# 1
# *Thinking in Probability*

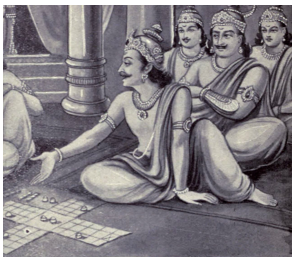

Figure 1.1: A portrait of Shakuni's game of dice from the Indian epic *Mahabharata* composed in the third century BCE. (*courtesy: Wikimedia Commons*).

CHANCE PERMEATES our physical and mental universe. While the role of chance in human lives has had a longer history, starting with the more authoritative influence of the nobility, the more rationally sound theory of probability and statistics has come into practice in diverse areas of science and engineering starting from the early to mid-twentieth century.[1] Practical applications of statistical theories proliferated to such an extent in the previous century that the American government-sponsored RAND corporation published a six hundred page book that wholly consisted of a random number table and a table of standard normal deviates.[2] One of the primary objectives of this book was to enable a computer simulated approximate solution of an exact but unsolvable problem by a procedure known as the *Monte Carlo method* devised by Fermi, von Neumann, and Ulam in the 1930-40s.[3]

Statistical methods are the mainstay of conducting modern scientific experiments. One such experimental paradigm is known as a *randomized control trial* that is widely used in a variety of fields like psychology, drug verification, testing efficacy of vaccines, agricultural sciences, demography, etc. These statistical experiments require sophisticated sampling techniques in order to nullify experimental biases. With the explosion of information in the modern era, the need to develop advanced and accurate predictive capabilities have grown manifold. This has led to the emergence of modern artificial intelligence (AI) technologies. Further, climate change has become a reality of the modern civilization. Accurate prediction of weather and climatic patterns relies on sophisticated AI and statistical techniques. It is impossible to think of a modern economy and social life without the influence and role of chance, and hence without the influence of technological interventions based on statistical principles. We must begin this journey by learning the foundational tenets of probability and statistics.

[1] *Philosophy of Probability and Statistical Modelling* by *Mauricio Suarez*, Cambridge University Press, 2020. DOI: 10.1017/9781108985826

[2] *A Million Random Digits with 100,000 Normal Deviates*, RAND corporation, 2001. (Originally published in 1955).

[3] *Randomness and the Twentieth Century* by *Alfred M. Bock*, The Antioch Review, **27** (1), pp. 40-61, 1967.

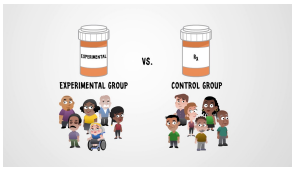


Figure 1.2: Schemata of a *randomized control trial* for evaluating the efficacy of a treatment intervention by a new drug launched in the market.

## 1.1 *Chapter objectives*

The chapter objectives are listed as follows.

1. Students will learn the fundamental axioms of probability.
2. Students will apply elementary principles of probability, permutation, and combination to solve simple numerical examples.
3. Students will learn the meaning of random variables and formulate solutions to problems involving random events.

4. Students will learn to apply the Bayes' theorem and the law of total probability to solve complex problems.

5. Students will learn to calculate statistical averages in terms of expectation of random variables.

6. Students will learn to use the techniques of computing probability and expectation of random events to solve a practical simulation project on one dimensional random walk.

## 1.2 *Chapter project: Random walk on a lonely island*

### 1.2.1 *Prologue: Will Squeaky drift off to the edge and fall off the cliff or keep hopping back and forth forever?*

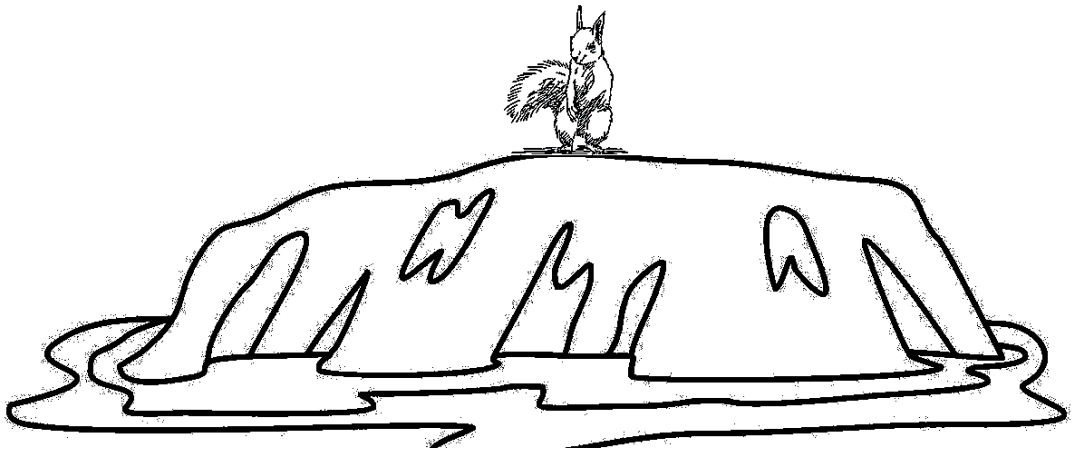

Figure 1.3: Squeaky is trapped in a lonely island hill with sharp cliffs on both sides.

**Our friend Squeaky is trapped somewhere in the middle of a lonely island hill with sharp cliffs on both sides. Squeaky is excited and jumps around in her merry way. At any given instance, she makes a decision to hop to the left or to the right independent of her past moves. Squeaky is unaware of the impending danger.**

**In this project, we will use calculations based on the principles of conditional probability, the law of total probability, and the law of total expectation to predict her fate. In other words, what are the odds that she will bounce around on the island hill, her left-sided moves balancing out her right-sided moves on an average, and never actually trip and fall off on either side? Or will chance play the devil's role and will she eventually drift off to one side and perish? And if the latter turns out to be true, then what is her life expectancy in terms of the total number of hops starting from her first move? Does a certain initial position on the hill give her the best chance to survive the longest?**

**In addition to our theoretical calculations, we will also build a computer simulation of her actions to corroborate our result. For convenience, we shall assume that the island is one dimensional, i.e. Squeaky's movements are restricted exclusively to lateral directions (left or right). While we build the computer-simulated solution, we will learn to apply a**

**random number generator using a computer software in order to mimic Squeaky's mental choices to hop either to the left or to the right independent of her past moves.**

## 1.3 *Deterministic vs probabilistic outcomes*

### 1.3.1 *Deterministic outcomes*


Figure 1.4: Number of different permutations of five differently colored buttons is 5! i.e., there are $5 \times 4 \times 3 \times 2 \times 1 = 120$ different ways of arranging these five buttons.

Permutations:

Consider an assortment of five differently colored buttons. A simple question may be to find out all the different ways in which we may be able to arrange the five buttons without piling them on top of each other.

This is a classic example of the number of *permutations* of $n$ distinct things. If we consider five empty slots in which to host the individual buttons, and if we begin with the leftmost slot; then this slot may be occupied by any of the five buttons. So depending on the color we choose, there are five different ways of filling the leftmost slot. Subsequently, we are left with four differently colored buttons, and hence the second slot can be filled in four different ways. This is followed by the third slot which can be occupied in three distinct manners. The penultimate and the ultimate slot can be filled in two and one different ways respectively. Therefore, the number of *permutations* of $n$ different things is $n! = n \times (n-1) \times \cdots \times 2 \times 1$.

However, if there are $r$ different *types* of $n = n_1 + n_2 + \cdots + n_r$ objects; then there are $\frac{n!}{n_1!n_2!n_3!...n_r!}$ different ways of arranging them. Here, the $i^{th}$ type has $n_i$ counts, $i = 1, 2, 3, ..., r$.

e.g., if we have nine buttons of which three are of red color, four are of green color, and two are of blue color; then there are $\frac{9!}{3!4!2!} = 1260$ ways of arranging them.

Alternatively, we may have $n$ different things and we may want to know the number of permutations by taking only $r \leq n$ of them at a time. The number of possible ways are

$$P_r^n = n \times (n-1) \times (n-2) \times \cdots \times (n-r+1) = \frac{n!}{(n-r)!}.$$

e.g., let us say there are nine slots and four differently colored buttons. There are $\frac{9!}{5!} = 3024$ ways of arranging them.

Combinations:

In many other situations, the order of arrangement is not so important. In such cases, we may only care about the number of subsets of $r$ items from amongst a total of $n$ items. The number of ways $n$ things can be combined by taking $r$ at a time is given by the formula

$$C_r^n = \binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

e.g., if there are nine slots and four identically colored buttons which can be placed in any of these slots; then there are $\binom{9}{4} = 126$ different designs/patterns that can emerge upon hosting any five buttons in the nine slots. Obviously, $C_r^n < P_r^n$ when $r > 1$.

### 1.3.2 *Probabilistic outcomes*

Most importantly, permutations and combinations belong to a class of experiments that have deterministic outcomes. There are a finite and fixed number of ways of arranging or collecting (combining) items. None of the aforementioned examples have a chance outcome. However, we may have to perform experiments whereby the outcome is not certain, at least not in the *a priori* sense. e.g., we may ask that in any given arrangement of the five distinctly colored buttons in the five available slots, what is the probability that the first slot is filled by a red color button? Implicit in this question is the fact that this particular arrangement of the five buttons is made blindfolded (without the person actually making a conscious decision of placing the red button in the first slot). Under such circumstances, the placement of the red button in the first slot is a matter of chance. The probability of such an outcome is $1/5$ because only one out of the possible five differently colored buttons that could have been placed in the first slot is red. We shall devote the rest of this book to the study of random events and statistical experiments that have a probabilistic outcome.

### 1.3.3 *A note of caution*

Statistical forecasts depend on good and reliable data. Biases in data can skew statistical predictions hugely as is often noticed in faulty exit poll results. In order to address these biases, statisticians are often concerned with appropriate design of their experiments.

Moreover, statistical inferences are based on the principles of probability (chance). They explain what outcome is likely to happen. However, in order to understand the rationale behind a particular outcome or the underlying principles responsible for a certain observation, one has to rely on physical theories that fall outside the scope of statistical techniques. Statistical theories shed light on idealized averages[4] of stochastic phenomena. Thus, the reach of statistical inferences may be far removed from individual experiences. A distinctive dimension of reality is its individual aspect which may not be gleaned from statistical approaches.[5]

[4] Here the word *averages* is used in a broader sense of all statistical moments and not just the mean value.

[5] **cf.** pg. 5 of *The Undiscovered Self* by *Carl Gustav Jung*, Routledge Classics, 2021 (Reprint of the 1958 edition).

## 1.4 *Foundations of probability*

### 1.4.1 *Definition: Probability*

It is the measure of likelihood that an event will occur. e.g., We may ask: what are the chances that it will rain today? Most weather prediction websites may give us an answer in terms of a probability measure, 75% (say).

### 1.4.2 *Definition: Statistics*

It is the branch of mathematics that deals with the collection (sampling), organization, analysis and interpretation of data including making inferences and forecasts. It relies on the principles of probability.

### 1.4.3 *Definition: Probability space*

A probability space comprises a triple $(\Omega, \mathfrak{F}, P)$. Here $\Omega$ denotes the *sample space* which is the set of all possible outcomes,[6] $\mathfrak{F}$ denotes the *$\sigma$-algebra* that is a collection of all events of

[6] An *outcome* is the result of a single realization of the model experiment.

concern to us in a certain statistical experiment and is generated by $\Omega$, and $P$ is the *probability measure* defined as a function $P : \mathfrak{F} \to [0, 1]$. We may think of $\mathfrak{F}$ as an *event space*.

e.g., for the coin tossing experiment illustrated in Figure 1.5, the $\sigma$-algebra generated by $\Omega_2 = \{HH, HT, TH, TT\}$ can be taken as the power set of $\Omega_2$,

$$\mathfrak{F} = 2^{\Omega_2} = \Big\{\{\}, \{HH\}, \{TT\}, \{HT\}, ..., \{HT, TT\}, ..., \{HH, HT, TH, TT\}\Big\}.$$

The cardinality of $\mathfrak{F}$ is $2^{|\Omega_2|} = 2^4 = 16$. The $\sigma$-algebra defined above is the largest such set. The smallest $\sigma$-algebra over $\Omega_2$ is $\Big\{\{\}, \{HH, HT, TH, TT\}\Big\}$ The probability of observing two successive heads is $1/4$.

It may be useful to state here that if we have disjoint sets (events) $E_1, E_2, ... \in \mathfrak{F}$, then $\cup_i E_i \in \mathfrak{F}$. A rigorous treatment of $\sigma$-algebra will be avoided in this introductory level text, wherever necessary we will loosely refer to the notion of an event space.

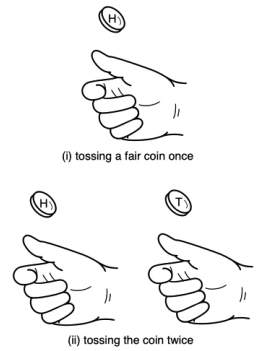


Figure 1.5: The outcome of a toss of a fair coin is *heads* or tails. Therefore, $\Omega_1 = \{H, T\}$ with the usual abbreviations for heads and tails. We may conduct an experiment whence we toss the coin twice whence $\Omega_2 = \{HH, HT, TH, TT\}$.

### 1.4.4 *Axioms of probability*

We begin this short section by asking: *why do we need axioms at all?* In fact, the first foundational axioms of mathematics appeared only as recent as 1879, courtesy Gottlob Frege.[7] Axioms may be regarded as *a priori propositions* whose veracity is accepted universally without requiring their validation by demonstration. The utility of axioms lies in the fact that they enable the deduction of realizable experiences that can be supported by sense perceptions.[8,9]

The axioms of probability were formulated by Andrey N. Kolmogorov in 1933.

1. $P(E) \geq 0$, for all $E \in \mathfrak{F}$ (non-negativity),
2. $P(\Omega) = 1$ (unitarity), and
3. $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$ for a countable sequence of disjoint events $E_1, E_2, ...$ ($\sigma$-additivity).

[7] `https://iep.utm.edu/frege/`

[8] *Russell's mathematical logic* by *Kurt Gödel* (1944), Benacerraf and Putnam, pp. 447-469, 1983.

[9] *The Role of Axioms in Mathematics* by *Kenny Easwaran*, Erkenn (Springer), **68**, pp.381-391, 2008.

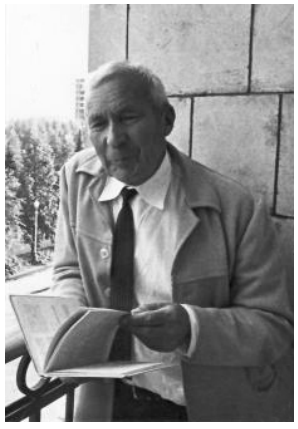

Figure 1.6: Russian Mathematician A. N. Kolmogorov (*courtesy: Wikimedia Commons*).

### 1.4.5 *Supplementary properties of probability measure*

In addition to the axioms of probability listed above, it is often helpful to consider the following properties of $P$ while performing calculations.

1. Consider $E_1, E_2 \in \mathfrak{F}$, then $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$. This result may be generalised to $n$ events $E_1, E_2, E_3, ..., E_n$ by induction. This result is known as the *principle of inclusion-exclusion.*
2. If $E_1$ and $E_2$ are *independent* events, then $P(E_1 \cap E_2) = P(E_1)P(E_2)$.
3. If $A^c$ stands for the complementary event of $A$, then $P(A^c) = 1 - P(A)$.
4. The probability of the impossible event is zero, i.e. $P(\{\}) = 0$.
5. A distinction must be made between *mutually exclusive* (disjoint) events and *independent* events. $E_1$ and $E_2$ are mutually exclusive when $E_1 \cap E_2 = \{\}$. In such a case, $P(E_1 \cap E_2) =$

$P(\{\}) = 0$, and $P(E_1|E_2) = 0$. On the other hand, if two events $A_1$ and $A_2$ are independent, then $P(A_1 \cap A_2) = P(A_1)P(A_2)$, and $P(A_1|A_2) = P(A_1)$. In essence, two events are mutually exclusive if they cannot happen concurrently; whereas two independent events may happen concurrently but the outcome of one does not influence the outcome of the other.[10]

The symbols $\cap$ and $\cup$ denote *overlapping* and *union* of events, respectively (**cf.** Figure 1.7).

[10] The prevailing weather pattern in a given locality may be either sunny or rainy because these are mutually exclusive weather events in commonly used terminology. The outcomes of tossing a fair coin twice are independent events.

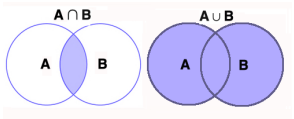


Figure 1.7: Venn diagram showing overlapping events and union of events

### 1.4.6 *Example: Defining events in probability space*

Two dice are thrown. Let E be the event that the sum of the dice is odd, let F be the event that the first die lands on 1, and let G be the event that the sum is 5. Describe the events $EF, E \cup F, FG, EF^c, EFG$.

Solution: $EF = \{(1,2),(1,4),(1,6)\}$
$E \cup F = \{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6)$ or any of the 15 possibilities where the first die is not 1 and the second die is odd when the first is even and even when the first is odd.$\}$
$FG = \{(1,4)\}$
$EF^c = \{$ any of the 15 possible outcomes where the first die is not 1 and the two dice are not either both even or both odd$\}$
$EFG = FG$.

### 1.4.7 *Example: Rolling two dice concurrently*

Consider an experiment comprising throws of two independent dice. The sample set is the Cartesian product comprising ordered pairs,

$$\Omega = \{1,2,3,4,5,6\} \times \{1,2,3,4,5,6\} = \{(1,1),(1,2),...,(2,1),(2,2),...,(6,5),(6,6)\}.$$

We may be interested in knowing the odds that the sum of the outcomes from each die is greater than equal to ten. In this case, the event space is

$$\mathfrak{E} = \{(4,6),(5,5),(5,6),(6,4),(6,5),(6,6)\}$$

and hence, the required probability is $\frac{|\mathfrak{E}|}{|\Omega|} = \frac{6}{36}$.

### 1.4.8 *Example: Probability of a complementary event*

$A$, $B$ and $C$ are 3 mutually exclusive and exhaustive event of a random experiment such that $P(B) = \frac{3}{2}P(A)$ and $P(C) = \frac{1}{2}P(B)$. What is probability of non-occurrence of event $A$.

Solution: $P(C) = \frac{1}{2} \times \frac{3}{2}P(A) = \frac{3}{4}P(A)$.

$$P(A) + P(B) + P(C) = 1.$$
$$P(A) + \frac{3}{2}P(A) + \frac{3}{4}P(A) = 1.$$
$$P(A)\left(1 + \frac{3}{2} + \frac{3}{4}\right) = 1.$$
$$P(A) = \frac{4}{13}.$$
$$P(A^c) = 1 - \frac{4}{13} = \frac{9}{13}.$$

*1.4.9 Example: Probabilities of composite events originating from rolling two dice.*

If 2 dice are thrown, what is the probability that sum is a) greater than 9, b) neither 7 nor 11.

Solution: The cardinality of the sample space, $n(\Omega) \equiv |\Omega| = 36$.
Let $S_i$ be the event when the sum of the outcomes of the two dice equal $i$.

**a)** $P(\text{sum is greater than } 9) = P(S_{10}) + P(S_{11}) + P(S_{12})$

$$= \frac{3}{36} + \frac{2}{36} + \frac{1}{36}$$
$$= \frac{1}{6}$$

**b)** Event $A \equiv S_7 \quad P(A) = \frac{1}{6}$.
Event $B \equiv S_{11} \quad P(B) = \frac{1}{18}$.

$$P\left(A^c \cap B^c\right) = P(\Omega) - P(A \cup B)$$
$$= 1 - (P(A) + P(B))$$
$$= 1 - \left(\frac{1}{6} + \frac{1}{18}\right)$$
$$= \frac{7}{9}.$$

## 1.5 Random variable

Consider a probability space $(\Omega, \mathfrak{F}, P)$. A *random variable*[11] is a *measurable* function, $X : \Omega \to \mathbb{R}$, that maps each outcome in the sample space to a real number, i.e.

$$\{\omega \in \Omega;\ X(\omega) \leq x\} \in \mathfrak{F}, \quad x \in \mathbb{R}.$$

A random variable may be *discrete* or *continuous* depending on whether it takes on discrete values or a continuum of values. Next, we will discuss some concrete examples.

[11] Notation: By convention, a random variable is denoted by an uppercase letter such as $X$.

### 1.5.1 Example: coin tossing experiments

Consider a simple experiment of tossing a fair coin. $\Omega_1 = \{H, T\}$. $X(H) = 1,\ X(T) = 0$ is a re-labelling of every outcome in $\Omega_1$ to a measurable space (often taken as $\mathbb{R}$). In the case of an experiment where we toss the coin twice, the outcomes are extracted from $\Omega_2 = \{HH, HT, TH, TT\}$. We may wish to know the number of heads observed in a given realization. Therefore, $X(HH) = 2,\ X(HT) = 1,\ X(TH) = 1,\ X(TT) = 0$.

### 1.5.2 Example: Indicator random variable

Often, in calculations, it is convenient to define an *indicator random variable*[12] as follows

$$\mathbb{1}_A(\omega) \equiv \mathbb{I}_{\{\omega \in A\}} = \begin{cases} 1; & \omega \in A \\ 0; & \omega \notin A \end{cases} \tag{1.1}$$

[12] It is also known as *Bernoulli random variable*.

In the second experiment described above where we toss a fair coin twice, we may be interested in an outcome where we observe at least one head. We may define $A = \{HH, HT, TH\}$ and use the indicator random variable to represent the events where we observe at least one head. In the latter section of this chapter, we will use the indicator random variable to solve a problem encountered by a hiring manager of a company.

### 1.5.3 Example: Defining random variables for events

Consider an fair coin being tossed thrice. Consider the number of heads obtained after three tosses. Find the sample space, and therefore define a random variable. Also find the probabilities associated for each value of the random variable.

Solution: Sample space, $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
Let $X$ be the number of heads obtained after three tosses.

$$X(\omega) = \begin{cases} 3, \text{if } \omega \in \{HHH\} \\ 2, \text{if } \omega \in \{HHT, HTH, THH\} \\ 1, \text{if } \omega \in \{HTT, THT, TTH\} \\ 0, \text{if } \omega \in \{TTT\} \end{cases}$$

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

### 1.5.4 Example: Defining an indicator random variable for a stochastic event

Consider an unbiased die being rolled once, where the outcome of interest is one where there is prime number. Find the sample space, the collection of events we are

interested in and therefore define an indicator random variable to represent when a prime number appears. Also find the probabilities associated for each value of the random variable.

Solution: Sample space, $\Omega = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\}\}$
Let $A = \{\{2\},\{3\},\{5\}\}$
Then, the indicator random variable is

$$\mathbb{1}_A(\omega) = \begin{cases} 1; \text{if } \omega \in \{\{2\},\{3\},\{5\}\} \\ 0; \text{if } \omega \in \{\{1\},\{4\},\{6\}\} \end{cases}$$

$$P(X=0) = \frac{3}{6} = \frac{1}{2},\ P(X=1) = \frac{3}{6} = \frac{1}{2}.$$

The examples discussed above are discrete type random variables. Some examples of continuous random variables are listed below.[13]

1. Time duration between successive arrivals of buses in a station (this random time interval follows an exponential distribution).
2. Distribution of wealth in a society follows a Pareto distribution unraveling the fact that a high proportion of wealth is held by a small fraction of people in a society.
3. Scores obtained by students in an engineering class may follow a bell-shaped curve (see Figure 1.8), etc.

[13] We will discuss different types of discrete and continuous random variables (and their probability distributions) in more detail in the next chapter.

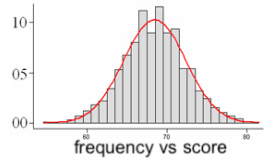


Figure 1.8: Scores obtained by students in a class may follow a Gaussian distribution. The scores can take on a continuum of values between the lowest score and the highest score.

## 1.6 *Conditional probability*

Occurrence of certain events may depend on the occurrence of other events. In fact, their likelihood of happening may be boosted (or diminished) by the outcomes of the preceding events. e.g., the chances of rain are certainly higher on a cloudy day than on a day with clear skies. In this simple example, knowledge of the prevailing weather (cloudy/sunny) can greatly enhance our ability to predict the chances of rain. This underscores the importance of calculating *conditional probability* where we may want to know the chance of occurrence of a certain event conditioned upon our knowledge of a preceding event.

Consider two events $A$ and $B$. The conditional probability of event $A$ given the occurrence of event $B$ is given by the following relation.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{1.2}$$

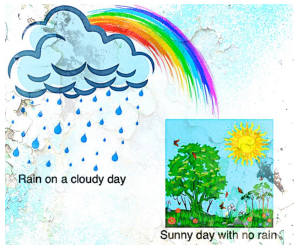


Figure 1.9: Likelihood of occurrence of an event (rain) may depend on another event (sunny/cloudy).

If $A$ and $B$ are independent events, then $P(A \cap B) = P(A)P(B) \implies P(A|B) = P(A)$. Let us understand this concept by considering another simple example.

Let $A$ be the event that we make the following observation on two successive tosses of a fair coin: "heads" followed by "tails". Let $B$ be the event that in any two succes-

sive tosses of a fair coin, the outcome of the first toss is "heads". Let us evaluate the conditional probability $P(A|B)$ using two different approaches.

1. Method 1: Given the knowledge of the event $B$, the only possible way that the event $A$ can happen is if the outcome of the second event turns out to be "tails".

$$\text{Therefore, } P(A|B) = \text{Prob(second toss turns up as "tails")} = \frac{1}{2}.$$

2. Method 2: An alternative approach would be to use the formula (1.2).

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/2 \times 1/2}{1/2} = \frac{1}{2}.$$

   Here, it is essential to explain the calculation of $P(A \cap B)$. First, we shall analyze the meaning of the event $A \cap B$. $A \cap B$ stands for the event which is common to both $A$ and $B$, i.e. it is that special event when each of event $A$ and event $B$ are guaranteed to have happened. A little introspection may reveal that this event must be the appearance of "heads" in the first toss and "tails" in the second toss which happens with a probability $1/2 \times 1/2 = 1/4$.[14]

[14] It may help to reflect if the event $B$ may be a candidate for the event $A \cap B$. It turns out that the occurrence of event $B$ does not guarantee the event $A$ as there is a possibility that the second toss may turn out to be "heads". You may proceed by a careful elimination process to comprehend the special event $A \cap B$.

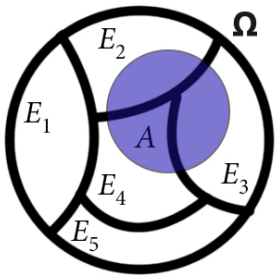


Figure 1.10: Here, the sample space $\Omega$ is partitioned into $E_1, E_2, ..., E_5$ in order to facilitate the computation of the probability of the event $A$ in terms of the conditional probabilities.

*Example: Simple concurrent events*

Suppose that a bag contains 6 pink balls and 2 grey balls and we draw 2 balls randomly from the bag without replacement. If at each draw, each ball in the bag is equally likely to be drawn, what is the probability that both balls that are drawn are pink.

Solution: Let $R_{1p}$ be the event that first ball drawn is pink and let $R_{2p}$ be the event that second ball drawn is pink.
$P\left(R_{1p}\right) = \frac{6}{8}$. Given that the first ball selected is pink, there remains 5 pink balls and 2 grey balls. Therefore, $P\left(R_{2p} \mid R_{1p}\right) = \frac{5}{7}$.

$$P\left(R_{1p} \cap R_{2p}\right) = P\left(R_{1p}\right) P\left(R_{2p}|R_{1p}\right) = \frac{6}{8} \times \frac{5}{7} = \frac{30}{56} = \frac{15}{28}.$$

*Example: Estimating chance of a defect in a supply line*

A box contain 2000 components of which 5% are defective, second box contain 500 components of which 40% are defective, 2 other boxes contains 1000 components each with 10% defective components. We select at random, one of the boxes and remove a single component from it. What a the probability that the component is defective.

Solution: Let $B_i$ be the event that denotes the selection of the $i^{th}$ box and let $A$ be the event that the selected component is defective. Then the required probability is $P(A)$.

$$\begin{aligned}
P(A) &= P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3) + P(A \cap B_4) \\
&= P(B_1)\,P(A \mid B_1) + P(B_2)\,P(A \mid B_2) + P(B_3)\,P(A \mid B_3) + P(B_4)\,P(A \mid B_4) \\
&= \frac{1}{4} \times \frac{5}{100} + \frac{1}{4} \times \frac{40}{100} + \frac{1}{4} \times \frac{10}{100} + \frac{1}{4} \times \frac{10}{100} \\
&= \frac{65}{400} \\
P(A) &= \frac{13}{80}.
\end{aligned}$$

### 1.6.1 *Law of total probability*

The sample space $\Omega$ may be partitioned into $k$ disjoint sets (events), namely $E_i$ where $i = 1, 2, ..., k$. The probability of a certain event $A \subset \Omega$ can then be computed by the weighted sum of the conditional probabilities, $P(A|E_i)$, where the weights are given by the probability of the partitioning events $P(E_i)$. This is the *law of total probability*, stated succinctly as follows.

$$P(A) = \sum_{i=1}^{k} P(A|E_i)P(E_i). \tag{1.3}$$

### 1.6.2 *Bayes' theorem*

Bayes' theorem helps us to compute *posterior* probability $P(A|B)$ by using the concepts of conditional probability and the law of total probability as follows.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}. \tag{1.4}$$

$P(A)$ and $P(B)$ are known as *prior* probabilities. The prior probabilities may be computed using the law of total probability. The formula in equation 1.4 can be deduced by using the definition of conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$.

Bayesian statistics is a model for capturing epistemological uncertainty within the framework of probability. The prior probabilities constitute our original belief sets which are conditioned (over time) by the diversity of our experiences (data) and manifest as posterior probabilities. These posterior probabilities constitute our refined and conditioned beliefs that form the basis of inferential decisions.[15] Let us understand the essence of this framework through a simple example.

### 1.6.3 *Example: Bayes' theorem and law of total probability*

A factory unit uses three automatic bolt threading machines (rollers), each accounting for 20%, 30%, and 50% of the factory output of ready-to-use bolts for the aerospace industry. The precision rating (number of non-defective parts produced per one hundred) of each of the rollers is 95%, 97%, and 99% respectively. If a part is picked up

[15] To put this in context of our formalism, $P(A)$ constitutes our original belief sets, and the posterior probabilities $P(A|B)$ are the updated beliefs that are attained by the process of conditioning over experiences and data (represented here by the event(s) $B$). This update is made possible through the *likelihood* model $P(B|A)$. Note: A detailed discussion on the likelihood function, used in estimation theory, is beyond the scope of this text.

at random from the production line and found to be defective, what is the probability that it was produced by the second machine?

Solution: Let us begin by defining the relevant events: $A_i$ is the event that a randomly picked bolt is manufactured by the $i^{th}$ machine, $i = 1, 2, 3$; $B$ is the event that a randomly chosen part is defective. Based on the information provided, we glean that the prior probabilities are $P(A_1) = 0.2$, $P(A_2) = 0.3$, $P(A_3) = 0.5$. Further, $P(B|A_1) = 0.05$, $P(B|A_2) = 0.03$, $P(B|A_3) = 0.01$. We are asked to find $P(A_2|B)$. Using Bayes' theorem,

$$\begin{aligned} P(A_2|B) = \frac{P(B|A_2)P(A_2)}{P(B)} &= \frac{(0.03)(0.3)}{\sum_{i=1}^{3} P(B|A_i)P(A_i)} \\ &= \frac{0.009}{(0.05)(0.2) + (0.03)(0.3) + (0.01)(0.5)} \\ &= \frac{9}{24} = \boxed{0.3750} \end{aligned} \tag{1.5}$$

Figure 1.11: A quality control engineer who understands the nuances of Bayesian statistics and its implications.

Albeit, as a toy example above, we have considered the case of a small factory that has only three operational rolling machines. In a more realistic setting, we may expect that the factory quality control engineer may have to deal with a large pool of machines producing bolts *en masse*. Her prior belief set may hint to her that there is a 30% chance this defective bolt came from the second machine because the second machine has a production rate of 30% of the total output. However, the extra information gleaned from randomly picking a part and noticing it to be defective has led to an update in her belief system that is manifested in terms of the posterior probability $P(A_2|B) = 0.375$. The update from 30% to 37.5% is a significant change of 25% that is likely to draw a greater attention of the engineer to the operational fitness of the second machine. Bayesian inference, thus, enables an enhancement of predictive knowledge of a phenomenon by synthesizing information and data from experiences.

### *1.6.4 Example: Diagnosis of disease*

The chance a doctor $D$ will diagnose a disease $X$ correctly is 60%. The chance that a patient will die by his treatment after correct diagnosis is 40%, and the chance of death by wrong diagnosis is 70%. A patient of doctor $D$ who had disease $X$ died. What is the chance that his disease was diagnosed correctly?

Solution: Let $B_1$ be the event that disease $X$ is diagnosed correctly by doctor $D$. Let $B_2$ be the event that disease $X$ is not diagnosed correctly by doctor $D$.

Let $A$ be the event that patient dies who had disease $X$.

$$\begin{aligned} P(B_1 \mid A) &= \frac{P(B_1)\,P(A \mid B_1)}{P(B_1)\,P(A \mid B_1) + P(B_2)\,P(A \mid B_2)} \\ &= \frac{\frac{60}{100} \times \frac{40}{100}}{\frac{60}{100} \times \frac{40}{100} + \frac{40}{100} \times \frac{70}{100}} \\ &= \frac{24}{24+28} \\ &= \frac{6}{13}. \end{aligned}$$

### *1.6.5 Example: Academic leadership and curriculum matters*

In late 2022, there are three candidates for the position of director at a University. Mr. $X$, Mr. $Y$ and Mr. $Z$ have chances of getting appointed to the post in the ratio $4:2:3$. If Mr. $X$ is selected, he could introduce a new syllabus in the university with a probability 0.3. The probabilities for Mr. $Y$ and Mr. $Z$ doing the same, if selected, are 0.5 and 0.8, respectively.

1. What is the probability that there will be a new syllabus in 2023?
2. If there is a new syllabus in 2023, what is the probability that Mr. $Z$ is the newly appointed director?

Solution: Let $A$ be the event that a new syllabus is introduced in 2023.
Let $X$, $Y$ and $Z$ be the event that Mr. $X$, Mr. $Y$ and Mr. $Z$ is the director respectively.

1.
$$\begin{aligned} P(A) &= P(X)P(A \mid X) + P(Y)P(A \mid Y) + P(Z)P(A \mid Z) \\ &= \frac{4}{9} \times 0.3 + \frac{2}{9} \times 0.5 + \frac{3}{9} \times 0.8 \\ &= \frac{12+10+24}{90} \\ &= \frac{46}{90} \\ &= \frac{23}{45}. \end{aligned}$$

2.
$$\begin{aligned} P(Z \mid A) &= \frac{P(Z)P(A \mid Z)}{P(A)} \\ &= \frac{\frac{3}{9} \times 0.8}{\frac{46}{90}} \\ &= \frac{24}{46} \\ &= \frac{12}{23}. \end{aligned}$$

## 1.7 Chapter project: Random walk on a lonely island

### 1.7.1 Interlude: Analytical calculations and computer simulations to predict the fate of Squeaky

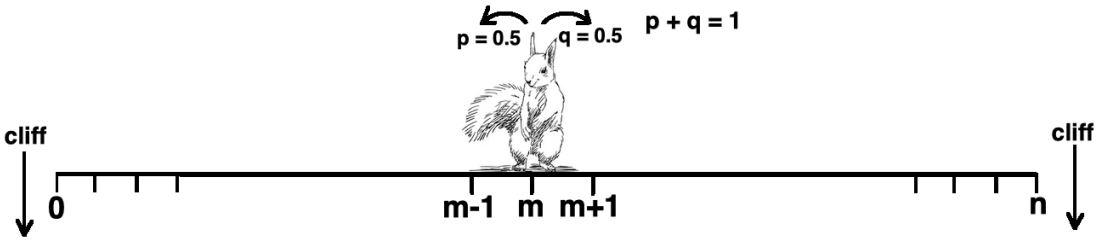


Figure 1.12: Schematic portrait of Squeaky's hopping adventure on the one dimensional island hill. At any given instance, she jumps to the left with probability 1/2 and jumps to the right with probability 1/2.

**Consider the schematic diagram of Squeaky's hopping adventure on the one dimensional island hill as shown in Figure 1.12. In order to make the calculations tractable, we may consider dividing the island into discrete grid points that can host Squeaky. The grid points run from location $0$ to location $n$. At a certain time, let us consider that Squeaky is at location $m$ and she makes a choice to jump to her left to location $(m+1)$ with probability $q = 0.5$ and to jump to her right with probability $p = 0.5$. The probabilities $p$ and $q$ are assigned the value $0.5$ because we have assumed that she does not have any inherent bias or preference in choosing between left and right-sided moves. In this example, we will take her decision instances to jump either to the left or right as the time stamps, i.e. in any given time point so defined, she does not decide to stay where she is.**

**Before we attempt the theoretical calculations to predict her fate, let us consider the case phenomenologically with the help of a spanning diagram as shown in Figure 1.13. Since it is quite obvious from this diagram that the decision paths span the entire breadth of the one-dimensional island, our hunch is Squeaky will make it all the way to the edge and trip. Let us see if the calculations below validate our intuition.**

**Let us define the following events. $W$ is the event that Squeaky falls in the pit to our left. We would like to compute the following probability.**

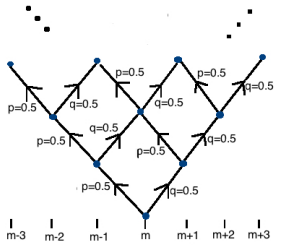


Figure 1.13: Spanning tree showing the possible decision paths that Squeaky could opt for starting from the location $m$.

$$P_m = P_m(\textbf{left pit}) = \textbf{Probability of the event } W \textbf{ when Squeaky starts at } X_0 = m, \qquad (1.6)$$

**with $P_0 = 1$, $P_n = 0$ for obvious reasons. Further, let $E$ be the event that the first hop is to the left. We will use the law of total probability and condition our computation upon this event as follows:**

$$
\begin{aligned}
P_m &= P(W \text{ and } E|X_0 = m) + P(W \text{ and } E^c|X_0 = m) \\
&\quad \nearrow \\
&E \text{ \& } E^c \text{ partition the choice space} \\
&= P(W|E \text{ and } X_0 = m)P(E|X_0 = m) \\
&\quad \nearrow \\
&\text{law of total probability} \\
&\qquad + \; P(W|E^c \text{ and } X_0 = m)P(E^c|X_0 = m) \\
&= P(W|X_1 = m-1) \times \frac{1}{2} + P(W|X_1 = m+1) \times \frac{1}{2} \\
&= \frac{1}{2}P(W|X_0 = m-1) + \frac{1}{2}P(W|X_0 = m+1) \\
&\quad \nearrow \\
&\text{independent hops} \\
P_m &= \frac{1}{2}P_{m-1} + \frac{1}{2}P_{m+1}. \qquad (1.7)
\end{aligned}
$$

**Equation 1.7 is a *recurrence* relation whose solution can be readily computed.**

**Questions:**

1. **Solve the recurrence relation 1.7 for $P_m = P_m(\text{left pit})$.**

2. **Use an argument based on symmetry to deduce the solution for $P'_m = P_m(\text{right pit})$.**

3. **Compute $P_m + P'_m$ and thereafter comment on the fate of Squeaky based on your theoretical calculations.**

4. **Build a computer simulation of Squeaky's exploration on the island hill and comment whether the results of the simulation corroborate with your theoretical calculations about Squeaky's fate. In order to develop the simulation, you may refer to the pseudo-code provided below and turn it into a computer executable code using a programming language of your choice.**

**Software Implementation**

**Pseudocode of the random walk algorithm:**

```
INPUT: grid_length, start_pos.

initialise curr_pos = start_pos;
initialise num_hops = 1;
while (curr_pos > 0 && curr_pos < grid_length)
   toss = rand(1);
   if (toss < 0.5)
      curr_pos = curr_pos - 1;
   elseif (toss >= 0.5)
      curr_pos = curr_pos + 1;
   end
   plot curr_pos and record graphic frame;
   num_hops = num_hops + 1;
end

OUTPUT: num_hops, play recorded animation.
```

**In case you prefer to use Matlab, some useful commands to build your code may be: `rand, stem, getframe, movie`.**

**We will return to the random walk expedition, undertaken by Squeaky, later in this chapter. Let us now get introduced to some new concepts on computing statistical averages.**

## *1.8 Expected values of random variables*

The most common entity of interest while conducting an experiment is perhaps its *average* output. Since we are largely interested in statistical experiments, we may expect the average output in terms of a statistical average due to the random or stochastic nature of the underlying process/model. What this means is the following: *if we were to repeat the same experiment many times over, each time recording the output of the model (or process), and subsequently take the average of all the recorded outputs; then, this ensemble average may be regarded as the mean behavior of the experimental model (or process).*

Figure 1.14: Let's flip some coins or do some math? Oh well, the law of large numbers will prevail!

e.g., Consider the case when we toss a fair coin. We may ask what the *expected* outcome is. In other words, if we toss the coin many times over, what is the mean outcome? Let us define the Bernoulli random variable associated with this experiment as has been suggested earlier: $X(H) = 1$ and $X(T) = 0$, each has a chance of 0.5 as an outcome of any given toss. We may compute the simple average of all the outcomes, i.e. if we tossed the coin a 100 times and if sixty two of those were heads, then the average is $62/100 = 0.62$. This would be perfectly fine as an estimate of average output

(or average behavior) if we had a lot of data (of the outcomes of the tosses). This may not be always readily available. In such a scenario, we may consider the weighted sum of all the possible outcomes (in this case 1 and 0), where the weights are the respective probabilities of individual outcomes (Prob(heads) = Prob(tails) = 0.5). So the expected value of this experiment is given by $E(X) = \frac{1}{2} \times 1 + \frac{1}{2} \times 0 = 0.5$.

In fact, had we conducted our coin tossing experiment many more times than one hundred (as was done above), then the average would be observed to converge to the value 0.5 with the increasing number of tosses. As a matter of fact, we have just stated a very important result of probability theory known as the *law of large numbers*.[16] We will revisit this law in greater detail in the next chapter after we introduce the notion of probability distributions.

[16] There are two variants of this law, viz., the *weak law of large numbers* and the *strong law of large numbers* depending on the nature of convergence of the sample mean to the expected value.

It is also essential to state that the expected value of a statistical experiment may take on a value that is not equal to the elements of the range of the random variable. In the above example, $X$ takes on values 0 and 1 but $E(X) = 0.5$ which does not belong to the range of $X$. This observation must be noted in conjunction with our comment earlier in section 1.3.3 that statistics deals with idealized averages that are far removed from individual experiences (outcomes).

The expected value of a random variable may thus be generalized as follows.

$$\mu_X = E(X) = \sum_{x \in \text{range}(X)} xP(X = x). \tag{1.8}$$

When it is understood, we will simply write $\mu$ and omit the subscript used to denote the relevant random variable. The formulation in equation 1.8 is simply a generalization of the explanation in the preceding paragraph where $\text{range}(X) = \{1, 0\}$. The expected value is the first statistical moment and the variation in the outcomes is given by the variance which is the second statistical moment and is defined as follows.

$$\sigma^2 = Var(X) = E\big((X - \mu)^2\big) = \sum_{x \in \text{range}(X)} (x - \mu)^2 P(X = x). \tag{1.9}$$

The variance is related to the expectation in yet another useful manner.

$$Var(X) = E(X^2) - \mu_X^2. \tag{1.10}$$

We will revisit the calculations of expectation and variance again with more rigor in the next chapter on probability distributions. Here we will simply state some useful results and study a few examples.

1. $E(cX) = cE(X)$, where $c$ is a constant.
2. $E(X + c) = E(X) + c$, where $c$ is again a constant.
3. $E(X + Y) = E(X) + E(Y)$.
4. $Var(cX) = c^2 Var(X)$, where $c$ is a constant.
5. $Var(X + c) = Var(X) + 0 = Var(X)$, where $c$ is a constant.
6. $Var(aX \pm bY) = a^2 Var(X) + b^2 Var(Y) \pm 2abCov(X, Y)$, where $a, b$ are constants. $Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$ is the *covariance* of $X$ and $Y$.

### 1.8.1 Example: Expectation of an indicator random variable

A very useful result we will employ in a subsequent example is the expected value of an indicator random variable defined in equation 1.1. By definition 1.8,

$$E(\mathbb{1}_A) = 1 \times P(A) + 0 \times P(A^c) = P(A).$$

### 1.8.2 Example: Success rate in an infinite series of independent trials

What is the expectation and variance of the number of failures preceding the first success in an infinite series of independent trials with probability $p$ of success in each trial?

Solution: $\Omega = \{S, FS, FFS, FFFS, FFFFS, \ldots\}$, where $S$=Success and $F$=Failure
Let X be the number of failures.

$$E(X) = \sum_{x=0}^{\infty} xP\{X = x\}$$

$$\begin{aligned} E(X) &= (1-p)p + 2(1-p)^2p + 3 \times (1-p)^3p + \cdots \\ (1-p)E(X) &= \ (1-p)^2p + 2 \times (1-p)^3p + \cdots \end{aligned}$$

We subtract the last two equations and get:

$$\begin{aligned} (1-1+p)E(X) &= (1-p)p + (1-p)^2p + (1-p)^3p + \\ pE(X) &= \frac{p(1-p)}{1-(1-p)} \\ pE(X) &= 1-p \\ E(X) &= \frac{1-p}{p} \end{aligned}$$

Now we calculate the variance,

$$Var(X) = E(X^2) - (E(X))^2$$

$$E(X^2) = \sum_{x=0}^{\infty} x^2P\{X = x\}$$

$$\begin{aligned} E(X^2) &= (1-p)p + 2^2(1-p)^2p + 3^2(1-p)^3p + 4^2(1-p)^4p + \cdots \\ E(X^2) &= (1-p)p + 4(1-p)^2p + 9(1-p)^3p + 16(1-p)^4p + \cdots \\ (1-p)E(X^2) &= \ (1-p)^2p + 4(1-p)^3p + 9(1-p)^4p + \cdots \end{aligned}$$

Subtract the last two equations and get:

$$\begin{aligned} (1-1+p)E(X^2) &= (1-p)p + 3(1-p)^2p + 5(1-p)^3p + 7(1-p)^3p + \cdots \\ pE(X^2) &= (1-p)p + 3(1-p)^2p + 5(1-p)^3p + 7(1-p)^3p + \cdots \\ (1-p)pE(X^2) &= \ (1-p)^2p + 3(1-p)^3p + 5(1-p)^4p + \cdots \end{aligned}$$

We subtract the last two equations and get:

$$
\begin{aligned}
(p - p + p^2)E(X^2) &= (1-p)p + 2\left[(1-p)^2 p + (1-p)^3 p + (1-p)^3 p + \cdots\right] \\
p^2 E(X^2) &= (1-p)p + 2\left[\frac{(1-p)^2 p}{1-(1-p)}\right] \\
p^2 E(X^2) &= (1-p)p + 2(1-p)^2 \\
p^2 E(X^2) &= (1-p)(p + 2 - 2p) \\
p^2 E(X^2) &= (1-p)(2-p) \\
E(X^2) &= \frac{(1-p)(2-p)}{p^2}
\end{aligned}
$$

Finally, we obtain the variance:

$$
\begin{aligned}
Var(X) =& \frac{(1-p)(2-p)}{p^2} - \left[\frac{1-p}{p}\right]^2 \\
Var(X) =& \frac{(1-p)(2-p)}{p^2} - \frac{(1-p)^2}{p^2} \\
Var(X) =& \frac{(1-p)\left[(2-p) - (1-p)\right]}{p^2} \\
Var(X) =& \frac{1-p}{p^2}.
\end{aligned}
$$

### 1.8.3 *Example: Expectation and variance of a mean subtracted normalized random variable*

Suppose $X$ is a random variable which takes the values $1, 2, 3$ and $4$ with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}$ and $\frac{1}{8}$ respectively. Let a new random variable $Y$ be defined as $Y = \frac{X-\mu_X}{\sigma_X}$, where $\mu_X$ is the mean and $\sigma_X^2$ is the variance of $X$. Use the properties of expectation to find the expectation and variance of $Y$. Is this true for all random variables $X$?

Solution: Take $\mu_X = \mu$ and $\sigma_X = \sigma$.

$$
\begin{aligned}
E(Y) &= E\left(\frac{X-\mu}{\sigma}\right) \\
&= E\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) \\
&= \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} \\
&= \frac{\mu}{\sigma} - \frac{\mu}{\sigma} \\
&= 0
\end{aligned}
$$

Since $Var(c) = 0$ and $Cov(X, c) = 0$ for all constants $c$;

$$\begin{aligned} Var(Y) &= Var\left(\frac{X-\mu}{\sigma}\right) \\ &= Var\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) \\ &= \frac{1}{\sigma^2} Var(X) + 0 \\ &= \frac{\sigma^2}{\sigma^2} \\ &= 1. \end{aligned}$$

Hence this result is true for all $X$.

### *1.8.4 Example: Variance of sum of two random variables*

Prove that $V(aX \pm bY) = a^2V(X) + b^2V(X) \pm 2abCov(X, Y)$, where $Cov(X, Y) = E(X - \mu_X)(Y - \mu_Y)$.

Solution: Note that $E(aX \pm bY) = aE(X) \pm bE(Y) = a\mu_X \pm b\mu_Y$.

$$\begin{aligned} Var(aX \pm bY) &= E\left(aX \pm bY - (a\mu_X \pm b\mu_Y)\right)^2 \\ &= E\left(a(X - \mu_X) \pm b(Y - \mu_Y)\right)^2 \\ &= E\left(a^2(X-\mu_X)^2 + b^2(Y-\mu_Y)^2 \pm 2ab(X-\mu_X)(Y-\mu_Y)\right) \\ &= a^2E(X-\mu_X)^2 + b^2E(Y-\mu_Y)^2 \pm 2abE(X-\mu_X)(Y-\mu_Y) \\ &= a^2Var(X) + b^2Var(X) \pm 2abCov(X, Y). \end{aligned}$$

Figure 1.15: Homer's probability of getting hired is $\frac{1}{513}$ because he is the $513^{th}$ candidate. How he wishes he had applied earlier! Had he been the first candidate to be interviewed, he would have most certainly been recruited because $E(\mathbb{1}_1) = p_1 = 1$. Moral of the story: *Act fast, do not procrastinate!*

### *1.8.5 Example: Expected number of new recruits per n hiring interviews*

Let us consider that a hiring manager has the responsibility of conducting interviews of $n$ candidates for the post of a peon over a certain period of time. The candidates appear for the interviews in a random fashion, i.e. from the perspective of the hiring manager; prior to the interview, there is an equal probability among candidates to be the most suitable candidate. The hires are made on a rolling basis in the sense that whenever he encounters a better candidate than the existing one, he hires that person and keeps him in the job until a better candidate is found. How many hires are made in this process? Can we give an estimate of the cost associated with this firing-recruiting process?

Consider an indicator random variable $\mathbb{1}_i = \begin{cases} 1; & \text{when the } i^{th} \text{ candidate is hired,} \\ 0; & \text{when the } i^{th} \text{ candidate is not hired.} \end{cases}$

For the $i^{th}$ candidate to be hired, the preceding $(i - 1)$ candidates must not have been better than this candidate. But since each of these $i$ candidates had an equal chance to

be hired, the probability that the $i^{th}$ candidate is hired is $p_i = \frac{1}{i}$. Thus, the total number of hires is given by

$$X = \sum_{i=1}^{n} \mathbb{1}_i. \tag{1.11}$$

We compute the expected value of $X$, and by linearity of expectation, we have

$$\begin{aligned} E(X) = E(\sum_{i=1}^{n} \mathbb{1}_i) = \sum_{i=1}^{n} E(\mathbb{1}_i) &= \sum_{i=1}^{n} p_i \\ \text{because } E(\mathbb{1}_A) = P(A) \text{ as per sec. 1.8.1} & \\ &= \sum_{i=1}^{n} \frac{1}{i} \\ &\to \log n + \mathcal{O}(1), \text{ as } n \to \infty. \end{aligned} \tag{1.12}$$

Euler–Mascheroni result

This means that for every $n$ interviews conducted by the hiring manager, approximately $\log n$ of them get hired on an average. The cost of the recruitment process is $\mathcal{O}(c_H \log n)$ where $c_H$ is the hiring cost factor.

### *1.8.6 Example: analysis of sorting algorithm*

One of the essential features while analyzing the cost of sorting an array is the number of existing inversions[17] in the array. We will denote an inversion by $\Im$. Let us define an indicator random variable $\mathbb{1}_{A[i]>A[j]}$ when $1 \leq i < j \leq n$ to analyze the average number of inversions in an array of length $n$. Let $X$ denote the total number of inversions in the array, $X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbb{1}_{A[i]>A[j]}$. Further, $P(\Im) = \frac{1}{2}$ because given any two distinct random numbers, the probability that one is bigger than the other is half. This entails $E(\mathbb{1}_{A[i]>A[j]}) = P(\Im) = 1/2$ for all $i < j$. Therefore,

$$\begin{aligned} E(X) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} = \frac{1}{2} \sum_{i=1}^{n-1} (n-i) = \frac{1}{2} \sum_{i=1}^{n-1} n - \frac{1}{2} \sum_{i=1}^{n-1} i &= \frac{1}{2}\big(n(n-1) - \frac{n(n-1)}{2}\big) \\ &= \frac{n(n-1)}{4}. \end{aligned} \tag{1.13}$$

The expected number of inversions in an array of size $n$ is $\mathcal{O}(n^2)$.

[17] An *inversion* in an array between a pair of entries is a condition when $i < j$ but $A[i] > A[j]$.

### *1.8.7 Law of total expectation*

Just like we could compute the probability of an event by conditioning over the partitioning events $E_i$ and calculating the weighted sum, we can perform a very similar calculation for computing the expected value. This is known as the *law of total expectation*.

$$E(\mathbb{1}_A) = \sum_{i=1}^{k} E(\mathbb{1}_A|\mathbb{1}_{E_i})P(E_i). \tag{1.14}$$

We will use the law of total expectation to estimate the life expectancy of Squeaky in terms of the average number of hops till the end.

A generalization of the law of total expectation is known as the *law of iterated expectations* (LIE).[18]

$$E(X) = E(E(X|Y)) = \sum_{y} E(X|Y=y)P(Y=y). \tag{1.15}$$

Here $X$ and $Y$ belong to the same probability space. $E(E(X|Y))$ must be understood as $E_Y(E_X(X|Y))$ to make the order of the expectation operator with respect to the random variables, $X$ and $Y$, explicitly clear.

[18] The law of iterated expectation implies the law of total probability as follows. Consider an event $A$ and a random variable $Y$. Then,

$$\begin{aligned} P(A) &= E(\mathbb{1}_A) \\ &= E(E(\mathbb{1}_A|Y)) \end{aligned}$$

law of iterated expectation

$$\begin{aligned} &= \sum_y E(\mathbb{1}_A|Y=y)P(Y=y) \\ P(A) &= \sum_y P(A|Y=y)P(Y=y) \end{aligned}$$

law of total probability

.

### 1.8.8 *Example: average number of mangoes eaten per week in a population of engineers*

Let us consider a dietary survey of 1000 engineers working in a factory. The population of the engineers has males and females. The number of young males is 300 and the number of old males is 500. The corresponding figures for female engineers are 50 and 150. The survey reveals that among the males, the younger folks eat 4 mangoes per week while the older folks eat 6 mangoes per week.[19] The corresponding figures for the women engineers are 0 and 4 per week. The question is to find the average number of mangoes eaten per week by any person from the whole population. Let $M$ be the number of mangoes eaten by a factory engineer per week. We proceed by first calculating the chance of encountering a male engineer, $p_m = \frac{800}{1000}$. Likewise, the corresponding estimate for a female engineer is $p_f = \frac{200}{1000}$. Here the subscripts $m$ and $f$ refer to males and females respectively, and the subscripts $y$ and $o$ refer to young and old respectively.[20] Further, $p_{y|m} = \frac{300}{300+500} = 3/8$. $p_{o|m} = 1 - p_{y|m} = 5/8$. Therefore, computing the expected value of mangoes eaten per week by male and female engineers as a weighted sum of the mangoes eaten by the respective age-groups,[21] we have the following estimates.

$$E(M|m) = M_{y|m} \times p_{y|m} + M_{o|m} \times p_{o|m} = 4 \times \frac{3}{8} + 6 \times \frac{5}{8} = \frac{21}{4} \tag{1.16}$$

for the male engineers. Similarly,

$$E(M|f) = M_{y|f} \times p_{y|f} + M_{o|f} \times p_{o|f} = 0 \times \frac{1}{4} + 4 \times \frac{3}{4} = 3 \tag{1.17}$$

for the female engineers.

Now, using the law of iterated expectation, we have

$$E(M) = E(E(M|gender)) = E(M|m)p_m + E(M|f)p_f = \frac{21}{4}\frac{800}{1000} + 3\frac{200}{1000} = 4.8. \quad (1.18)$$

So, on an average, the number of mangoes eaten by an engineer from the whole population in the factory is 4.8 per week.

[21] For simplicity, let us assume that within the same age group, the answers in the survey are consistent and identical.
[21] This is one of the very rare instances in the book where we have used both uppercase ($M$) and lowercase letters ($y, o, m, f$) to denote random variables.
[21] **cf.** explanation given in the introductory paragraphs of section 1.8

Figure 1.16: Should the mango vendor bring more mangoes to the factory during the lunch breaks?
[22] There is a similar law of total covariance which we will not discuss in this text.

### 1.8.9 *Law of total variance*

The *law of total variance*[22] is another useful result that we will simply state here as follows.

$$Var(Y) = E(Var(Y|X)) + Var(E(Y|X)). \quad (1.19)$$

## 1.9 *Chapter project: Random walk on a lonely island*

### 1.9.1 *Epilogue: Life expectancy of Squeaky*

**Given that we have predicted Squeaky's fate, our next question of interest is: *what is her life expectancy in terms of the number of hops from the beginning till the end?* Let $D$ be the number of hops till the end. We will use the law of total expectation and once again condition upon the event $E$ as follows:**

$$\begin{aligned}
E_m &= E(D|X_0 = m) \\
&= E(D|E \text{ and } X_0 = m)P(E|X_0 = m) + E(D|E^c \text{ and } X_0 = m)P(E^c|X_0 = m) \\
&= \frac{1}{2}E(D|X_1 = m-1) + \frac{1}{2}E(D|X_1 = m+1) \\
&\overset{\text{reset chain}}{=} \frac{1}{2}\Big(1 + E(D|X_0 = m-1)\Big) + \frac{1}{2}\Big(1 + E(D|X_0 = m+1)\Big) \\
E_m &= 1 + \frac{1}{2}E_{m-1} + \frac{1}{2}E_{m+1}. \quad (1.20)
\end{aligned}$$

**Equation 1.20 is a non-homogeneous recurrence relation.**

**Questions:**

1. **Explain the appearance of the numeral 1 in the recurrence relation 1.20.**
2. **Solve the non-homogeneous recurrence relation 1.20 to estimate the life expectancy of Squeaky $E_m$.**
3. **Find the starting location of Squeaky's hopping expedition to maximize her life span.**
4. **Compare the estimate of $E_m$ from the computer simulation you developed earlier in the chapter with the theoretical estimate of $E_m$ above. Comment on the origin of any discrepancy you observe in the comparison.**

## 1.10 Selected bibliography


1. *Foundations of the Theory of Probability* by *Andrey N. Kolmogorov*, Chelsea Publishing Co. (second English translation), 1956.
2. *A First Look at Rigorous Probability Theory* by *Jeffrey S. Rosenthal*, World Scientific (second edition), 2006.
3. *An Introduction to Probability Theory and Its Applications: Vol. 1* by *William Feller*, John Wiley & Sons, Inc. (third edition), 1968.
4. *Applied Statistics and Probability for Engineers* by *Douglas C. Montgomery* and *George C. Runger*, Wiley (sixth edition), 2014.
5. *Fundamentals of Mathematical Statistics* by *S. C. Gupta* and *V. K. Gupta*, Sultan Chand and Sons (eleventh edition), 2017.
6. *Introduction to Probability and Statistics for Engineers and Scientists* by *Sheldon M. Ross*, Academic Press (Elsevier) (sixth edition), 2021.
7. *Probability Theory* by *Alfred Renyi*, Dover Publications Inc. (Illustrated edition), 2007.
8. *Probability and Random Processes with Applications to Signal Processing* by *Henry Stark* and *John W. Woods*, Pearson Education (fourth edition), 2011.


## 1.11 Exercise problems

1. (***Combinatorics***) Out of a population of 10 digits running from 0 through 9, what is the probability that five consecutive random digits are all different?

2. (***Occupancy problems: Bose-Einstein, Fermi-Dirac, and Maxwell-Boltzmann statistics***) This model relates to placing randomly $r$ indistinguishable balls (particles) into $n$ cells (quantum states). Consider the *occupancy numbers* $r_1, r_2, ..., r_n$[23] satisfying $\sum_{i=1}^{n} r_i = r$. Two distributions of the balls are distinguishable only if the $n$ tuples $(r_1, r_2, ..., r_n)$ are not identical.

   [23] Occupancy number $r_k$ stands for the number of balls in the $k^{th}$ cell.

   2.I Bose-Einstein statistics: This is a model for photons, nuclei, and atoms containing an even number of particles.

   (a) Find an expression for the number of distinguishable distributions, $A_{r,n}$.
   (b) Find an expression for the number of distinguishable distributions in which no cells are empty.
   (c) What is the probability of each distribution in (a)?
   (d) Given $n = 5$ quantum states (cells) and $r = 3$ indistinguishable particles, what is the probability of the distribution $(*|\_|*|*|\_)$ in Bose-Einstein statistics? Here $*$ represents a particle, an empty orbital is denoted by _ , and the barrier between two successive orbitals is denoted by the symbol $|$.

   2.II Fermi-Dirac statistics: This is a model for electrons, neutrons, and protons. This model assumes (i) it is impossible for two or more identical particles to be in the same quantum state[24], i.e. $r \leq n$; and (ii) all distinguishable distributions have equal probabilities.

   [24] Pauli's exclusion principle.

(a) How many such distributions are possible?

(b) What is the probability of the distribution $(*|\_|*|*|\_)$ in Fermi-Dirac statistics?

2.III Maxwell-Boltzmann statistics: This is a model for material particles distributed over various energy states in thermal equilibrium in classical mechanics. This model does not apply to quantum particles. In this model, the number of ways we can place $r$ distinguishable particles in $n$ cells is certainly $\underbrace{n \times n \times \cdots \times n}_{r \text{ times}} = n^r$ when sampling with replacement is permissible.[25]

(a) Find the number of ways in which a population of $r$ particles can be partitioned into $n$ cells such that $r_1 + r_2 + \cdots r_n = r$.

(b) Given occupancy numbers $r_1, ..., r_n$, what is the probability of this distribution in Maxwell-Boltzmann statistics?

(c) What is the probability of the distribution $(*|\_|*|*|\_)$ in Maxwell-Boltzmann statistics?

3. (***Quadratic equation with stochastic coefficients***) Each coefficient of a quadratic equation $ax^2 + bx + c = 0$ is determined by the throw of a regular die. Find the probability that the equation will have at least one real root?

4. (***Medical diagnosis of prostate cancer***) Prostate cancer is the most common type of cancer found in males. As an indicator of whether a male has prostate cancer, doctors often perform a test that measures the level of the PSA protein (prostate specific antigen) that is produced only by the prostate gland. The test is highly unreliable even though there is a strong correlation between high PSA value and incidence of cancer. Indeed, the probability that a non-cancerous man will have an elevated PSA level is approximately 0.115, with this probability increasing to approximately 0.273 if the man does have cancer. If, based on other factors, a physician is 81 percent certain that a male has prostate cancer, what is the conditional probability that he has the cancer given that

(a) the test indicates a high PSA level;

(b) the test does not indicate a high PSA value?

Re-estimate your probabilities if the physician initially believes there is a 29 percent chance the man has prostate cancer.

5. (***Spacecraft control system***)[26] The flight control computer on a spacecraft employs four independent flight computers that work in parallel - a much required redundancy built into the system. During every critical flight path decision, the computers "vote" to decide on the most important step. e.g., in case of a "roll" decision, the probability that a computer will make an error is 0.0002. Let $X$ denote the number of computers that vote for an erroneous roll movement. Compute $E(X)$ and $Var(X)$.

6. (***A game of dart***) In a game of dart, a participant is given three attempts to hit a target. On each try, she either scores a hit, $H$, or a miss, $M$. The game requires that the player must alternate which hand she uses in successive attempts. That is, if she makes her first attempt with her right hand, she must use her left hand for the second attempt and her right hand for the third. Her chance of scoring a hit with her right hand is 0.7 and

[25] To illustrate this further, consider that the $n$ cells are in a *bag*, we randomly select a cell from this bag and place one of the $r$ particles in it. Consequently, we put this cell back in the bag and continue sampling, accounting for the fact that the same cell (as was chosen from the bag before) may be sampled again to be filled by yet another particle. This is akin to sampling with replacement. This way of placing $r$ particles in $n$ cells (with replacement) is similar to throwing an $n$-sided dice $r$ times.

[26] *Architecture of the Space Shuttle Primary Avionics Software System* by *Gene D. Carlow*, Communications of the ACM, **27** (9), 1984.

with her left hand is 0.4. Assume that the results of successive attempts are independent and that she wins the game if she scores at least two hits in a row. If she makes her first attempt with her right hand, what is the probability that she wins the game?

7. (***Random sums***) Let $X_1, X_2, \ldots$ be i.i.d.[27] random variables and let $E(X_1) = \mu_X$. Let $N$ be a non-negative integer valued random variable that is independent of the sequence of $X$s and let $E(N) = \mu_N$. Define the random sum $S$ to be $S = X_1 + X_2 + \ldots + X_N$ where $S = 0$ if $N = 0$.[28] What is $E(S)$?
*Hint: Use the law of iterated expectation, i.e. condition your computation on $N = n$.*

[27] independent and identically distributed random variables

[28] Notice that $S$ is the sum of a random number of terms.

8. (***Population growth model with random progeny and no deaths***) Consider a population of tumor cells where each cell has a random number of progeny. Consider that the number of progeny of the proliferating cells is i.i.d. with mean $\mu$. Suppose the process starts with one cell in generation zero. For simplicity let us assume there are no deaths. What is the expected total number of tumor cells in $n$ generations? For $n \to \infty$, find a condition for arresting the rate of growth of tumor cells.
*Hint: In the formulation of the question above, consider $T_k \equiv S =$ number of cells in generation $k$. $T_k = X_1 + X_2 + ... + X_{T_{k-1}}$. Here $X_1, X_2$, etc.... are the numbers of progeny of the first, second, etc. ... cells in generation $k-1$. This gives $E(T_k) = \mu E(T_{k-1})$. The final answer is $\sum_{k=0}^{n} E(T_k)$.*

9. (***Matching probability***) Consider $n$ letters that are designated for $n$ envelopes. However, the letters and envelopes are not in order (they are randomly arranged) and hence the letters may not go in the correct envelop. Let $A_k$ denote an event when a match occurs in the $k^{th}$ place. What is the probability of the event $A_k$? What happens as $n \to \infty$?

10. (***Ordering of events and their probability***) Consider a probability measure $P$ defined on an appropriate probability space $(\Omega, \mathfrak{F}, P)$. Let $E_1, E_2$ be events from the event space $\mathfrak{F}$ such that $E_1 \subset E_2$. Deduce a relationship between $P(E_1)$ and $P(E_2)$.

□