at least one Ace,

1.  $P(A|F_1) = 1/13$ ,
2. In general, for  $i > 1$ ,  $P(A|F_i) = 1 - \binom{48}{i} / \binom{52}{i}$ .

Therefore, by 1st Bayes' formula,

$$P(A) = \frac{1}{6} \left( \frac{1}{13} + 1 - \frac{\binom{48}{2}}{\binom{52}{2}} + 1 - \frac{\binom{48}{3}}{\binom{52}{3}} + 1 - \frac{\binom{48}{4}}{\binom{52}{4}} + 1 - \frac{\binom{48}{5}}{\binom{52}{5}} + 1 - \frac{\binom{48}{6}}{\binom{52}{6}} + 1 \right)$$

□



**Example 4.8** (*Coupon collector problem - revisited*). As promised, we will develop a computationally much better formula than the one in Example 3.9. This will be another example of *conditioning*, whereby you (1) *reinterpret* the problem as a sequential experiment and (2) use Bayes' formula with "conditions"  $F_i$  being relevant events at the first stage of the experiment.

Let  $P_{k,r}$  be the probability that exactly  $r$  (out of a total of  $n$ ) birthdays are represented among  $k$  people, we call the event  $A$ ; that is,

$$P(A) = p_{k,r} \quad (r \leq n \leq k).$$

We will fix  $n$  and let  $k$  and  $r$  be variable. Note that  $p_{k,r}$  is what we computed by the **inclusion-exclusion formula**.

At the first stage you have  $k - 1$  people; then the  $k$ -th person arrives on the scene. Let

- $F_1$  : the events that there are  $r$  birthdays represented among the  $k - 1$  people,
- $F_2$  : the events that there are  $r - 1$  birthdays represented among the  $k - 1$  people,
- $F_3$  : the events that any other number of birthdays represented among the  $k - 1$  people.

Clearly,  $P(A|F_3) = 0$  as the newcomer contributes either **0** or **1** new birthdays. Moreover,  $P(A|F_1) = r/n$ , the probability that the newcomer duplicates one of the existing  $r$  birthdays, and  $P(A|F_2) = (n - r)/n$  the probability that the newcomer does not duplicate any of the existing  $r - 1$  birthdays.

Therefore,

$$\begin{aligned} p_{k,r} = P(A) &= P(A|F_1)P(F_1) + P(A|F_2)P(F_2) \\ &= \frac{r}{n}p_{k-1,r} + \frac{n - r + 1}{n} \cdot p_{k-1,r-1} \end{aligned}$$

for  $k, r \geq 1$ , and this, together with the boundary conditions

$$\begin{aligned} p_{0,0} &= 1 \\ p_{k,r} &= 0, \text{ for } 0 \leq k < r, \\ p_{k,0} &= 0, \text{ for } k > 0, \end{aligned}$$

makes the computation fast and precise. In Example 3.9, by using the *inclusion-exclusion formula*, we have obtained that for

$$P(\text{complete collection}) = 1 - P\left(\bigcup_{i=1}^n A_i\right),$$

where  $A_i = \{i\text{-th birthday is missing}\}$ ,

$$\begin{aligned} 1 - P\left(\bigcup_{i=1}^n A_i\right) &= 1 - n \left(\frac{n-1}{n}\right)^k + \binom{n}{2} \left(\frac{n-2}{n}\right)^k - \cdots + (-1)^n \binom{n}{n-1} \left(\frac{1}{n}\right)^k \\ &= \sum_{i=0}^n \binom{n}{i} (-1)^i \left(1 - \frac{i}{n}\right)^k. \end{aligned}$$

**Theorem 4.2 (Second Bayes' Formula).** Let  $F_1, \dots, F_n$  and  $A$  be as in Theorem 4.1. Then

$$P(F_j|A) = \frac{P(F_j \cap A)}{P(A)} = \frac{P(A|F_j)P(F_j)}{P(A|F_1) + \dots + P(A|F_n)P(F_n)}.$$

An event  $F_j$  is often called a **hypothesis**,  $P(F_j)$  its **prior probability**, and  $P(F_j|A)$  its **posterior probability**.

Note that

$$P(A|B) = \frac{\text{Likelihood} \cdot \text{Prior}}{\text{Evidence}} = \frac{P(B|A) \cdot P(A)}{P(B)}.$$

**Example 4.9.** We have a fair coin and an unfair coin, which always comes out Heads. Choose one at random, toss it twice. It comes out Head both times. What is the probability that the coin is fair?

**Sol.** The relevant events are  $F = \{\text{fair coin}\}$ ,  $U = \{\text{unfair coin}\}$ , and  $B = \{\text{both tosses } H\}$ . Then  $P(F) = P(U) = 1/2$  (as each coin is chosen with equal probability). Moreover,  $P(B|F) = 1/4$ , and  $P(B|U) = 1$ . Our probability then is

$$P(F|B) = \frac{P(F)P(B|F)}{P(F)P(B|F) + P(U)P(B|U)} = \frac{1/2 \cdot 1/4}{1/2 \cdot 1/4 + 1/2 \cdot 1} = \frac{1}{5}.$$

□

**Example 4.10.** A factory has three machines,  $M_1$ ,  $M_2$ , and  $M_3$ , that produce items (say, lightbulbs). It is impossible to tell which machine produced a particular item, but some are defective. Here are the known numbers:

machine	proportion of item made	prob. any made item is defective
$M_1$	0.2	0.001
$M_2$	0.3	0.002
$M_3$	0.5	0.003

You pick an item, test it, and find it is defective. What is the probability that it was made by machine  $M_2$ ?

**Sol.** The best way to think about this random experiment is as a two-stage procedure. First you choose a machine with the probabilities given by the proportion. Then, that machine produces an item, which you then proceed to test. (Indeed, this is the same as choosing the item from a large number of them and testing it.)

Let  $D$  be the event that an item is defective and let  $M_i$  also denote the event that the item was made by machine  $i$ . Then,  $P(D|M_1) = 0.001$ ,  $P(D|M_2) = 0.002$ ,  $P(D|M_3) = 0.003$ ,  $P(M_1) = 0.2$ ,  $P(M_2) = 0.3$ ,  $P(M_3) = 0.5$ , and so

$$P(M_2|D) = \frac{0.002 \cdot 0.3}{0.001 \cdot 0.2 + 0.002 \cdot 0.3 + 0.003 \cdot 0.5} \approx 0.26.$$

□

**Example 4.11.** Assume 10% of people have a certain disease. A test gives the correct diagnosis with probability of 0.8; that is, if the person is sick, the test will be positive with probability 0.8, but if the person is not sick, the test will be positive with probability 0.2. A *random person* from the population has tested positive for the disease. What is the probability that he is actually sick? (No, it is not 0.8.)

**Sol.** Let us define the three relevant events:  $S = \{\text{sick}\}$ ,  $H = \{\text{healthy}\}$ ,  $T = \{\text{tested positive}\}$ . Now,  $P(H) = 0.9$ ,  $P(S) = 0.1$ ,  $P(T|H) = 0.2$ , and  $P(T|S) = 0.8$ . We are interested in  $S^c = H$ . Thus,

$$P(S|T) = \frac{P(T|S)P(S)}{P(T|S)P(S) + P(T|H)P(H)} = \frac{8}{26} \approx 0.31.$$

□

**Example 4.12** (*O. J. Simpson's first trial*, 1995). The famous sports star and media personality O. J. Simpson was on trial in Los Angeles for murder of his wife and her boyfriend. One of the main issues was whether Simpson's history of spousal abuse could be presented by prosecution at the trial; that is, whether this history was "probative," i.e., it had some evidentiary value, or whether it was merely "prejudicial" and should be excluded. Alan Dershowitz, a famous professor of law at Harvard and a consultant for the defense, was claiming the latter, citing the statistics that  $< 0.1\%$  of men who abuse their wives end up killing them. As J. F. Merz and J. C. Caulkins pointed out in the journal *Chance* (Vol. 8, 1995, pg. 14), this was the wrong probability to look at!

Dershowitz states that "a considerable number" of wife murderers had previously assaulted them, although "some" did not. So, we will (conservatively) say that  $P(A|M) = 0.5$ . (The two-stage experiment then is: choose a murdered woman at random; at the first stage, she is murdered by her partner, or not, with stated probabilities; at the second stage, she is among the abused women, or not, with probabilities depending on the outcome of the first stage.) By Bayes' formula,

$$P(M|A) = \frac{P(M)P(A|M)}{P(M)P(A|M) + P(M^c)P(A|M^c)} = \frac{(0.29)(0.5)}{(0.29)(0.5) + (0.71)(0.05)} = \frac{29}{36.1} \approx 0.8.$$

The law literature studiously avoids quantifying concepts such as probative value and reasonable doubt. Nevertheless, we can probably say that 80% is considerably too high compared to the prior probability of 29%, to use as a *sole* argument that the evidence is not probative.

## 4.4 Independent Events

**Definition (Independence).** Events  $A$  and  $B$  are **independent** if  $P(A \cap B) = P(A)P(B)$  and **dependent** (or **correlated**) otherwise.

Assume that  $P(B) > 0$ , one could rewrite the condition for independence,

$$P(A|B) = P(A),$$

so the probability of  $A$  is unaffected by knowledge that  $B$  occurred. Also, if  $A$  and  $B$  are independent.

**Note.** We claim that  $A$  and  $B$  are independent if and only if  $A$  and  $B^c$  are independent.

$$P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^c),$$

so  $A$  and  $B^c$  are also independent.

**Example 4.13.** Pick a random card from a full deck. Let  $A = \{\text{card is an Ace}\}$  and  $R = \{\text{card is red}\}$ . Are  $A$  and  $R$  independent?

**Sol.** We have  $P(A) = 1/13$ ,  $P(R) = 1/2$ , and, as there are two red Aces,  $P(A \cap R) = 2/52 = 1/26$ . The two events are independent—the proportion of aces among red card is the same as the proportion among all cards. That is,  $P(A \cap R) = P(A)P(R)$ .

Now, pick two cards out of the deck sequentially without replacement. Are  $F = \{\text{first card is an Ace}\}$  and  $S = \{\text{second card is an Ace}\}$  independent? Now  $P(F) = P(S) = 1/13$  and  $P(S|F) = 3/51$ , so they are not independent. Note that

$$P(S) = P(F)P(S|F) + P(F^c)P(S|F^c).$$

□

**Example 4.14.** Toss 2 fair coins and let  $F = \{\text{Heads on 1st toss}\}$ ,  $S = \{\text{Headson2ndtoss}\}$ . These are independent. You will notice that here the independence is in fact and assumption.

How do we define independence of more than two events? We say that events  $A_1, A_2, \dots, A_n$  are *independent* if

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2})P(A_{i_k}),$$

where  $1 \leq i_1 < i_2 < \dots < i_k \leq n$  are arbitrary indices. The occurrence of any combination of events does not influence the probability of others. Again, it can be shown that, in such a collection of independent events, we can replace an  $A_i$  by  $A_i^c$  and the events remain independent.

**Example 4.15.** Roll a four sided fair die, that is, choose one of the numbers 1, 2, 3, 4 at random. Let  $A = \{1, 2\}$ ,  $B = \{1, 3\}$ ,  $C = \{1, 4\}$ . Check that these are pairwise independent (each pair is independent), but not independent.

**Sol.** Indeed,  $P(A) = P(B) = P(C) = 1/2$  and  $P(A \cap B) = P(A \cap C) = P(B \cap C) = 1/4$  and pairwise independence follows. However,

$$P(A \cap B \cap C) = \frac{1}{4} \neq \frac{1}{8}.$$

The simple reason for lack of independence is  $(A \cap B) \subset C$ .

□

**Example 4.16.** You roll a die, your friend tosses a coin.

- If you roll 6, you win outright.
- If you do not roll 6 and your friend tossed Heads, you lose outright.
- If neither, the game is repeated until decided.

What is the probability that you win?

**Sol.** One way to solve this problem certainly is this:

$$\begin{aligned}
 P[\text{win}] &= P[\text{win on 1st round}] + P[\text{win on 2nd round}] + P[\text{win on 3rd round}] + \cdots \\
 &= \frac{1}{6} + \left(\frac{5}{6} \cdot \frac{1}{2}\right) \frac{1}{6} + \left(\frac{5}{6} \cdot \frac{1}{2}\right)^2 \frac{1}{6} + \cdots \\
 &= \frac{1}{6} \left[ 1 + \left(\frac{5}{12}\right) + \left(\frac{5}{12}\right)^2 + \cdots \right] = \frac{1}{6} \left( \frac{1}{1 - 5/12} \right) = \frac{2}{7}.
 \end{aligned}$$

*Important note: we have implicitly assumed independence between the coin and the die, as well as between different tosses and rolls. This is very common in problems such as this!*

You can avoid the nuisance, however, by the following trick. Let

$$\begin{aligned}
 D &= \{\text{game is decided on 1st round}\}, \\
 W &= \{\text{you win}\}.
 \end{aligned}$$

The events  $D$  and  $W$  are independent, which one can certainly check by computation, but, in fact, there is a very good reason to conclude so immediately. The crucial observation is that, provided that the game is not decided in the 1st round, you are thereafter facing the same game with the same winning probability; thus

$$P(W|D^c) = P(W).$$

In other words,  $D^c$  and  $W$  are independent and then so are  $D$  and  $W$ . and therefore

$$P(W) = P(W|D).$$

This means that one can solve this problem by computing the relevant probabilities for the 1st round:

$$P(W) = P(W|D) = \frac{P(W \cap D)}{P(D)} = \frac{1/6}{1/6 + 5/6 + 1/2} = \frac{2}{7},$$

which is our answer. □

**Example 4.17 (Craps).** Many casinos allow you to bet even money on the following game. Two dice are rolled and the sum  $S$  is observed.

- If  $S \in \{7, 11\}$ , you win immediately.
- If  $S \in \{2, 3, 12\}$ , you lose immediately.
- If  $S \in \{4, 5, 6, 8, 9, 10\}$ , the pair of dice is rolled repeatedly until one of the following happens:
  - $S$  repeats, in which case you win.
  - 7 appears, in which case you lose.

What is the winning probability?

**Sol.** Let us look at all possible ways to win:

1. For  $P(W) = P(F) + P(D)$ , where  $W$  is event of winning,  $F$  is win on 1<sup>st</sup>, and  $D$  is win on after 2<sup>nd</sup>, we have  $P(F) = 8/36$ .
2. Otherwise,  $P(D) = P(D_4) + P(D_5) + \cdots + P(D_{10})$ , where  $D_i$  is a event of “1<sup>st</sup> roll =  $i$  and win”.

Using Bayes’ formula  $P(D_i) = P(\{1^{\text{st}} \text{ roll}\} \cap \{\text{win}\}) = P(\text{roll } i | \text{roll } i \text{ or } 7)$ ,

$$P(\text{win}) = \frac{8}{36} + 2 \left( \frac{3}{36} \cdot \frac{1}{3} + \frac{4}{36} \cdot \frac{2}{5} + \frac{5}{36} \cdot \frac{5}{11} \right) \approx 0.4929,$$

a decent game by casino standards. □

## 4.5 Bernoulli trials

Assume  $n$  independent experiments, each of which is a success with probability  $p$  and, thus, failure with probability  $1 - p$ . In sequence of  $n$  **Bernoulli trials**,

$$P[\text{exactly } k \text{ successes}] = \binom{n}{k} p^k (1 - p)^{n-k}.$$

This is because the successes can occur on any subset  $S$  of  $k$  trials out of  $n$ , i.e., on any  $S \subset \{1, \dots, n\}$  with cardinality  $k$ . These possibilities are disjoint, as exactly  $k$  successes cannot occur on two different such sets. There are  $\binom{n}{k}$  such subsets; if we fix such an  $S$ , then successes must occur on  $k$  trials in  $S$  and failures on all  $n - k$  trials not in  $S$ ; the probability that this happens, by the assumed independence, is  $p^k (1 - p)^{n-k}$ .

**Example 4.18.** A machine produces items which are independently defective with probability  $p$ . Let us compute a few probabilities:

1.  $P[\text{exactly two items among the first 6 are defective}] = \binom{6}{2} p^2 (1 - p)^4.$
2.  $P[\text{exactly one items among the first 6 is defective}] = 1 - P[\text{no defects}] = 1 - (1 - p)^6.$
3.  $P[\text{at least 2 items among the first 6 are defective}] = 1 - (1 - p)^6 - 6p(1 - p)^5.$
4.  $P[\text{exactly 100 items are made before 6 defective are found}]$  equals  
 $P[100\text{th item defective, exactly 5 items among 1st 99 defective}] = p \binom{99}{5} p^5 (1 - p)^{94}.$

**Example 4.19 (Problem of Points).** This involves finding the probability of  $n$  successes before  $m$  failures in a sequence of Bernoulli trials. Let us call this probability  $p_{n,m}$ .

$$\begin{aligned} p_{n,m} &= P[\text{in the first } m + n - 1 \text{ trials, the number of successes is } \geq n] \\ &= \sum_{k=n}^{n+m-1} \binom{n+m-1}{k} p^k (1 - p)^{n+m-1-k}. \end{aligned}$$

The problem is solved, but it needs to be pointed out that computationally this is not the best formula. It is much more efficient to use the recursive formula obtained by conditioning on the outcome of the first trial. Assume that  $m, n \geq 1$ . Then,

$$\begin{aligned} p_{n,m} &= P[\text{first trial is a success}] \cdot P[n - 1 \text{ successes before } m \text{ failures}] \\ &= P[\text{first trial is a success}] \cdot P[n \text{ successes before } m - 1 \text{ failures}] \\ &= p \cdot p_{n-1,m} + (1 - p) \cdot p_{n,m-1}. \end{aligned}$$

Together with boundary conditions, valid for  $m, n \geq 1$ ,

$$p_{n,0} = 0, \quad p_{0,m} = 1,$$

which allows for very speedy and precise computations for large  $m$  and  $n$ .

**Example 4.20** (*Best of 7*). Assume that two equally matched teams,  $A$  and  $B$ , play a series of games and that the first team that wins four games is the overall winner of the series. As it happens, team  $A$  lost the first game. What is the probability it will win the series? Assume that the games are Bernoulli trials with success probability  $1/2$ . We have

$$\begin{aligned} P[A \text{ wins the series}] &= P[4 \text{ successes before 3 failures}] \\ &= p_{n,m}(n=4, m=3) \\ &= \sum_{k=4}^6 \binom{6}{k} \left(\frac{1}{2}\right)^6 = \frac{15+6+1}{2^6} \approx 0.3438. \end{aligned}$$

**Example 4.21 (Branch Matchbox Problem).** A mathematician carries two matchboxes, each originally containing  $n$  matches. Each time he needs a match, he is equally likely to take it from either box. What is the probability that, upon reaching for a box and finding it empty, there are exactly  $k$  matches still in the other box? Here,  $0 \leq k \leq n$ .

Let  $A_1$  be the event that matchbox 1 is the one discovered empty and that, at that instant, matchbox 2 contains  $k$  matches. Before this point, he has used  $n + n - k$  matches,  $n$  from matchbox 1 and  $n - k$  from matchbox 2. This means that he has reached for box 1 exactly  $n$  times in  $(n + n - k)$  trials and for the last time at the  $(n + 1 + n - k)$ -th trial. Therefore, our probability is

$$2P(A_1) = 2 \cdot \frac{1}{2} \binom{2n-k}{n} \frac{1}{2^{2n-k}} = \binom{2n-k}{n} \frac{1}{2^{2n-k}}.$$

**Example 4.22.** Each day, you independently decide, with probability  $p$ , to flip a fair coin. Otherwise, you do nothing. (a) What is the probability of getting exactly 10 Heads in the first 20 days? (b) What is the probability of getting 10 Heads before 5 Tails?

For (a), the probability of getting Heads is  $p/2$  independently each day, so the answer is

$$\binom{20}{10} \left(\frac{p}{2}\right)^{10} \left(1 - \frac{p}{2}\right)^{10}.$$

For (b), you can disregard days at which you do not flip, to get

$$\sum_{k=10}^{14} \binom{14}{k} \frac{1}{2^{14}} = p_{n,m}(n=10, m=5).$$

**Example 4.23.** You roll a die and your score is the number shown on the die. Your friend rolls five dice and his score is the number of 6's shown. Compute (a) the probability of event  $A$  that the two scores are equal and (b) the probability of event  $B$  that your friend's score is strictly larger than yours.

**Sol.** In both cases we will condition on your friend's score—this works a little better in case (b) than conditioning on your score. Let  $F_i (i = 0, \dots, 5)$  be the event that your friend's score is  $i$ . Then,  $P(A|F_i) = \frac{1}{6}$  if  $i \geq 1$  and  $P(A|F_0) = 0$ . Then, by the first Bayes' formula, we get

$$P(A) = \sum_{i=1}^5 P(F_i) \cdot \frac{1}{6} = \frac{1}{6} = \frac{1}{6}(1 - P(F_0)) = \frac{1}{6} - \frac{5^5}{6^5} \approx 0.0997.$$

Moreover,  $P(B|F_i) = i - 1/6$  if  $i \geq 2$  and 0 otherwise, and so

$$\begin{aligned} P(B) &= \sum_{i=1}^5 P(F_i) \cdot \frac{i-1}{6} \\ &= \frac{1}{6} \sum_{i=1}^5 i \cdot P(F_i) - \frac{1}{6} \sum_{i=1}^5 P(F_i) \\ &= \frac{1}{6} \sum_{i=1}^5 i \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{5-i} - \frac{1}{6} + \frac{5^5}{6^5}. \end{aligned}$$

Here, since

$$\begin{aligned} \frac{1}{6} \sum_{i=1}^5 i \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{5-i} &= \frac{1}{6} \sum_{i=1}^5 \frac{5 \cdot 4!}{(i-1)!(5-i)!} \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{5-i} \\ &= \frac{5}{6} \sum_{i=1}^5 \frac{4!}{(i-1)!(4-(i-1))!} \left(\frac{1}{6}\right) \left(\frac{1}{6}\right)^{i-1} \left(\frac{5}{6}\right)^{4-(i-1)} \\ &= \frac{5}{6} \cdot \frac{1}{6} \sum_{j=0}^4 \frac{4!}{j!(4-j)!} \left(\frac{1}{6}\right)^j \left(\frac{5}{6}\right)^{4-j} \quad (j = i-1) \\ &= \frac{5}{6} \cdot \frac{1}{6} \left(\frac{1}{6} + \frac{5}{6}\right)^4 \\ &= \frac{5}{6} \cdot \frac{1}{6}, \end{aligned}$$

thus, we have

$$\begin{aligned} P(B) &= \frac{1}{6} \sum_{i=1}^5 i \left(\frac{1}{6}\right)^i \left(\frac{5}{6}\right)^{5-i} - \frac{1}{6} + \frac{5^5}{6^5} \\ &= \frac{1}{6} \cdot \frac{5}{6} - \frac{1}{6} + \frac{5^5}{6^5} \approx 0.0392. \end{aligned}$$

□

**Note.** Prove or disprove that two events  $A$  and  $B$  are independent if and only if they are mutually exclusive.

**Disproof. Independence does not imply exclusiveness:** Toss two fair coins. Then  $\Omega = \{\text{HH}, \text{HT}, \text{TH}, \text{TT}\}$ . Let  $A = \{\text{HH}, \text{TT}\}$ ,  $B = \{\text{HH}, \text{TH}\}$ . Then

$$P(A \cap B) = P(\{\text{HH}\}) = 1/4 = P(A)P(B),$$

but  $A \cap B = \{\text{HH}\} \neq \emptyset$ ; that is,  $A$  and  $B$  are independent but not exclusive.

**Also exclusiveness does not imply independence:** Let  $A = \{\text{HH}\}$ ,  $B = \{\text{HT}, \text{TH}\}$ . Then

$$P(A \cap B) = P(\emptyset) = 0,$$

but  $P(A)P(B) \neq 0$ ; that is,  $A$  and  $B$  are exclusive (disjoint) but not independent.

□



## 4.6 Problems

## 5 Discrete Random Variables

### 5.1 Random Variables

**Definition.** A **random variable** is a number whose value depends upon the outcome of a random experiment. Mathematically, a random variable  $X$  is real-valued function on  $\Omega$  the space of outcomes:

$$X : \Omega \rightarrow \mathbb{R}.$$

Sometimes, when convenient, we also allow  $X$  to have the value  $\infty$  or, more rarely,  $-\infty$ , but this will not occur in this chapter. The crucial theoretical property that  $X$  should have is that, for each interval  $B$ , the set of outcomes for which  $X \in B$  is an event, so we are able to talk about its probability,  $P(X \in B)$ . Random variables are traditionally denoted by capital letters to distinguish them from deterministic quantities.

**Example 5.1.** Here are some examples of random variables.

1. Toss a coin 10 times and let  $X$  be the number of Heads.
2. Choose a random point in the unit square  $\{(x, y) : 0 \leq x, y \leq 1\}$  and let  $X$  be its distance from the origin.
3. Choose a random person in a class and let  $X$  be the height of the person, in inches.
4. Let  $X$  be value of the NASDAQ stock index at the closing of the next business day.

### 5.2 Discrete Random Variables

**Definition.** A **discrete random variable**  $X$  has finitely or countably many values  $x_i$ ,  $i = 1, 2, \dots$ , and  $p(x_i) = P(X = x_i)$  with  $i = 1, 2, \dots$ , is called the **probability mass function** of  $X$ . Sometimes  $X$  is added as the subscript of its p. m. f.  $p = p_X$ .

A probability mass function  $p$  has the following properties:

1. For all  $i$ ,  $p(x_i) > 0$ ; that is, we do not list values of  $X$  which occur with probability 0.
2. For any interval  $B$ ,  $P(X \in B) = \sum_{x_i \in B} p(x_i)$ .
3. As  $X$  must have some value,  $\sum_i p(x_i) = 1$ .

**Example 5.2.** Let  $X$  be the number of Heads in 2 fair coin tosses. Determine its probability mass function.

**Sol.** Possible values of  $X$  are 0, 1, and 2. Their probabilities are  $P(X = 0) = 1/4$ ,  $P(X = 1) = 1/2$ , and  $P(X = 2) = 1/4$ . You should note that random variable  $Y$ , which counts the number of Tails in the 2 tosses, has the same p. m. f., that is,  $p_X = p_Y$  but  $X$  and  $Y$  are far from being the same random variable! In general, random variables may have the same p. m. f., but may not even be defined on the same set of outcomes.  $\square$

**Example 5.3.** An urn contains 20 balls numbered  $1, \dots, 20$ . Select 5 balls at random, without replacement. Let  $X$  be the largest number among selected balls. Determine its p. m. f. and the probability that at least one of the selected number is 15 or more.

**Sol.** The possible values are  $5, \dots, 20$ . To determine the p. m. f., note that we have  $\binom{20}{5}$  outcomes, and, then,

$$P(X = i) = \frac{\binom{i-1}{4}}{\binom{20}{5}}, \quad i = 5, \dots, 20.$$

Finally,  $P(\text{at least one number of 15 or more}) = P(X \geq 15) = \sum_{i=15}^{20} P(X = i)$ .  $\square$

**Example 5.4.** An urn contains 11 balls, 3 white, 3 red, and 5 blue balls. Take out 3 balls at random, without replacement. You win \$1 for each red ball you select and lose a \$1 for each white ball you select. Determine the p. m. f. of  $X$ , the amount you win.

**Sol.** The number of outcomes is  $\binom{11}{3}$ .  $X$  can have values  $-3, -2, -1, 0, 1, 2$ , and  $3$ . Let us start with 0. This can occur with one ball of each color or with 3 blue balls:

$$P(X = 0) = \frac{3 \cdot 3 \cdot 5 + \binom{5}{3}}{\binom{11}{3}} = \frac{55}{165}.$$

To get  $X = 1$ , we can have 2 red and 1 white, or 1 red and 2 blue:

$$P(X = 1) = P(X = -1) = \frac{\binom{3}{2}\binom{3}{1} + \binom{3}{1}\binom{5}{2}}{\binom{11}{3}} = \frac{39}{165}.$$

The probability that  $X = -1$  is the same because of symmetry between the roles that the red and the white balls play. Next, to get  $X = 2$  we must have 2 red balls and 1 blue:

$$P(X = -2) = P(X = 2) = \frac{\binom{3}{2}\binom{5}{1}}{\binom{11}{3}} = \frac{15}{165}.$$

Finally, a sing outcome (3 red balls) produces  $X = 3$ :

$$P(X = -3) = P(X = 3) = \frac{1}{\binom{11}{3}} = \frac{1}{165}.$$

All the seven probabilities should add to 1, which can be used either to ehck the computations or to compute the seventh probability (say,  $P(X = 0)$ ) from the other six.  $\square$

### 5.3 Expectations and Uniform Discrete Random Variable

**Definition.** Assume that  $X$  is a discrete random variable with possible values  $x_i$ ,  $i = 1, 2, \dots$ . Then, the **expected value**, also called **expectation** of  $X$  is

$$EX = \sum_i x_i P(X = x_i) = \sum_i x_i p(x_i).$$

For any function  $g : \mathbb{R} \rightarrow \mathbb{R}$ ,

$$Eg(x) = \sum_i g(x_i) P(X = x_i).$$

**Example 5.5.** Let  $X$  be a random variable with  $P(X = 1) = 0.2$ ,  $P(X = 2) = 0.3$ , and  $P(X = 3) = 0.5$ . What is the expected value of  $X$ ?

**Sol.** We can, of course, just use the formula, but let us instead proceed intuitively and see that the definition makes sense. What, then, should the average of  $X$  be?

Imagine a large number  $n$  of repetitions of the experiment and measure the realization of  $X$  in each. By the frequency interpretation of probability, about  $0.2n$  realizations will have  $X = 1$ , about  $0.3n$  will have  $X = 2$ , and about  $0.5n$  will have  $X = 3$ . The average value of  $X$  then should be

$$\frac{1 \cdot 0.2n + 2 \cdot 0.3n + 3 \cdot 0.5n}{n} = 2.3$$

which of course is the same as the formula gives. □

**Note.** The most natural way would certainly be  $E[X - \mu]$ . The only problem with this is that absolute values are annoying. Instead, we define the **variance** of  $X$

$$\text{Var}(X) = E(x - \mu)^2.$$

The quantity that has the correct units is the **standard deviation**

$$\sigma(X) = \sqrt{\text{Var}(X)} = \sqrt{E(X - \mu)^2}.$$

We will give another, more convenient, formula for variance that will use the following property of expectation, called *linearity*:

$$E(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 E X_1 + \alpha_2 E X_2,$$

valid for any random variable  $X_1$  and  $X_2$  and non-random constant  $\alpha_1$  and  $\alpha_2$ . This property will be explained and discussed in more detail later. Then

$$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - \mu^2 \quad (\because E(X) = \mu) \\ &= E(X^2) - (EX)^2. \end{aligned}$$

In computation, bear in mind that variance cannot be negative! Furthermore, the only way that a random variable has 0 variance is when it is equal to its expectation  $\mu$  with probability 1 (so it is not really random at all):  $P(X = \mu) = 1$ . Here is the summary:

$$\text{Var}(X) = E(X - EX)^2 = E(X^2) - (EX)^2.$$

**Example 5.6.** Previous example, continued. Compute  $\text{Var}(X)$ .

**Sol.** Since  $E(X^2) = 1^2 \cdot 0.2 + 2^2 \cdot 0.3 + 3^2 \cdot 0.5 = 5.9$  and  $(EX)^2 = (2.3)^2 = 5.29$ ,

$$\text{Var}(X) = 0.61 \quad \text{and} \quad \sigma(X) \approx 0.7810.$$

□

**Definition. Uniform discrete random variable** is a random variable with values  $x_1, \dots, x_n$ , each with equal probability  $1/n$ . Such a random variable is simply the random choice of one among  $n$  numbers.

✓ **Properties:**

$$1. EX = \frac{\sum_{i=1}^n x_i}{n}.$$

$$2. \text{Var}X = \frac{\sum_{i=1}^n x_i^2}{n} - \left( \frac{\sum_{i=1}^n x_i}{n} \right)^2.$$

**Example 5.7.** Let  $X$  be the number shown on a rolled fair die. Compute  $EX$ ,  $E(X^2)$ , and  $\text{Var}(X)$ .

*Sol.*

$$\begin{aligned} EX &= \frac{1 + 2 + \dots + 6}{6} = \frac{7}{2}, \\ EX^2 &= \frac{1^2 + 2^2 + \dots + 6^2}{6} = \frac{91}{6}, \\ \text{Var}X &= \frac{91}{6} - \left( \frac{7}{2} \right)^2 = \frac{35}{12}. \end{aligned}$$

□

## 5.4 Bernoulli Random Variable and Binomial Random Variable

**Definition.** Bernoulli random variable is also called **indicator random variable**. Assume that  $A$  is an event with probability  $p$ . Then,  $I_A$ , the indicator of  $A$ , is given by

$$I_A = \begin{cases} 1, & \text{if } A \text{ happens} \\ 0, & \text{otherwise} \end{cases}.$$

Other notations for  $I_A$  include  $1_A$  and  $\chi_A$ . Although simple, such random variables are very important as building blocks for more complicated random variables.

✓ **Properties:**

$$1. EI_A = p.$$

$$2. \text{Var}(I_A) = p(1 - p).$$

For the variance, note that  $I_A^2 = I_A$ , so that  $E(I_A^2) = EI_A = p$ .

A **Binomial** $(n, p)$  **random variable** counts the number of successes in  $n$  independent trials, each of which is a success with probability  $p$ .

✓ **Properties:**

1. Probability mass function:  $P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i}, \quad i = 0, 1, \dots, n.$
2.  $EX = np$
3.  $\text{Var}(X) = np(1 - p).$

**Example 5.8.** Let  $X$  be the number of Heads in 50 tosses of a fair coin. Determine  $EX$ ,  $\text{Var}(X)$  and  $P(X \leq 10)$ ? As  $X$  is Binomial $(50, 1/2)$ , so  $E(X) = 25$ ,  $\text{Var}(X) = 12.5$ , and

$$P(X \leq 10) = \sum_{i=0}^{10} \binom{50}{i} \frac{1}{2^{50}}.$$

From the definition of expectation,

$$E(X) = \sum_{x \in \Omega_X} x \Pr(X = x).$$

Thus,

$$\begin{aligned}
 E(X) &= \sum_{k=0}^n k \binom{n}{k} p^k q^{n-k} && \text{Definition of Binomial Distribution with } p + q = 1 \\
 &= \sum_{k=1}^n k \binom{n}{k} p^k q^{n-k} && \text{for } k = 0, k \binom{n}{k} p^k q^{n-k} = 0 \\
 &= \sum_{k=1}^n n \binom{n-1}{k-1} p^k q^{n-k} && \text{Factors of Binomial Coefficient: } k \binom{n}{k} = n \binom{n-1}{k-1} \\
 &= np \sum_{k=1}^n \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} && \text{Taking out } np, \text{ using } n - k = (n - 1) - (k - 1) \\
 &= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} && \text{Substituting } m = n - 1, j = k - 1 \\
 &= np(p + q)^{n-1} && \text{By the Binomial Theorem} \\
 &= np && p + q = 1
 \end{aligned}$$

Note that

$$\begin{aligned}
E(X^2) &= \sum_{k=0}^n \binom{n}{k} p^k q^{n-k} && \text{Definition of Binomial Distribution: } p + q = 1 \\
&= \sum_{k=0}^n k n \binom{n-1}{k-1} p^k q^{n-k} && \text{Factors of Binomial Coefficient: } k \binom{n}{k} = n \binom{n-1}{k-1} \\
&= np \sum_{k=1}^n k \binom{n-1}{k-1} p^{k-1} q^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{m=n-1} (j+1) \binom{m}{j} p^j q^{m-j} && \text{Substituting } j = k-1, m = n-1 \\
&= np \left( \sum_{j=0}^m j \binom{m}{j} p^j q^{m-j} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) \\
&= np \left( \sum_{j=0}^m m \binom{m-1}{j-1} p^j q^{m-j} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) && \because j \binom{m}{j} = m \binom{m-1}{j-1} \\
&= np \left( (n-1)p \sum_{j=0}^m \binom{m-1}{j-1} p^{j-1} q^{(m-1)-(j-1)} + \sum_{j=0}^m \binom{m}{j} p^j q^{m-j} \right) \\
&= np ((n-1)p(p+q)^{m-1} + (p+q)^m) && \text{Binomial Theorem} \\
&= np((n-1)p + 1) && \because p + q = 1 \\
&= n^2 p^2 + np(1-p).
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\text{Var}(X) &= E(X^2) - (EX)^2 \\
&= np(1-p) + n^2 p^2 - (np)^2 \\
&= np(1-p).
\end{aligned}$$

**Example 5.9.** Denote by  $d$  the dominant gene and by  $r$  the recessive gene at a single locus. Then  $dd$  is called the pure dominant genotype,  $dr$  is called the hybrid, and  $rr$  the pure recessive genotype. The two genotypes with at least one dominant gene,  $dd$  and  $dr$ , result in the phenotype of the dominant gene, while  $rr$  result in a recessive phenotype.

Assuming that both parents are hybrid and have  $n$  children, what is the probability that at least two will have the recessive phenotype? Each child, independently, gets one of the genes at random from each parent.

For each child, independently, the probability of the  $rr$  genotype is  $1/4$ . If  $X$  is the number of  $rr$  children, then  $X$  is  $\text{Binomial}(n, 1/4)$ . Therefore,

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - \left(\frac{3}{4}\right)^n - n \cdot \frac{1}{4} \left(\frac{3}{4}\right)^{n-1}.$$

## 5.5 Poisson Random Variable

**Definition.** A random variable is **Poisson( $\lambda$ )**, with parameter  $\lambda > 0$ , if it has the probability mass function given as follows:

$$P(X = i) = \frac{\lambda^i}{i!} e^{-\lambda} \quad \text{for } i = 0, 1, 2, \dots$$

✓ **Properties:**

$$E(X) = \text{Var}(X) = \lambda.$$

*Proof.*

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} = \sum_{x=1}^{\infty} x \frac{e^{-\lambda} \lambda^x}{x!} \\ &= \sum_{x=1}^{\infty} \frac{e^{-\lambda} \lambda^x}{(x-1)!} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} \left( \frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \dots \right) \\ &= \lambda e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda. \end{aligned}$$

Taylor Series Expansion for Exponential Function

And since

$$\begin{aligned} E(X^2) &= \sum_{k \geq 0} k^2 \frac{1}{k!} \lambda^k e^{-\lambda} = \lambda e^{-\lambda} \sum_{k \geq 1} k \frac{1}{(k-1)!} \lambda^{k-1} \\ &= \lambda e^{-\lambda} \left( \sum_{k \geq 1} (k-1) \frac{1}{(k-1)!} \lambda^{k-1} + \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \right) \\ &= \lambda e^{-\lambda} \left( \lambda \sum_{k \geq 1} \frac{1}{(k-2)!} \lambda^{k-2} + \sum_{k \geq 1} \frac{1}{(k-1)!} \lambda^{k-1} \right) \\ &= \lambda e^{-\lambda} \left( \lambda \sum_{i \geq 0} \frac{1}{i!} \lambda^i + \sum_{j \geq 0} \frac{1}{j!} \lambda^j \right) \quad \text{Substituting } i = k-2, j = k-1 \\ &= \lambda e^{-\lambda} (\lambda e^{\lambda} + e^{\lambda}) \\ &= \lambda(\lambda + 1) = \lambda^2 + \lambda, \end{aligned}$$

we have

$$\begin{aligned} \text{Var}(X) &= E(X^2) - (E(X))^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda. \end{aligned}$$

□



**Note.** The Poisson random variable is useful as an approximation to a Binomial random variable when *the number of trials is large and the probability of success is small*. In this context it is often called the **law of rare events**, first formulated by L. J. Bortkiewicz (in 1898), who studied deaths by horse kicks in the Prussian cavalry.

**Theorem 5.1 (Poisson approximation to Binomial).** *When  $n$  is large,  $p$  is small, and  $\lambda = np$  is of moderate size,  $\text{Binomial}(n, p)$  is approximately  $\text{Poisson}(\lambda)$ . That is, if  $X$  is  $\text{Binomial}(n, p)$ , with  $p = \lambda/n$ , then, as  $n \rightarrow \infty$ ,*

$$P(X = i) \rightarrow e^{-\lambda} \frac{\lambda^i}{i!}$$

for each fixed  $i = 0, 1, 2, \dots$ .

*Proof.*

$$\begin{aligned} P(X = i) &= \binom{n}{i} p^i (1-p)^{n-i} = \binom{n}{i} \left(\frac{\lambda}{n}\right)^i \left(1 - \frac{\lambda}{n}\right)^{n-i} \\ &= \frac{n(n-1) \cdots (n-i+1)}{i!} \frac{\lambda^i}{n^i} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-i} \\ &= \frac{\lambda^i}{i!} \cdot \left(1 - \frac{\lambda}{n}\right)^n \cdot \frac{n(n-1) \cdots (n-i+1)}{n^i} \cdot \frac{1}{\left(1 - \frac{\lambda}{n}\right)^i} \\ &\rightarrow \frac{\lambda^i}{i!} \cdot e^{-\lambda} \cdot 1 \cdot 1, \quad \text{as } n \rightarrow \infty. \end{aligned}$$

□

**Example 5.10.** Suppose that the probability that a person is killed by lightning in a year is, independently,  $1/500M$ . Assume that the US population is  $300M$ .

1. Compute  $P(3 \text{ or more people will be killed by lightning next year})$  exactly.

If  $X$  is the number of people killed by lightning, then  $X$  is  $\text{Binomial}(n, p)$ , where  $n = 300M$ , and  $p = 1/500M$ , and the answer is

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \\ &= 1 - (1-p)^n - np(1-p)^{n-1} - \binom{n}{2} p^2 (1-p)^{n-2} \\ &\approx 0.02311530. \end{aligned}$$

2. Approximate the above probability.

As  $\lambda = np = 3/5$ ,  $X$  is approximately  $\text{Poisson}\left(\frac{3}{5}\right)$ , and the answer is

$$1 - e^{-\lambda} - \lambda e^{-\lambda} - \frac{\lambda^2}{2} e^{-\lambda} \approx 0.02311529.$$

3. Approximate  $P(\text{two or more people are killed by lightning within the first 6 months of next year})$ , that is,  $P(Y \geq 2) = 1 - P(Y = 0) - P(Y = 1)$ ,  $Y \sim \text{Poisson}(\lambda')$ .

This highlights the interpretation of  $\lambda$  as a *rate*. If lightning deaths occur at the rate of  $3/5$  a year, they should occur at half that rate in 6 months. Indeed, assuming that lightning deaths occur as a result of independent factors in disjoint time intervals, we can imagine that they operate on different people in disjoint time intervals.

Thus, *doubling* the time interval is the same as doubling the number  $n$  of people (while keeping  $p$  the same), and then  $np$  also doubles. Consequently, having the time interval has the same  $p$ , but half as many trials, so  $np$  changes to  $3/10$  and so  $\lambda' = 3/10$  as well. The answer is, for  $\lambda' = (n/2)p = np/2 = 3/10$ ,

$$1 - e^{-\lambda'} - \lambda' e^{-\lambda'} \approx 0.0369.$$

4. Approximate  $P(\text{in exactly 3 of next 10 years exactly 3 people are killed by lightning})$ .

In every year, the probability of exactly 3 deaths is approximately  $\frac{\lambda^3}{3!} e^{-\lambda}$ , where, again,  $\lambda = 3/5$ . Assuming year-to-year independence, the number of years with exactly 3 people killed is approximately  $U \sim \text{Binomial}\left(10, \frac{\lambda^3}{3!} e^{-\lambda}\right)$ . Then answer is

$$P(U = 3) = \binom{10}{3} \left(\frac{\lambda^3}{3!} e^{-\lambda}\right)^3 \left(1 - \frac{\lambda^3}{3!} e^{-\lambda}\right)^7 \approx 4.34 \cdot 10^{-6}.$$

5. Compute the expected number of years, among the next 10, in which 2 or more people are killed by lightning; that is,  $E[V] = np$  where  $V \sim \text{Binomial}(n, p)$ ,  $n = 10$ , and  $p = P[x \geq 2]$ .

By the same logic as above and the formula for Binomial expectation, the answer is

$$10(1 - e^{-\lambda} - \lambda e^{-\lambda}) \approx 0.3694.$$

**Example 5.11** (*Poisson distribution and law*). Assume a crime has been committed. It is known that the perpetrator has certain characteristics, which occur with a small frequency  $p$  (say,  $10^{-8}$ ) in a population of size  $n$  (say,  $10^8$ ). A person who matches these characteristics has been found at random (e.g., at a routine traffic stop or by airport security) and, since  $p$  is so small, charge with the crime. *There is no other evidence*. What should the defence be?

**Sol.** Let us start with a mathematical model of this situation. Assume that  $N$  is the number of people with given characteristics. This is a Binomial random variable but, given the assumptions, we can easily assume that it is Poisson with  $\lambda = np$ . Choose a person from among these  $N$ , label that person by  $C$ , the criminal. Then, choose at random another person,  $A$ , who is arrested. The question is whether  $C = A$ , that is, whether the arrested person is guilty. Mathematically, we can formulate the problem as follows:

$$P(C = A | N \geq 1) = \frac{P(C = A, N \geq 1)}{P(N \geq 1)}$$

We need to condition as the experiment cannot even be performed when  $N = 0$ . Now, by the first Bayes' formula,

$$\begin{aligned} P(C = A, N \geq 1) &= \sum_{k=1}^{\infty} P(C = A, N \geq 1 | N = k) \cdot P(N = k) \\ &= \sum_{k=1}^{\infty} P(C = A | N = k) \cdot P(N = k) \end{aligned}$$

and  $P(C = A | N = k) = 1/k$ , so  $P(C = A, N \geq 1) = \sum_{k=1}^{\infty} \frac{1}{k} \frac{\lambda^k}{k!} e^{-\lambda}$ .

The probability that the arrested person is guilty then is

$$P(C = A | N \geq 1) = \frac{e^{-\lambda}}{1 - e^{-\lambda}} \sum_{k=1}^{\infty} \frac{\lambda^k}{k k!}.$$

There is no closed-form expression for the sum, but it can be easily computed numerically. The defense may claim that the probability of innocence,  $1 -$  (the above probability), is about 0.2330 when  $\lambda = 1$ , presumably enough for a reasonable doubt.  $\square$

## 5.6 Geometric Random Variables

**Definition.** A Geometric( $p$ ) random variable  $X$  counts the number trials required for the first success in independent trials with success probability  $p$ . The probability mass function is

$$P(X = n) = n(1 - p)^{n-1},$$

where  $n = 1, 2, \dots$ .

✓ **Properties:**

$$EX = \frac{1}{p}$$

*Proof.* Suppose that  $Y \simeq \text{Geom}(p)$ , i.e.,  $Y$  has probability function  $p(y) = q^{y-1}p$ , where  $y = 1, 2, \dots$ . Then

$$\begin{aligned} E(Y) &= \sum_y yp(y) \\ &= \sum_{y=1}^{\infty} yq^{y-1}p = p \sum_{y=1}^{\infty} yq^{y-1} \\ &= \end{aligned}$$

□

## 5.7 Problems

## 6 Continuous Random Variables

### 6.1 Continuous Random Variables

**Definition.** A random variable  $X$  is **continuous** if there exists a non-negative function  $f$  so that, for every interval  $B$ ,

$$P(X \in B) = \int_B f(x) dx.$$

The function  $f = f_X$  is called the **density of  $X$**  (p. d. f.).

We will assume that a density function  $f$  is continuous, apart from finitely many (possibly infinite) jumps. Clearly, it must hold that

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Note that

$$P(X \in [a, b]) = P(a \leq X \leq b) = \int_a^b f(x) dx,$$

$$P(X = a) = \int_a^a f(x) dx = 0,$$

$$P(X \leq b) = P(X < b) = \int_{-\infty}^b f(x) dx.$$

**Definition.** The function  $F = F_X$  given by

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt$$

is called the **distribution function of  $X$** . On an open interval where  $f$  is continuous,

$$F'_X(x) = f_X(x).$$

Density has the same role as the probability mass function for discrete random variables: it tells which values  $x$  are relatively more probable for  $X$  than others. Namely, if  $h$  is very small, then

$$P(X \in [x, x + h]) = F(x + h) - F(x) \approx F'(x) \cdot h = f(x) \cdot h$$

by *Mean Value Theorem*. By analogy with discrete random variables, we define

$$EX = \int_{-\infty}^{\infty} x \cdot f(x) dx,$$

$$Eg(X) = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx,$$

and variance is computed by the same formula:  $\text{Var}(X) = [E(X - EX)]^2 = E(X^2) - (EX)^2$ .

**Example 6.1.** Let

$$f(x) = \begin{cases} cx & \text{if } 0 < x < 4, \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine  $c$ . (b) Compute  $P(1 \leq X \leq 2)$ . (c) Determine  $EX$  and  $\text{Var}(X)$ .

**Sol.** (a) We must find  $c$  such that  $\int_0^4 cx \, dx = 1$ . Then

$$\int_0^4 cx \, dx = c \frac{1}{2} x^2 \Big|_0^4 = 8c = 1.$$

Thus,  $c = 1/8$ . Note that

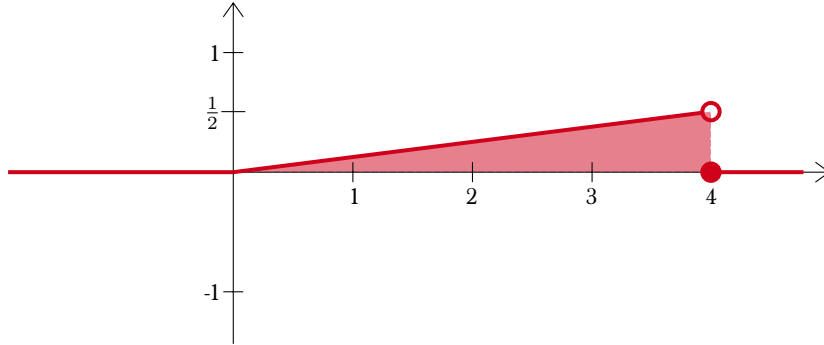


Figure 5

$$(b) \int_1^2 \frac{1}{8} x \, dx = \frac{1}{8} \frac{1}{2} x^2 \Big|_1^2 = \frac{3}{16}.$$

(c) Since

$$EX = \int_0^4 \frac{1}{8} x^2 \, dx = \frac{1}{8} \frac{1}{3} x^3 \Big|_0^4 = \frac{8}{3}$$

and

$$E(X^2) = \int_0^4 \frac{1}{8} x^3 \, dx = \frac{1}{8} \frac{1}{4} x^4 \Big|_0^4 = 8,$$

$$\text{we have } \text{Var}(X) = 8 - \left(\frac{8}{3}\right)^2 = \frac{8}{9}.$$

□

**Example 6.2.** Assume that  $X$  has density

$$f_X(x) = \begin{cases} 3x^2 & \text{if } x \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Compute the density  $f_Y$  of  $Y = 1 - X^4$ .

**Sol.** Note that the density  $f_Y(y)$  will be non-zero only when  $y \in [0, 1]$ , as the values of  $Y$  are restricted to that interval. Now, for  $y \in (0, 1)$ ,

$$F_Y(y) = P(Y \leq y) = P(1 - X^4 \leq y) = P(1 - y \leq X^4) = P[(1 - y)^{1/4} \leq X] = \int_{(1-y)^{1/4}}^1 3x^2 \, dx.$$

It follows that

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \left[ \int_{(1-y)^{1/4}}^1 3x^2 \, dx \right] = \frac{d}{dy} \left[ 1 - (1 - y)^{3/4} \right] = -\frac{3}{4} (1 - y)^{-1/4} (-1) = \frac{3}{4} \frac{1}{(1 - y)^{1/4}}$$

for  $y \in (0, 1)$ , and  $f_Y(y) = 0$  otherwise. Observe that it is immaterial how  $f(y)$  is defined at  $y = 0$  and  $y = 1$ , because those two values contribute nothing to any integral. □

## 6.2 Uniform Random Variable

Uniform random variable represents the choice of a **random number** in  $[\alpha, \beta]$ . For  $[\alpha, \beta] = [0, 1]$ , this is ideally the output of a *computer random number generator*.

**Definition.** A random variable is uniform if it has the probability density function given below.

$$\text{Density : } f(x) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } x \in [\alpha, \beta], \\ 0 & \text{otherwise.} \end{cases}$$

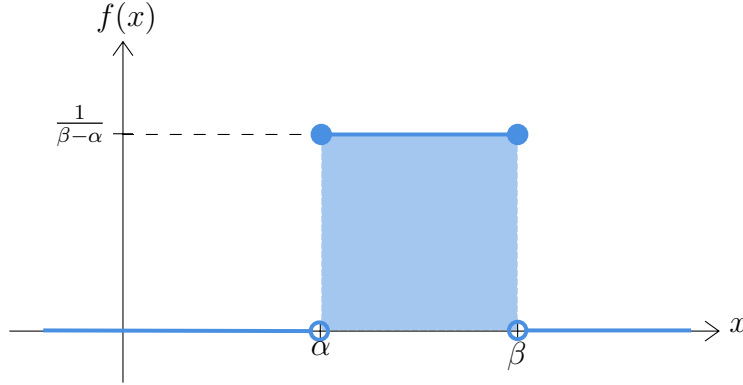


Figure 6

### Properties of Uniform Random Variable:

1.  $EX = \frac{\alpha + \beta}{2}$ .
2.  $\text{Var}(X) = \frac{(\beta - \alpha)^2}{12}$ .

*Proof.* 1.

$$EX = \int_{\alpha}^{\beta} x \cdot \frac{1}{\beta - \alpha} dx = \left( \frac{1}{\beta - \alpha} \right) \frac{1}{2} x^2 \Big|_{\alpha}^{\beta} = \frac{\beta^2 - \alpha^2}{2(\beta - \alpha)} = \frac{\alpha + \beta}{2}.$$

2. Note that

$$EX^2 = \int_{\alpha}^{\beta} x^2 \cdot \frac{1}{\beta - \alpha} dx = \left( \frac{1}{\beta - \alpha} \right) \frac{1}{3} x^3 \Big|_{\alpha}^{\beta} = \frac{\beta^3 - \alpha^3}{3(\beta - \alpha)} = \frac{\alpha^2 + \beta^2 + \alpha\beta}{3}.$$

Thus,

$$\begin{aligned} \text{Var}(X) &= EX^2 - (EX)^2 = \frac{\alpha^2 + \beta^2 + \alpha\beta}{3} - \left( \frac{\alpha + \beta}{2} \right)^2 \\ &= \frac{\alpha^2 + \beta^2 + \alpha\beta}{3} - \frac{\alpha^2 + \beta^2 + 2\alpha\beta}{4} \\ &= \frac{4(\alpha^2 + \beta^2 + \alpha\beta) - 3(\alpha^2 + \beta^2 + 2\alpha\beta)}{12} \\ &= \frac{\alpha^2 + \beta^2 - 2\alpha\beta}{12} = \frac{(\beta - \alpha)^2}{12}. \end{aligned}$$

□

**Example 6.3.** Assume that  $X$  is uniform on  $[0, 1]$ . (a) What is  $P(X \in \mathbb{Q})$ ? (b) What is the probability that the binary expansion of  $X$  starts with 0.010?

**Sol.** (a) As  $\mathbb{Q}$  is countable, it has an enumeration, say,  $\{q_1, q_2, \dots\}$ . By Axiom 3 of Probability,

$$P(X \in \mathbb{Q}) = P\left(\bigcup_{i \in I} \{X = q_i\}\right) = \sum_{i \in I} P(X = q_i) = 0.$$

Note that  $P(X \in \mathbb{R}) = P(\Omega) = 1$ .

(b) As  $X$  is, with probability 1, irrational, its binary expansion is uniquely defined, so there is no ambiguity about what the second question means.

Divide  $[0, 1)$  into  $2^n$  intervals of equal length. If the binary expansion of a number  $x \in [0, 1)$  is  $0.x_1x_2\dots$ , the first  $n$  binary digits determine which of the  $2^n$  subintervals  $x$  belong to: if you know that  $x$  belongs to an interval  $I$  based on the first  $n-1$  digits, then  $n$ -th digit 1 means that  $x$  is in the right half of  $I$  and  $n$ -th digit 0 means that  $x$  is in the left half of  $I$ . Thus, if the expansion starts with 0.010, then number is in  $[0, \frac{1}{2}]$ , then is  $[\frac{1}{4}, \frac{1}{2}]$ , and then finally in  $[\frac{1}{4}, \frac{3}{8}]$ . Our answer is  $1/8$ .  $\square$

**Example 6.4.** A uniform random number  $X$  divides  $[0, 1]$  into two segments. Let  $R$  be the ratio of the smaller versus the large segment. Compute the density of  $R$ .

**Sol.** As  $R$  has value in  $(0, 1)$ , the density  $f_R(r)$  is non-zero only for  $r \in (0, 1)$  and we will deal only with such  $r$ 's.

$$F_R(r) = P(R \leq r) = P\left(X \leq \frac{1}{2}, \frac{X}{1-X} \leq r\right) + P\left(X > \frac{1}{2}, \frac{1-X}{X} \leq r\right).$$

Note that

$$\begin{aligned} \frac{X}{1-X} \leq r &\implies \frac{1-X}{X} \geq \frac{1}{r} \implies \frac{1}{X} - 1 \geq \frac{1}{r} \implies \frac{1}{X} \geq \frac{r+1}{r} \implies X \leq \frac{r}{r+1}, \\ \frac{1-X}{X} \leq r &\implies \frac{1}{X} - 1 \leq r \implies \frac{1}{X} \leq r+1 \implies X \geq \frac{1}{r+1}. \end{aligned}$$

Thus,

$$\begin{aligned} F_R(r) &= P(R \leq r) = P\left(X \leq \frac{1}{2}, X \leq \frac{r}{r+1}\right) + P\left(X > \frac{1}{2}, X \geq \frac{1}{r+1}\right) \\ &= P\left(X \leq \frac{r}{r+1}\right) + P\left(X \geq \frac{1}{r+1}\right) \quad \left(\because \frac{r}{r+1} \leq \frac{1}{2} \text{ and } \frac{1}{r+1} \geq \frac{1}{2}\right) \\ &= \frac{r}{r+1} + 1 - \frac{1}{r+1} = \frac{2r}{r+1}. \end{aligned}$$

For  $r \in (0, 1)$ , the density, thus, equals  $f_R(r) = \frac{d}{dr}F_R(r) = \frac{2}{(r+1)^2}$ . Hence

$$f_R(r) = \begin{cases} \frac{2}{(r+1)^2} & \text{if } r \in (0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

$\square$



## 6.3 Exponential Random Variable

**Definition.** A random variable is Exponential( $\lambda$ ), with parameter  $\lambda > 0$ , if it has the probability density function given below.

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

This is a distribution for the **waiting time** for some random event, for example, for a lightbulb to burn out or for the next earthquake of a t lest some given magnitude.

### Properties of Exponential Random Variable:

1.  $EX = \frac{1}{\lambda}$ .
2.  $\text{Var}(X) = \frac{1}{\lambda^2}$ .
3.  $P(X \geq x) = e^{-\lambda x}$ .
4. Memoryless Property:  $P(X \geq x + y | X \geq y) = e^{-\lambda x}$ .

The last property means that, if the event has not occurred by some given time (no matter large), the distribution of the remaining waiting time is the same as it was at the beginning. There is no “aging”.

*Proof.* Note that

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}.$$

1.

$$\begin{aligned} EX &= \int_0^\infty x \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \int_0^\infty y e^{-y} dy \quad \text{where } y = \lambda x \\ &= \frac{1}{\lambda} [-e^{-y} - y e^{-y}]_0^\infty = \frac{1}{\lambda}. \end{aligned}$$

2. Since

$$\begin{aligned} EX^2 &= \int_0^\infty x^2 \lambda e^{-\lambda x} dx \\ &= \int_0^\infty x \cdot x \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda^2} \int_0^\infty y^2 e^{-y} dy \quad \text{where } y = \lambda x \\ &= \frac{1}{\lambda^2} [-2e^{-y} - 2y e^{-y} - y^2 e^{-y}]_0^\infty \\ &= \frac{2}{\lambda^2}, \end{aligned}$$

$$\text{we obtain that } \text{Var}(X) = EX^2 - (EX)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.$$

3.  $P(X \geq x) = 1 - P(X \leq x) = e^{-\lambda x}.$

4.

$$\begin{aligned}
 P(X \geq x+y|X \geq y) &= \frac{P(X > x+y, X > y)}{P(X > y)} \\
 &= \frac{P(X > x+y)}{P(X > y)} \\
 &= \frac{1 - F_X(x+y)}{1 - F_X(y)} \\
 &= \frac{1 - (1 - e^{-\lambda(x+y)})}{1 - (1 - e^{-\lambda y})} \\
 &= \frac{e^{-\lambda(x+y)}}{e^{-\lambda y}} \\
 &= \exp[-\lambda(x+y) - (-\lambda y)] \\
 &= e^{-\lambda x} (= P(X > x)).
 \end{aligned}$$

□

**Example 6.5.** Assume that a lightbulb lasts on average 100 hours. Assuming exponential distribution, compute the probability that it lasts more than 200 hours and the probability that it lasts less than 50 hours.

**Sol.** Let  $X$  be the waiting time for the bulb to burn out. Then,  $X$  is Exponential with  $\lambda = 1/100$  ( $\because EX = 1/\lambda = 100$ ). And then

$$F_X(x) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = -e^{-\lambda t} \Big|_0^x = 1 - e^{-\lambda x},$$

so  $F_X(x) \Big|_{\lambda=\frac{1}{100}} = 1 - \exp\left(-\frac{x}{100}\right)$ . Thus,

$$P(X \geq 200) = 1 - F_X(200) = e^{-2} \approx 0.1353,$$

$$P(X \leq 50) = F_X(50) = 1 - e^{-1/2} \approx 0.3935.$$

□

## 6.4 Normal Random Variable

**Definition.** A random variable is **Normal with parameter**  $\mu \in \mathbb{R}$  **and**  $\sigma^2 > 0$  or, in short,  $X$  is  $N(\mu, \sigma^2)$ , if its density is the function given below.

$$\text{Density : } f(x) = f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right],$$

where  $x \in (-\infty, \infty)$ .

### Properties of Normal Random Variable:

1. For  $f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right]$ ,  $\int_{-\infty}^{\infty} f(x) dx = 1$ .
2.  $EX = \mu$ .
3.  $\text{Var}(X) = \sigma^2$ .

*Proof.*

1. Let  $\frac{x - \mu}{\sigma} = t$ , i.e.,  $\frac{1}{\sigma} dx = dt$ . Then

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp \left[ -\frac{(x - \mu)^2}{2\sigma^2} \right] dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left( -\frac{t^2}{2} \right) dt.$$

Let  $S = \int_{-\infty}^{\infty} e^{-ax^2} dx = \int_{-\infty}^{\infty} e^{-ay^2} dy$ . Then

$$\begin{aligned} S^2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-a(x^2+y^2)} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-ar^2} r dr d\theta \\ &= \int_0^{2\pi} -\frac{1}{2a} \left[ e^{-r^2} \right]_0^{\infty} d\theta \\ &= \left[ \frac{1}{2a} \theta \right]_0^{2\pi} = \frac{\pi}{a}. \end{aligned}$$

Thus,  $S = \sqrt{\frac{\pi}{a}}$ . Hence

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left( -\frac{t^2}{2} \right) dt = \frac{1}{\sqrt{2\pi}} \sqrt{2\pi} = 1.$$

2.

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{\infty} x \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \int_{-\infty}^{\infty} [(x-\mu) + \mu] \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \int_{-\infty}^{\infty} (x-\mu) \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx + \int_{-\infty}^{\infty} \mu \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} (-\sigma^2) \int_{-\infty}^{\infty} \frac{-x+\mu}{\sigma^2} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx + \mu \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx \\
&= \frac{1}{\sigma\sqrt{2\pi}} (-\sigma^2) \left[ \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \right]_{-\infty}^{\infty} + \mu \\
&= \mu.
\end{aligned}$$

3.

$$\text{Var}(X) = E[(X-E[X])^2] = \int_{-\infty}^{\infty} (x-\mu)^2 \cdot f(x) dx = \int_{-\infty}^{\infty} (x-\mu)^2 \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx.$$

Let  $\omega = \frac{x-\mu}{\sigma\sqrt{2}}$  (here,  $d\omega = \frac{1}{\sigma\sqrt{2}} dx$ ), then

$$\text{Var}(X) = \int_{-\infty}^{\infty} \sigma^2 2\omega^2 \frac{1}{\sigma\sqrt{2\pi}} e^{-\omega^2} \sigma\sqrt{2} d\omega = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} \omega \cdot \omega e^{-\omega^2} d\omega.$$

Using integration by parts:

Differentiation	Integration
$\omega$	$\omega e^{-\omega^2}$
1	$-1/2 e^{-\omega^2}$

we have

$$\begin{aligned}
\text{Var}(X) &= \frac{2\sigma^2}{\sqrt{\pi}} \left( \cancel{\omega \frac{-1}{2} e^{-\omega^2}} \Big|_{-\infty}^{\infty} \overset{\text{0 by L'Hôpital's rule}}{\rightarrow} - \int_{-\infty}^{\infty} \frac{-1}{2} e^{-\omega^2} d\omega \right) \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \left( \frac{1}{2} \int_{-\infty}^{\infty} e^{-\omega^2} d\omega \right) \\
&= \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{1}{2} \cdot \sqrt{\pi} = \sigma^2.
\end{aligned}$$

□

If  $X \sim N(\mu_X, \sigma_X^2)$  and  $Y = aX + b$  where  $a, b \in \mathbb{R}$ , then

$$Y \sim N(\mu_Y, \sigma_Y^2) \quad \text{where} \quad \mu_Y = a\mu_X + b \quad \text{and} \quad \sigma_Y^2 = a^2\sigma_X^2.$$

*Proof.*

- $EY = E(aX + b) = aEX + b = a\mu_X + b.$
- $\text{Var}(Y) = \text{Var}(aX + b) = a^2\text{Var}(X) = a^2\sigma_X^2.$

□

**Example 6.6.** Let  $X$  be a  $N(\mu, \sigma^2)$  random variable and let  $Y = \alpha X + \beta$ , with  $\alpha > 0$ . How is  $Y$  distributed?

**Sol.** From

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\alpha X + \beta \leq y) \\ &= P\left(X \leq \frac{y - \beta}{\alpha}\right) \\ &= \int_{-\infty}^{\frac{y - \beta}{\alpha}} f_X(x) dx, \end{aligned}$$

we have

$$\begin{aligned} \frac{d}{dt} F_Y(y) &= f_Y(y) = f_X\left(\frac{y - \beta}{\alpha}\right) \cdot \frac{1}{\alpha} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{\alpha} \exp\left[-\frac{(y - \beta - \alpha\mu)^2}{2\alpha^2\sigma^2}\right] = \frac{1}{\sqrt{2\pi}(\alpha\sigma)} \exp\left[-\frac{(y - (\alpha\mu + \beta))^2}{2(\alpha\sigma)^2}\right]. \end{aligned}$$

Therefore,  $Y$  is normal with  $EY = \alpha\mu + \beta$  and  $\text{Var}(Y) = (\alpha\sigma)^2$ . □

**Note.** In particular,

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X + \left(-\frac{\mu}{\sigma}\right)$$

has  $EZ = 0$  and  $\text{Var}(Z) = 1$ . Such a  $N(0, 1)$  random variable is called *standard* Normal. Its distribution function  $F_Z(z)$  is denoted by  $\Phi(z)$ . Note that

$$\begin{aligned} \Phi(z) &= F_Z(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z e^{-x^2/2} dx, \\ f_Z(z) &= \frac{1}{\sqrt{2\pi}} e^{-z^2/2}. \end{aligned}$$

You should also note that it is enough to know these values for  $z > 0$ , as in this case, by using the fact that  $f_Z(x)$  is an even function:

$$\Phi(-z) = \int_{-\infty}^{-z} f_Z(x) dx = \int_z^{\infty} f_Z(x) dx = 1 - \int_{-\infty}^z f_Z(x) dx = 1 - \Phi(z).$$

In particular,  $\Phi(0) = 1/2$  because  $\Phi(0) = 1 - \Phi(0)$ . Another way to write this is  $P(Z \geq -z) = P(Z \leq z)$  for all  $z \in \mathbb{R}$ .

**Example 6.7.** What is the probability that a Normal random variable differs from its mean  $\mu$  by more than  $\sigma$ ? More than  $2\sigma$ ? More than  $3\sigma$ ?

**Sol.** For  $X \sim N(\mu, \sigma^2)$ , we need to compute

$$\begin{aligned} P(|X - \mu| \geq \sigma), \\ P(|X - \mu| \geq 2\sigma), \\ P(|X - \mu| \geq 3\sigma). \end{aligned}$$

We have

$$P(|X - \mu| \geq \sigma) = P\left(\left|\frac{X - \mu}{\sigma}\right| \geq 1\right) = P(|Z| \geq 1) = 2P(Z \geq 1) = 2(1 - \Phi(1)) \approx 0.3713.$$

Similarly,

$$\begin{aligned} P(|X - \mu| \geq 2\sigma) &= 2(1 - \Phi(2)) \approx 0.0455, \\ P(|X - \mu| \geq 3\sigma) &= 2(1 - \Phi(3)) \approx 0.0027. \end{aligned}$$

□

**Example 6.8.** Assume that  $X$  is Normal with mean  $\mu = 2$  and variance  $\sigma^2 = 25$ . Compute the probability that  $X$  is between 1 and 4.

**Sol.**

$$\begin{aligned} P(1 \leq X \leq 4) &= P\left(\frac{1-2}{5} \leq \frac{X-2}{5} \leq \frac{4-2}{5}\right) = P(-0.2 \leq Z \leq 0.4) \\ &= P(Z \leq 0.4) - P(Z \leq -0.2) \\ &= \Phi(0.4) - (1 - \Phi(0.2)) \approx 0.2347. \end{aligned}$$

□

**Theorem 6.1 (De Moivre-Laplace Central Limit Theorem).** Let  $S_n$  be Binomial( $n, p$ ), where  $p$  is fixed and  $n$  is large. Then,  $\frac{S_n - np}{\sqrt{np(1-p)}} \approx N(0, 1)$ ; more precisely,

$$P\left(\frac{S_n - np}{\sqrt{np(1-p)}} \leq x\right) \rightarrow \Phi(x)$$

as  $n \rightarrow \infty$ , for every real number  $x$ .

We should also note that the above theorem is an analytical statement; it says that

$$P\left(S_n \leq np + x\sqrt{np(1-p)}\right) = \sum_{0 \leq k \leq np + x\sqrt{np(1-p)}} \binom{n}{k} p^k (1-p)^{n-k} \rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{t^2/2} dt$$

as  $n \rightarrow \infty$ , for every  $x \in \mathbb{R}$ .

**Remark.** Remember that, when  $n$  is large and  $p$  is small, say  $n = 100$  and  $p = 1/100$ , the Poisson approximation (with  $\lambda = np$ ) is much better!

**Example 6.9.** A roulette wheel has 38 slots: 18 red, 18 black, and 2 green. The ball ends at one of these at random. You are a player who plays a large number of games and makes an even bet of \$1 on red in every game. After  $n$  games, what is the probability that you are ahead? Answer this for  $n = 100$  and  $n = 1000$ .

**Sol.** Let  $S_n$  be the number of times you win. This is a Binomial  $(n, \frac{9}{19})$  random variable.

$$\begin{aligned} P(\text{ahead}) &= P(\text{win more than half of the games}) \\ &= P\left(S_n > \frac{n}{2}\right) \\ &= P\left(\frac{S_n - np}{\sqrt{np(1-p)}} > \frac{1/2n - np}{\sqrt{np(1-p)}}\right) \\ &\approx P\left(Z > \frac{(1/2 - p)\sqrt{n}}{\sqrt{p(1-p)}}\right). \end{aligned}$$

For  $n = 100$ , we get

$$P(Z > 5/\sqrt{90}) \approx 0.2990,$$

and for  $n = 1000$ , we get

$$P(Z > 5/3) \approx 0.0478.$$

□

**Example 6.10.** What would the answer to the previous example be if the game were fair, i.e., you bet even money on the outcome of a fair coin toss each time.

**Sol.** Then  $p = 1/2$  and

$$P(\text{ahead}) \rightarrow P(Z > 0) = 0.5,$$

as  $n \rightarrow \infty$ .

□

**Example 6.11.** How many times do you need to toss a fair coin to get 100 heads with probability 0.9.

**Sol.** Let  $n$  be the number of tosses that we are looking for. For  $S_n$ , which is Binomial( $n, 1/2$ ), we need to find  $n$  so that

$$P(S_n \geq 100) \approx 0.9.$$

We will use below that  $n > 200$ , as the probability would be approximately 1/2 for  $n = 200$  (see the previous example). Here is the computation:

$$\begin{aligned} P\left(\frac{S_n - 1/2n}{1/2\sqrt{n}} \geq \frac{100 - 1/2n}{1/2\sqrt{n}}\right) &\approx P\left(Z \geq \frac{100 - 1/2n}{1/2\sqrt{n}}\right) = P\left(Z \geq \frac{200 - n}{\sqrt{n}}\right) \\ &= P\left(Z \geq -\left(\frac{n - 200}{\sqrt{n}}\right)\right) \\ &= P\left(Z \leq \frac{n - 200}{\sqrt{n}}\right) \\ &= \Phi\left(\frac{n - 200}{\sqrt{n}}\right) \\ &= 0.9. \end{aligned}$$

According to the tables,  $\Phi(1.28) \approx 0.9$ , thus we need to solve  $\frac{n-200}{\sqrt{n}} = 1.28$ , that is,

$$n - 1.28\sqrt{n} - 200 = 0.$$

This is a quadratic equation in  $\sqrt{n}$ , with the only positive solution

$$\sqrt{n} = \frac{1.28 + \sqrt{1.28^2 + 800}}{2}.$$

Thus,  $n = 219$ .

□



## 6.5 Exercises for Chapter 6

1. A random variable  $X$  has the density function

$$f(x) = \begin{cases} c(x + \sqrt{x}) & x \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

(a) Determine  $c$ . (b) Compute  $E(1/X)$ . (c) Determine the p. d. f. of  $Y = X^2$ .

**Sol.** (a) We must find  $c$  such that  $\int_0^1 c(x + \sqrt{x}) dx = 1$ . Then

$$\int_0^1 c(x + \sqrt{x}) dx = c \left[ \frac{1}{2}x^2 + \frac{2}{3}x^{3/2} \right]_0^1 = c \left( \frac{1}{2} + \frac{2}{3} \right) = c \cdot \frac{7}{6} = 1.$$

Thus,  $c = 6/7$ .

(b)

$$\begin{aligned} E(1/X) &= \int_0^1 \frac{1}{x} \cdot f_X(x) dx = \int_0^1 \frac{1}{x} \cdot \frac{6}{7} (x + \sqrt{x}) dx \\ &= \frac{6}{7} \int_0^1 (1 + x^{-1/2}) dx \\ &= \frac{6}{7} \left[ x + 2x^{1/2} \right]_0^1 \\ &= \frac{6}{7} (1 + 2) = \frac{18}{7}. \end{aligned}$$

(c) For  $y \in (0, 1)$ ,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(X^2 \leq y) = P(X \leq \sqrt{y}) = \int_0^{\sqrt{y}} \frac{6}{7} (x + \sqrt{x}) dx \\ &= \frac{6}{7} \left[ \frac{1}{2}x^2 + \frac{2}{3}x^{3/2} \right]_0^{\sqrt{y}} \\ &= \frac{6}{7} \left( \frac{1}{2}y + \frac{2}{3}y^{3/4} \right). \end{aligned}$$

It follows that

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} \left[ \frac{6}{7} \left( \frac{1}{2}y + \frac{2}{3}y^{3/4} \right) \right] = \frac{6}{7} \left( \frac{1}{2} + \frac{2 \cdot 3}{3 \cdot 4} y^{-1/4} \right) = \frac{6}{7} \left( \frac{1}{2} + \frac{1}{2} y^{-1/4} \right)$$

for  $y \in (0, 1)$ , and  $f_Y(y) = 0$  otherwise; that is,

$$f_Y(y) = \begin{cases} \frac{3}{7} \left( 1 + y^{-1/4} \right) & y \in (0, 1), \\ 0 & \text{otherwise.} \end{cases}$$

□

2. The density function of a random variable  $X$  is given by

$$f(x) = \begin{cases} a + bx & x \in [0, 2], \\ 0 & \text{otherwise.} \end{cases}$$

We also know that  $E(X) = 7/6$ . (a) Compute  $a$  and  $b$ . (b) Compute  $\text{Var}(X)$ .

**Sol.** (a) Note that

$$\int_0^2 (a + bx) dx = \left[ ax + \frac{b}{2}x^2 \right]_0^2 = 2a + 2b = 1 \quad \dots \quad (1)$$

and that

$$EX = \int_0^2 x(a + bx) dx = \int_0^2 (ax + bx^2) dx = \left[ \frac{a}{2}x^2 + \frac{b}{3}x^3 \right]_0^2 = 2a + \frac{8}{3}b = \frac{7}{6}. \quad \dots \quad (2)$$

Doing (2) – (1), we have

$$a = b = \frac{1}{4}.$$

(b) Remind that  $f_X(x) = \frac{1}{4}(1 + x)$  for  $x \in [0, 2]$  and that  $EX = 7/6$ . Since

$$EX^2 = \int_0^2 x^2 \left[ \frac{1}{4}(1 + x) \right] dx = \frac{1}{4} \int_0^2 (x^2 + x^3) dx = \frac{1}{4} \left[ \frac{1}{3}x^3 + \frac{1}{4}x^4 \right]_0^2 = \frac{1}{4} \left( \frac{8}{3} + 4 \right) = \frac{2}{3} + 1 = \frac{5}{3},$$

we obtain

$$\text{Var}(X) = EX^2 - (EX)^2 = \frac{5}{3} - \left( \frac{7}{6} \right)^2 = \frac{5}{3} - \frac{49}{36} = \frac{60 - 49}{36} = \frac{11}{36}.$$

□

3. After your complaint about their service, a representative of an insurance company promised to call you “between 7 and 9 this evening. Assume that this means that the time  $T$  of the call is uniformly distributed in the specified interval.

- (a) Compute the probability that the call arrives between 8:00 and 8:20.
- (b) At 8:30, the call still hasn’t arrived. What is the probability that it arrives in the next 10 minutes?
- (c) Assume that you know in advance that the call will last *exactly* 1 hour. From 9 to 9:30, there is a game show on TV that you wanted to watch. Let  $M$  be the amount of time of the show that you miss because of the call. Compute the expected value of  $M$ .

**Sol.** (a) Remind that the time  $T$  is uniformly distributed in the interval 120 minutes (7 p.m.—9 p.m.); that is,  $T$  is uniform on  $[0, 120]$ . Thus,

$$P(\text{the call arrives between 8:00 and 8:20}) = \frac{1}{120} \cdot 20 = \frac{1}{6} \approx 0.167.$$

(b)

$$P(T \leq 100 | T \geq 90) = \frac{P(T \leq 100, T \geq 90)}{P(T \geq 90)} = \frac{10/120}{30/120} = \frac{1}{3}.$$

(c) Note that

$$M = \begin{cases} 0 & \text{if } T \in [0, 60], \\ T - 60 & \text{if } T \in [60, 90], \\ 30 & \text{if } T \in [90, 120]. \end{cases}$$

Since the density of  $T$  is  $1/120$  for  $T \in [0, 120]$ ,

$$\begin{aligned} EM &= \int_{60}^{90} (t - 60) \frac{1}{120} dt + \int_{90}^{120} 30 \cdot \frac{1}{120} dt \\ &= \frac{1}{120} \left[ \frac{1}{2} t^2 - 60t \right]_{60}^{90} + 30t \Big|_{90}^{120} \\ &= \frac{30}{120} \left[ \frac{1}{60} t^2 - 2t \right]_{60}^{90} + t \Big|_{90}^{120} \\ &= \frac{1}{4} \left( \frac{8100}{60} - 180 - (60 - 120) + 30 \right) \\ &= \frac{1}{4} \left( \frac{8100}{60} - 90 \right) \end{aligned}$$

□

4. Toss a fair coin twice. You win \$1 if *at least one* of the two tosses comes out heads.
- (a) Assume that you play this game 300 times. What is, approximately, the probability that you win at least \$250?
- (b) Approximately how many times do you need to play so that you win at least \$250 with probability at least 0.99?
5. Roll a dice  $n$  times and let  $M$  be the number of times you roll 6. Assume that  $n$  is large.
- (a) Compute the expectation  $EM$ .
- (b) Write down an approximation, in terms on  $n$  and  $\Phi$ , of the probability that  $M$  differs from its expectation by less than 10%.
- (c) How large should  $n$  be so that the probability in (b) is larger than 0.99?

## 7 Joint Distributions and Independence

### 7.1 Discrete Case

Assume that you have a pair  $(X, Y)$  of discrete random variable  $X$  and  $Y$ . Their *joint probability mass function* is given by

$$p(x, y) = P(X = x, Y = y)$$

so that

$$P((X, Y) \in A) = \sum_{(X, Y) \in A} p(x, y) = \sum_x \sum_y p(x, y).$$

The *marginal probability mass function* are the p. m. f.'s of  $X$  and  $Y$ , given by

$$P(X = x) = \sum_y P(X = x, Y = y) = \sum_y p(x, y),$$

$$P(Y = y) = \sum_x P(X = x, Y = y) = \sum_x p(x, y).$$

**Example 7.1.** An urn has 2 red, 5 white, and 3 green balls. Select 3 balls at random and let  $X$  be the number of red balls and  $Y$  be the number of white balls. Determine (a) joint p.m.f. of  $(X, Y)$ , (b) marginal p.m.f.'s, (c)  $P(X \geq Y)$ , and (d)  $P(X = 2 | X \geq Y)$ .

**Sol.** The joint p.m.f. is given by  $P(X = x, Y = y)$  for all possible  $x$  and  $y$ . Then

$y \backslash x$	0	1	2	$P(Y = y)$
0	1/120	2 · 3/120	3/120	10/120
1	5 · 3/120	2 · 5 · 3/120	5/120	50/120
2	10 · 3/120	10 · 2/120	0	50/120
3	10/120	0	0	10/120
$P(X = x)$	56/120	56/120	8/120	1

Table 1

The last row and column entries are the respective column and row sums and, therefore, determine the marginal p.m.f.'s. To answer (c) we merely add the relevant probabilities,

$$P(X \geq Y) = \frac{1 + 6 + 3 + 30 + 5}{120} = \frac{3}{8},$$

and, to answer (d), we compute

$$\frac{P(X = 2, X \geq Y)}{P(X \geq Y)} = \frac{8/120}{3/8} = \frac{8}{45}.$$

□

**Definition.** Two random variables  $X$  and  $Y$  are **independent** if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all intervals  $A$  and  $B$ . In the discrete case,  $X$  and  $Y$  are independent exactly when

$$P(X = x, Y = y) = P(X = x)P(Y = y)$$

for all possible values  $x$  and  $y$  of  $X$  and  $Y$ , that is, the joint p.m.f. is the product of the marginal p.m.f.'s.

**Example 7.2.** In the previous example, are  $X$  and  $Y$  independent?

**Sol.** No, the 0's in the table are dead giveaways. For example,  $P(X = 2, Y = 2) = 0$ , but neither  $P(X = 2)$  nor  $P(Y = 2)$  is 0.  $\square$

**Example 7.3.** Most often, independence is an assumption. For example, roll a die twice and let  $X$  be the number on the first roll and let  $Y$  be the number on the second roll. Then,  $X$  and  $Y$  are independent.

## 7.2 Continuous Case

We say that  $(X, Y)$  is a *joint density*  $f(x, y) \geq 0$  so that

$$P((X, Y) \in S) = \iint_S f(x, y) \, dx \, dy,$$

where  $S$  is some nice (say, open or closed) subset of  $\mathbb{R}^2$ .

**Example 7.4.** Let  $(X, Y)$  be a random point in  $S$ , where  $S$  is compact (that is, closed and bounded) subset of  $\mathbb{R}^2$ . This means that

$$f(x, y) = \begin{cases} \frac{1}{\text{Area}(S)} & \text{if } (x, y) \in S, \\ 0 & \text{otherwise.} \end{cases}$$

This simplest example is a square of side length 1 where

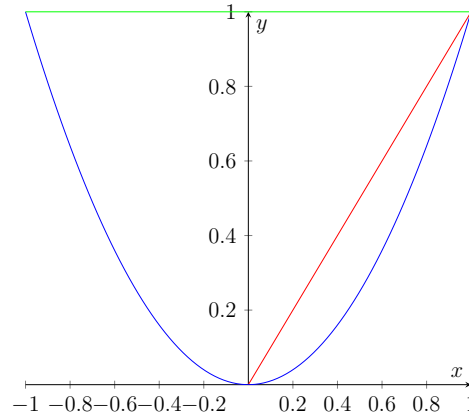
$$f(x, y) = \begin{cases} 1 & \text{if } x \in [0, 1] \text{ and } y \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

**Example 7.5.** Let

$$f(x, y) = \begin{cases} cx^2y & \text{if } x^2 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Determine (a) the constant  $c$ , (b)  $P(X \geq Y)$ , (c)  $P(X = Y)$ , and (d)  $P(X = 2Y)$ .

**Sol.** Note that



For (a),

$$\begin{aligned} \int_{-1}^1 \int_{x^2}^1 cx^2y \, dy \, dx &= \int_{-1}^1 cx^2 \left[ \frac{1}{2}y^2 \right]_{y=x^2}^{y=1} dx \\ &= \int_{-1}^1 \frac{cx^2}{2} (1 - x^4) \, dx \\ &= \frac{c}{2} \int_{-1}^1 (x^2 - x^6) \, dx \\ &= \frac{c}{2} \left[ \frac{1}{3}x^3 - \frac{1}{7}x^7 \right]_{-1}^1 \\ &= \frac{c}{2} \left( \frac{1}{3} - \frac{1}{7} + \left( \frac{1}{3} - \frac{1}{7} \right) \right) = c \cdot \frac{4}{21} = 1, \end{aligned}$$

and so  $c = \frac{21}{4}$ . For (b), let  $S$  be the region between the graphs  $y = x^2$  and  $y = x$ , for  $x \in (0, 1)$ . Then

$$\begin{aligned} P(X \geq Y) &= P((X, Y) \in S) = \int_0^1 \int_{x^2}^x \frac{21}{4} \cdot x^2y \, dy \, dx \\ &= \frac{21}{4} \int_0^1 x^2 \left[ \frac{1}{2}y^2 \right]_{x^2}^x dx \\ &= \frac{21}{8} \int_0^1 x^2(x^2 - x^4) \, dx \\ &= \frac{21}{8} \left[ \frac{1}{5}x^5 - \frac{1}{7}x^7 \right]_0^1 = \frac{21}{8} \cdot \frac{2}{35} = \frac{3}{20}. \end{aligned}$$

Both probabilities in (c) and (d) are 0 because a two-dimensional integral over a line is 0.  $\square$

If  $f$  is the joint density of  $(X, Y)$ , then the two *marginal densities*, which are the densities of  $X$  and  $Y$ , are computed by integrating out the other variable:

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

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y) dx.$$

Indeed, for an interval  $A$ ,  $X \in A$  means that  $(X, Y) \in S$ , where  $S = A \times \mathbb{R}$ , and, therefore,

$$P(X \in A) = \int_A \int_{-\infty}^{\infty} f(x, y) dy dx.$$

**Definition.** Two jointly continuous random variables  $X$  and  $Y$  are **independent** exactly when the joint density is the product of the marginal ones:

$$f(x, y) = f_X(x)f_Y(y),$$

for all  $x$  and  $y$ .

**Example 7.6.** Previous example, continued. Compute marginal densities and determine whether  $X$  and  $Y$  are independent.

**Sol.** We have

$$f_X(x) = \int_{x^2}^1 \frac{21}{4} \cdot x^2 y dy = \frac{21}{8} x^2 (1 - x^4),$$

for  $x \in [-1, 1]$ , and 0 otherwise. Moreover,

$$f_Y(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4} \cdot x^2 y dx = \frac{7}{2} y^{5/2},$$

where  $y \in [0, 1]$ , and 0 otherwise. The two random variables  $X$  and  $Y$  are clearly not independent, as  $f(x, y) \neq f_X(x)f_Y(y)$ .  $\square$

**Example 7.7.** Let  $(X, Y)$  be a random point in a square of length 1 with the bottom left corner at the origin. Are  $X$  and  $Y$  independent?

**Sol.** Note that

$$f(x, y) = \begin{cases} 1 & \text{if } (x, y) \in [0, 1] \times [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

The marginal densities are

$$\begin{cases} f_X(x) = 1 & \text{if } x \in [0, 1], \\ f_Y(x) = 1 & \text{if } y \in [0, 1], \\ 0 & \text{otherwise.} \end{cases}$$

Therefore,  $X$  and  $Y$  are independent.  $\square$

**Example 7.8.** Let  $(X, Y)$  be a random point in the triangle  $\{(x, y) \mid 0 \leq y \leq x \leq 1\}$ . Are  $X$  and  $Y$  independent?

**Sol.** Note that

$$f(x, y) = \begin{cases} 2 & 0 \leq y \leq x \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The marginal densities are

$$\begin{cases} f_X(x) = 2x & \text{if } x \in [0, 1] \\ f_Y(y) = 2(1 - y) & \text{if } y \in [0, 1] \\ 0 & \text{otherwise.} \end{cases}$$

So  $X$  and  $Y$  are no longer distributed uniformly and no longer independent.  $\square$

**Note.** We can make a more general conclusion. Assume that  $(X, Y)$  is a jointly continuous pair of random variables, uniform on a compact set  $S \subset \mathbb{R}^2$ . If they are to be independent, their marginal densities have to be constant, thus uniform on some sets, say  $A$  and  $B$ , and then  $S = A \times B$ . (If  $A$  and  $B$  are both interval, then  $S = A \times B$  is a rectangle, which is the most common example of independence.)

**Example 7.9.** Mr. and Mrs. Smith agree to meet at specified location “between 5 and 6 p.m.” Assume that they both arrive there at a random time between 5 and 6 and that their arrivals are independent. (a) Find the density for the time one of them will have to wait for the other. (b) Mrs. Smith later tells you she had to wait; given this information, compute the probability that Mr. Smith arrived before 5:30.

**Sol.** Let  $X$  be the time when Mr. Smith arrives and let  $Y$  be the time when Mrs. Smith arrives, with the time unit 1 hour. The assumptions imply that  $(X, Y)$  is uniform on  $[0, 1] \times [0, 1]$ . For (a), let  $T = |X - Y|$ , which has possible values in  $[0, 1]$ . So fix  $t \in [0, 1]$  and compute

$$\begin{aligned} P(T \leq t) &= P(|X - Y| \leq t) \\ &= P(-t \leq X - Y \leq t) \\ &= P(X - t \leq Y \leq X + t) \\ &= 1 - (t - t)^2 \\ &= 2t - t^2, \end{aligned}$$

and so  $f_T(t) = 2 - 2t$ . For (b),

$$P(X \leq 0.5 \mid X > Y) = \frac{P(X \leq 0.5, X > Y)}{P(X > Y)} = \frac{1/8}{1/2} = \frac{1}{4}.$$

$\square$

**Example 7.10.** Assume that  $X$  and  $Y$  are independent, that  $X$  is uniform on  $[0, 1]$ , and that  $Y$  has density  $f_Y(y) = 2y$ , for  $y \in [0, 1]$ , and 0 elsewhere. Compute  $P(X + Y \leq 1)$ .

**Sol.** The assumptions determine the joint density of  $(X, Y)$ :

$$f_X(x)f_Y(y) = f(x, y) = \begin{cases} 2y & \text{if } (x, y) \in [0, 1] \times [0, 1] \\ 0 & \text{otherwise.} \end{cases}$$



Then

$$P(X + Y \leq 1) = \iint_0^1 \int_0^{1-x} 2y \, dy \, dx = \frac{1}{3},$$

or

$$P(X + Y \leq 1) = \iint_0^1 \int_0^{1-y} 2y \, dx \, dy = \frac{1}{3}.$$

Thus, the answer is  $1/3$ . □

**Example 7.11.** Assume that you are waiting for two phone calls, from Alice and from Bob. The waiting time  $T_1$  for Alice's call has expectation 10 minutes and the waiting time  $T_2$  for Bob's call has expectation 40 minutes. Assume  $T_1$  and  $T_2$  are independent exponential random variables. What is the probability that Alice's call will come first?

**Sol.** We need to compute  $P(T_1 < T_2)$ . Assuming our unit is 10 minutes, we have, for  $t_1, t_2 > 0$ ,

$$f_{T_1}(t_1) = e^{-t_1} \quad \text{and} \quad f_{T_2}(t_2) = \frac{1}{4}e^{-t_2/4},$$

so that the joint density is  $f(t_1, t_2) = \frac{1}{4}e^{-t_1-t_2/4}$ , for  $t_1, t_2 \geq 0$ . Therefore

$$\begin{aligned} P(T_1 < T_2) &= \int_0^\infty \int_{t_1}^\infty \frac{1}{4} \exp\left(-t_1 - \frac{t_2}{4}\right) dt_2 dt_1 \\ &= \int_0^\infty \exp\left(-\frac{5t_1}{4}\right) dt_1 \\ &= \frac{4}{5}. \end{aligned}$$

□

**Example 7.12 (Buffon Needle Problem).** Parallel lines at a distinct 1 are drawn on a large sheet of paper. Drop a needle of length  $l$  onto the sheet. Compute the probability that it intersects one of the lines.

**Sol.** Let  $D$  be the distance from the center of the needle to the nearest line and let  $\Theta$  be the acute angle relative to the lines. We will, reasonably, *assume* that  $D$  and  $\Theta$  are independent and uniform on their respective intervals  $0 \leq D \leq 1/2$  and  $0 \leq \Theta \leq \pi/2$ . Then

$$f_{D,\Theta}(d, \theta) = f_D(d) \cdot f_\Theta(\theta) = \frac{4}{\pi},$$

for  $(d, \theta) \in S$ . And then

$$\begin{cases} f_D(d) = 2 & 0 \leq d \leq 1/2, \\ f_D(d) = 0 & \text{otherwise.} \\ f_\Theta(\theta) = 2/\pi & 0 \leq \theta \leq \pi/2, \\ f_\Theta(\theta) = 0 & \text{otherwise.} \end{cases}$$

Thus,

$$P(\text{the needle intersects a line}) = P\left(D < \frac{l}{2} \sin \Theta\right).$$

- Case I:  $l \leq 1$ . Then, the probability equals

$$P\left(D < \frac{l}{2} \sin \Theta\right) = \int_0^{\pi/2} \int_0^{1/2 \sin \theta} \frac{4}{\pi} dD d\theta = \frac{4}{\pi} \int_0^{\pi/2} \frac{l}{2} \sin \theta d\theta = \frac{4}{4\pi} \cdot \frac{l}{2} = \frac{2l}{\pi}.$$

When  $l = 1$ , you famously get  $\frac{2}{\pi}$ , which can be used to get (very poor) approximations for  $\pi$ .

- Case II:  $l > 1$ . Now, the curved  $d = \frac{l}{2} \sin \theta$  intersects  $d = 1/2$  at  $\theta = \arcsin \frac{1}{l}$ . The probability equals

$$\frac{4}{\pi} \left[ \frac{l}{2} \int_0^{\arcsin \frac{1}{l}} \sin \theta d\theta + \left( \frac{\pi}{2} - \arcsin \frac{1}{l} \right) \cdot \frac{1}{2} \right] = \frac{4}{\pi} \left[ \frac{l}{2} - \frac{1}{2} \sqrt{l^2 + 1} + \frac{\pi}{4} - \frac{1}{2} \arcsin \frac{1}{l} \right].$$

□

**Example 7.13.** Assume  $X_1, X_2, X_3$  are uniform on  $[0, 1]$  and independent. What is  $P(X_1 + X_2 + X_3 \leq 1)$ ?

**Sol.** The joint density is

$$f_{X_1, X_2, X_3}(x_1, x_2, x_3) = f_{X_1}(x_1)f_{X_2}(x_2)f_{X_3}(x_3) = \begin{cases} 1 & \text{if } (x_1, x_2, x_3) \in [0, 1]^3, \\ 0 & \text{otherwise.} \end{cases}$$

Here is how we get the answer:

$$P(X_1 + X_2 + X_3 \leq 1) = \int_0^1 \int_0^{1-x_1} \int_0^{1-x_1-x_2} dx_3 dx_2 dx_1 = \frac{1}{6}.$$

In general, if  $X_1, \dots, X_n$  are independent and uniform on  $[0, 1]$ ,

$$P(X_1 + \dots + X_n \leq 1) = \frac{1}{n!}.$$

□

### 7.3 Conditional Distributions

The conditional p. m. f. of  $X$  given  $Y = y$  is, in the discrete case, given simply by

$$p_X(x|Y = y) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

This is trickier in the continuous case, as we cannot divide by  $P(Y = y) = 0$ .

**Definition.** For a jointly continuous pair of random variables  $X$  and  $Y$ , we define the *conditional density of  $X$  given  $Y = y$*  as follows:

$$f_X(x|Y = y) = \frac{f(x, y)}{f_Y(y)},$$

where  $f(x, y)$  is, of course, the joint density of  $(X, Y)$ .

Observe that when  $f_Y(y) = 0$ ,  $\int_{-\infty}^{\infty} f(x, y) dx = 0$ , and so  $f(x, y) = 0$  for every  $x$ . So we have a  $\frac{0}{0}$  expression, which we define to be 0.

Here is a “physicist’s proof” why this should be the conditional density formula:

$$\begin{aligned} P(X = x + dx | Y = y + dy) &= \frac{P(X = x + dx, Y = y + dy)}{P(Y = y + dy)} \approx \frac{f(x, y) dx dy}{f_Y(y) dy} \\ &= \frac{f(x, y)}{f_Y(y)} dx \\ &= f_X(x | Y = y) dx. \end{aligned}$$

**Example 7.14.** Let  $(X, Y)$  be a random point in the triangle  $\{(x, y) \mid x, y \geq 0, x + y \leq 1\}$ . Compute  $f_X(x | Y = y)$ .

**Sol.** The joint density  $f(x, y)$  equals 2 on the triangle. For a given  $y \in [0, 1]$ , we know that, if  $Y = y$ ,  $X$  is between 0 and  $1 - y$ . Moreover,

$$f_Y(y) = \int_0^{1-y} 2 dx = 2(1 - y).$$

Therefore,

$$f_X(x | Y = y) = \begin{cases} \frac{1}{1 - y} & 0 \leq x \leq 1 - y. \\ 0 & \text{otherwise.} \end{cases}$$

In other words, given  $Y = y$ ,  $X$  is distributed uniformly on  $[0, 1 - y]$ , which is hardly surprising.  $\square$

**Example 7.15.** Suppose  $(X, Y)$  has joint density

$$f(x, y) = \begin{cases} \frac{21}{4}x^2y & x^2 \leq y \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

Compute  $f_X(x | Y = y)$ .

**Sol.** We compute first

$$f_Y(y) = \frac{21}{4}y \int_{-\sqrt{y}}^{\sqrt{y}} x^2 dx = \frac{7}{2}y^{5/2}.$$

for  $y \in [0, 1]$ . Then

$$f_X(x | Y = y) = \frac{\frac{21}{4}x^2y}{\frac{7}{2}y^{5/2}} = \frac{3}{2}x^2y^{-3/2},$$

where  $-\sqrt{y} \leq x \leq \sqrt{y}$ . Suppose we are asked to compute  $P(X \geq Y | Y = y)$ . This makes no literal sense because the probability  $P(Y = y)$  of the condition is 0. We reinterpret this expression as

$$P(X \geq y | Y = y) = \int_y^\infty f_X(x | Y = y) dx(y \text{ fixed}),$$

which equals

$$\int_y^{\sqrt{y}} \frac{3}{2}x^2y^{-3/2} dx = \frac{1}{2}y^{-3/2} (y^{3/2} - y^3) = \frac{1}{2}(1 - y^{3/2}).$$

$\square$

## 8 Properties of Expectation and Variance

The Proba. Distribution	Parameter	Prob. Distribution Function	Mean	Variance
Binomial Distribution	$(n, p)$	$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$	$np$	$np(1-p)$
Geometric Distribution	$(p)$	$\Pr(X = k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{(1-p)}{p^2}$
Normal Distribution	$(\mu, \sigma^2)$	$f(x \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	$\mu$	$\sigma^2$
Uniform Distribution	$(a, b)$	$f(x a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b, \\ 0 & x < a \text{ or } x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponential Distribution	$(\lambda)$	$f(x \lambda) = \lambda e^{-\lambda x}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Poisson Distribution	$(\lambda)$	$f(k \lambda) = \frac{e^{-\lambda} \lambda^k}{k!}$	$\lambda$	$\lambda$

### 8.1 Expectations

**Definition.** Given a pair of random variables  $(X, Y)$  with joint density  $f$  and another function  $g$  of two variables,

$$E_g(X, Y) = \iint g(x, y) f(x, y) dx dy;$$

if instead  $(X, Y)$  is a discrete pair with joint probability mass function  $p$ , then

$$E_g(X, Y) = \sum_{x, y} g(x, y) p(x, y).$$

**Example 8.1.** Assume that two among the 5 items are defective. Put the items in a random order and inspect them one by one. Let  $X$  be the number of inspections needed to find the first defective item and  $Y$  the number of additional inspections needed to find the second defective item. Compute  $E|X - Y|$ .

**Sol.** The joint p. m. f. of  $(X, Y)$  is given by the following table, which lists  $P(X = i, Y = j)$ , together with  $|i - j|$  in parentheses, whenever the probability is non-zero:

$$P(X = 1, Y = 1) = P(X = 1)P(Y = 1|X = 1) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10},$$

$$P(X = 1, Y = 2) = P(X = 1)P(Y = 2|X = 1) = \frac{2}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{10}.$$

$i \backslash j$	1	2	3	4
1	.1(0)	.1(1)	.1(2)	.1(3)
2	.1(1)	.1(0)	.1(1)	0
3	.1(2)	.1(1)	0	0
4	.1(3)	0	0	0

The answer is, therefore,  $E|X - Y| = 1.4$ . □

**Example 8.2.** Assume that  $(X, Y)$  is a random point in the right triangle  $\{(x, y) \mid x, y \geq 0, x + y \leq 1\}$ . Compute  $EX$ ,  $EY$ , and  $EXY$ .

**Sol.** Note that the density is 2 on the triangle, and so,

$$\begin{aligned} EX &= \int_0^1 \int_0^{1-x} x \cdot 2 \, dy \, dx \\ &= \int_0^1 2x(1-x) \, dx \\ &= \frac{1}{3}, \end{aligned}$$

and, therefore, by symmetry,  $EY = EX = 1/3$ . And

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^{1-x} xy \cdot 2 \, dy \, dx \\ &= \int_0^1 2x \frac{(1-x)^2}{2} \, dx \\ &= \frac{1}{12}. \end{aligned}$$

□

## 8.2 Linearity and Monotonicity of Expectation

**Theorem 8.1.** *Expectation is linear and monotone:*

1. For constants  $a$  and  $b$ ,  $E(aX + b) = aE(X) + b$ .
2. For arbitrary random variables  $X_1, \dots, X_n$  whose expected values exists,

$$E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n).$$

3. For two random variables  $X \leq Y$ , we have  $EX \leq EY$ .

*Proof.* 2. We will check the second property for  $n = 2$  and the continuous.

$$E(X+Y) = \iint (x+y) f(x, y) \, dx \, dy = \iint x f(x, y) \, dx \, dy + \iint y f(x, y) \, dx \, dy = EX + EY.$$

3. Note that  $Z \geq 0 \implies EZ \geq 0$ . Let  $Z = Y - X$ , then it is proved.

□

**Note.** We emphasize again that linearity holds for arbitrary random variables which *do not need to be independent!* This is very useful. For example, we can often write a random variable  $X$  as a sum of (possibly dependent) indicators,  $X = I_1 + \dots + I_n$ . An instance of this method is called the *indicator trick*.

**Example 8.3.** Assume that an urn contains 10 black, 7 red, and 5 white balls. Select 5 balls (a) with and (b) without replacement and let  $X$  be the number of red balls selected. Compute  $EX$ .

**Sol.** Let  $I_i$  be the indicator of the event that the  $i$ -th ball is red, that is,

$$I_i = I_{i\text{-th ball is red}} = \begin{cases} 1 & \text{if } i\text{-th ball is red,} \\ 0 & \text{otherwise.} \end{cases}$$

In both cases,  $X = \sum_{i=1}^5 I_i$ .

1. In (a),  $X$  is Binomial( $5, \frac{7}{22}$ ), so we know that  $EX = 5 \cdot 7/22$ , but we will not use this knowledge. Instead, it is clear that

$$EI_1 = 1 \cdot P(\text{1st ball is red}) = \frac{7}{22} = EI_2 = \dots = EI_5.$$

Therefore, by additivity,  $EX = 5 \cdot 7/22$ .

2. For (b), one solution is to compute the p. m. f. of  $X$ ,

$$P(X = i) = \frac{\binom{7}{i} \binom{15}{15-i}}{\binom{22}{5}},$$

where  $i = 0, 1, \dots, 5$ , and then

$$EX = \sum_{i=0}^5 i \frac{\binom{7}{i} \binom{15}{5-i}}{\binom{22}{5}}.$$

However, the indicator trick *works exactly as before*. (the fact that  $I_i$  are now dependent does not matter) and so the answer is also exactly the same,  $EX = 5 \cdot 7/22$ .

□

**Example 8.4 (Matching Problem - revisited).** Assume that  $n$  people buy  $n$  gifts, which are then assigned at random, and let  $X$  be the number of people who receive their own gift. What is  $EX$ ?

**Sol.** This is another problem very well suited for the indicator trick. Let  $I_i = I_{\text{person } i \text{ receives own gift}}$ . Then,  $X = \sum_{i=1}^n I_i$ . Moreover,  $EI_i = 1/n$ , for all  $i$ , and so  $EX = 1$ . □

**Example 8.5.** Five married couples are seated around a table at random. Let  $X$  be the number of wives who sit next to their husbands. What is  $EX$ ?

**Sol.** Now, let  $I_i = I_{A_i}$ , where  $A_i = \{\text{wife } i \text{ sits next to her husband}\}$  such that  $P(A_i) = (8! \cdot 2!)/9!$ . Then,  $X = \sum_{i=1}^5 I_i$ , and  $EI_i = 2/9$ , so that  $EX = 10/9$ . □