Introduction to Probability: Lecture Notes

Dr Rachel Quinlan
School of Mathematical and Statistical Sciences
NUI Galway

Contents

Chapter 1

Probability Spaces

1.1 What is Probability?

Probability is concerned with quantifying the likelihoods of various events in situations involving elements of randomness or uncertainty.

Example 1.1.1 *50 people are gathered in a room for a probability lecture.*

How likely is it that at least two of these people have the same birthday? Is it

- Extremely unlikely?
- Unlikely?
- Likely?
- Almost certainly true?

How likely is it that two (or more) of these people were born on June 15?

Example 1.1.2 *In poker, a full house (3 cards of one rank and two of another, e.g. 3 fours and 2 queens) beats a flush (five cards of the same suit).*

A player is more likely to be dealt a flush than a full house. We will be able to precisely quantify the meaning of "more likely" here.

Example 1.1.3 *A coin is tossed repeatedly.*

Each toss has two possible outcomes:

$$\text{Heads (H)} \quad \text{or} \quad \text{Tails (T)},$$

both equally likely. The outcome of each toss is unpredictable; so is the sequence of H and T.
However, as the number of tosses gets large, we expect that the number of H (heads) recorded will fluctuate around 1/2 of the total number of tosses. We say the *probability* of a H is $\frac{1}{2}$, abbreviated by

$$\mathsf{P}(\mathsf{H}) = \frac{1}{2}.$$

Of course $\mathsf{P}(\mathsf{T}) = \dfrac{1}{2}$ also.

Example 1.1.4 *Suppose three coins are tossed. What is the probability of the outcome "2H, 1T"?*

> If the "experiment" of tossing three coins is repeated a large number of times, how often should we expect the result "2H, 1T"?

Possible outcomes are :

Coin 1	Coin 2	Coin 3	
H	H	H	
H	H	T	$\leftarrow$
H	T	H	$\leftarrow$
H	T	T	
T	H	H	$\leftarrow$
T	T	H	
T	H	T	
T	T	T	

All of these 8 outcomes are equally likely, and three of them involve the combination "2H, 1T". So the probability of this combination is 3/8.

$$\boxed{\mathsf{P}(2\mathsf{H},\ 1\mathsf{T}) = \tfrac{3}{8}.}$$

Problem 1.1.5 *If 4 coins are tossed, what is the probability of getting 3 heads and 1 tail?*

NOTES:

- In general, an event has associated to it a *probability*, which is a real number between 0 and 1.
- Events which are unlikely have low (close to 0) probability, and events which are likely have high (close to 1) probability.
- The probability of an event which is certain to occur is 1; the probability of an impossible event is 0.

Example 1.1.6 *(Probability and Addition)*

(a) *A fair die is cast. What is the probability that it will show either a 5 or a 6?*

(b) *Two fair dice are cast. What is the probability that at least one of them will show a 6?*

SOLUTION:
(a) There are six possible outcomes :

$$1, 2, 3, 4, 5, 6,$$

all equally likely - each has probability 1/6. So the probability of the event "5 or 6" is 1/3 : we expect this outcome in 1/3 of cases.

$$\mathsf{P}(5 \text{ or } 6) = \frac{1}{3}.$$

Note : $\mathsf{P}(5) = \frac{1}{6}$ and $\mathsf{P}(6) = \frac{1}{6}$. So

$$\mathsf{P}(5 \text{ or } 6) = \frac{1}{3} = \mathsf{P}(5) + \mathsf{P}(6).$$

In this example the probability of one or other of the events "5" and "6" is the sum of their individual probabilities.

(b) We need

$$\text{P(Die 1 shows 6 or Die 2 shows 6)}$$

There are $36 = 6^2$ possible outcomes, all equally likely. In $25 = 5^2$ of these, neither die shows a 6 (each of them shows 1,2,3,4 or 5). In the remaining 11, at least one die shows a 6. Thus

$$\mathsf{P}(\text{At least one } 6) = \frac{11}{36}.$$

Note : P(Die 1 shows 6)$= \dfrac{1}{6}$

P(Die 2 shows 6)$= \dfrac{1}{6}$

but P(Die 1 shows 6 *or* Die 2 shows 6) $= \dfrac{11}{36} \neq \dfrac{1}{6} + \dfrac{1}{6}$

Why does addition not apply here as in Part (a)?

What is different about the questions in (a) and (b)?
<u>Answer</u>: The two events in (a), namely "5" and "6" are *mutually exclusive* - they can never occur simultaneously (i.e. they cannot occur in the same instance of the

experiment).

The events in (b), “Die 1 shows 6” and “Die 2 shows 6” are not mutually exclusive - it is possible for both to occur simultaneously in the experiment described in (b), if both dice show 6.

GENERAL PROPERTY: If A and B are mutually exclusive events, then

$$P(A \text{ or } B) = P(A) + P(B).$$

1.2 Sample Spaces and Events

The purpose of the following definition is to establish some of the language that is commonly used to discuss problems in probability theory.

Definition 1.2.1

- *An* experiment *is a process whose outcome can be observed.*
- *The* sample space, S, *of an experiment is the set of possible outcomes for the experiment. (The sample space may be finite or infinite).*
- *An* EVENT *is a subset of the sample space, i.e. a (possibly empty) collection of possible outcomes to the experiment. We say that an event* A OCCURS *if the outcome is an element of* A.

NOTE: The difference between an *outcome* and an *event* in this context is that an outcome is an individual possible result of the experiment; an event may include many different possible outcomes.
The following examples describe how these definitions might apply in various applications.

Example 1.2.2 ***Experiment:*** *Toss a coin three times.*

The *Sample Space* is $S = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$
The event "2 heads, one tail" is the subset $A = \{HHT, HTH, THH\}$ of S.

Example 1.2.3 ***Experiment:*** *Roll a die.*

The *Sample Space* is the set of possible outcomes : $S = \{1, 2, 3, 4, 5, 6\}$

Define $A = \{1, 2\} \subseteq S$.
The set A is the event "the outcome is less than 3".

Example 1.2.4
Experiment: *Roll two dice (one red, one blue) and record the scores.*

The *Sample Space* is $S = \{(1, 1), (1, 2), \dots, (1, 6), (2, 1), \dots, (6, 1), \dots, (6, 6)\}$ (listed with the score on the red die first)
(36 = 6^2 possible outcomes, all equally likely if both dice are fair).

Example 1.2.5
Experiment: *Roll two dice and observe the total score.*

The Sample Space is $S = \{2, \dots, 12\}$
(11 possible outcomes, not all equally likely - see 1.2.4

Example 1.2.6
Experiment: *Continue tossing a coin until a tail appears, and observe the number of tosses.*

The *Sample Space* is $\{1, 2, \dots\} = \mathbb{N}$.
This sample space is infinite.
The subset $S_1 = \{5, 6, \dots\}$ of the sample space is the event that the first four tosses all produce heads.

Example 1.2.7
Experiment: *Measuring the hours a computer disk runs.*

The *Sample Space* is $S = [0, \infty)$.
The event $A = [0, 168]$ represents "breakdown in the first week".

Some Notation from Set Theory

Let S be a set with subsets A and B.

1. $\emptyset$ denotes the *empty set* : $\emptyset$ contains no elements and is a subset of *every* set.

2. Union : $A \cup B = \{a \in S : a \in A \text{ or } a \in B \text{ or both}\}$.

3. Intersection : $A \cap B = \{a \in S : a \in A \text{ and } a \in B\}$.

4. The *complement* of A in S is
$$\bar{A} = \{a \in S : a \notin A\}.$$
$\bar{A}$, also denoted $S \backslash A$, contains exactly those elements of S which are not in A.

5. A and B are called *disjoint* if $A \cap B = \emptyset$.

   Note - if S is the sample space of an experiment, then $A \cap B = \emptyset$ means precisely that the events A and B cannot occur simultaneously - they are mutually exclusive. Events $A_1, A_2, \dots$ are called *pairwise mutually exclusive* if $A_i \cap A_j = \emptyset$ for each choice of i and j.

   Note - For every subset A of S we have
$$A \cap \bar{A} = \emptyset \text{ and } S = A \cup \bar{A},$$
   so S is the *disjoint union* of A and its complement.

6. The *power set* of S, denoted $\mathcal{P}(S)$, is the set whose elements are all the subsets of S. If S is the sample space of an experiment, then $\mathcal{P}(S)$ is the set of possible events.

In Probability Theory, a *probability* $P(A)$ is assigned to every subset A of the sample space S of an experiment (i.e. to every event). The number $P(A)$ is a measure of how likely the event A is to occur and ranges from 0 to 1. We insist that the following two properties be satisfied :

1. $P(S) = 1$: The outcome is *certain* to be an element of the sample space.
2. If $A_1, A_2, A_3 \dots$ are *pairwise mutually exclusive events*, then

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = P(A_1) + P(A_2) + P(A_3) + \dots$$

Remark on Property 2 : Property 2 says :

> If $A_1, A_2, \dots$ are events no two of which can occur simultaneously, then the probability of the event "one of $A_1, A_2, \dots$" is the sum of the probabilities of the events $A_1, A_2, \dots$.

Take some time, bearing in mind Example 1.1.6, to convince yourself that this is a sensible property to insist on.

Definition 1.2.8 *A* probability measure *on a sample space* S *is a function*

$$P : \mathcal{P}(S) \longrightarrow [0, 1],$$

which satisfies Properties 1 and 2 above.

Example 1.2.9 *Suppose a die is unfairly balanced, so that it shows a 6 with probability* $\frac{1}{2}$*, the other 5 outcomes being equally likely.*

Experiment: Roll this unfair die.
Sample Space : $S = \{1, 2, 3, 4, 5, 6\}$
What should the probability measure be?
Solution:

$$\begin{aligned} P(6) &= \frac{1}{2} \\ \Longrightarrow P(\{1,2,3,4,5\}) &= \frac{1}{2} \\ \Longrightarrow \ P(1) = P(2) = P(3) = P(4) = P(5) &= \frac{1}{5} \times \frac{1}{2} = \frac{1}{10}. \end{aligned}$$

If A is a subset of S (there are $2^6 = 64$ subsets), we need to indicate what $P(A)$ is :

$$P(A) = \begin{cases} \frac{1}{2} + (|A| - 1)\frac{1}{10} & \text{if } \ 6 \in A \\ \\ |A|\frac{1}{10} & \text{if } \ 6 \notin A \end{cases}$$

Note : Here $|A|$ denotes the *number of elements* in A. The above definition of the probability measure P says

- If the event A includes the outcome 6 then its probability is $\frac{1}{2}$ (for the "6") plus $\frac{1}{10}$ for each remaining outcome in A, of which there are $|A| - 1$.
- If the event A does not include the outcome 6 then its probability is $|A|\frac{1}{10}$ - the number of outcomes included in A is $|A|$, and each of these has the same probability $\frac{1}{10}$.

Problem 1.2.10 *Suppose a die is unfairly balanced so that*

(i) $P(1) = P(2) = P(3) = P(4)$ (i.e. outcomes 1,2,3,4 are all equally likely).

(ii) $P(5) = P(6) = 2P(1)$ (i.e. outcomes 5 and 6 are equally likely, but each is twice as likely as any of the other four possible outcomes).

Describe as in Example 1.2.9 the probability measure on the sample space $S = \{1, 2, 3, 4, 5, 6\}$ for rolling this die.

Properties 1 and 2 of Definition 1.2.3 are the only Properties which we insist must be satisfied. In Section 1.1.3 we will derive some consequences of these properties, and see that they are consistent with what we would expect based on examples.

1.3 Further Properties of Probability

Throughout this section, let S be a sample space (for some experiment), and

$$P : \mathcal{P}(S) \longrightarrow [0,1]$$

a probability measure. We will refer to properties 1 and 2 of Definition 1.2.3 as *Property 1* and *Property 2* respectively.

1. **Union and Intersection** ("OR" and "AND")
   Let A and B be events (subsets of S). Then

   - The event "A or B" occurs if the outcome belongs either to A or to B, i.e. to their *union* $A \cup B$.
   - The event "A and B" occurs if the outcome belongs both to A and to B, i.e. to their *intersection* $A \cap B$.

2. **Complement of A** (not A)
   For $A \subseteq S$, $\bar{A}$ (the *complement* of A) is the set of outcomes which do not belong to A. So the event $\bar{A}$ occurs precisely when A does not. Since A and $\bar{A}$ are mutually exclusive (disjoint) and $A \cup \bar{A} = S$, we have by properties 1 and 2

   $$P(A) + P(\bar{A}) = P(S) = 1.$$

   Thus

   $$P(\bar{A}) = 1 - P(A) \text{ for any event } A. \quad \text{(Property 3)}$$

Example 1.3.1 *Suppose A and B are subsets of S with $P(A) = 0.2$, $P(B) = 0.6$ and $P(A \cap B) = 0.1$. Calculate*

(a) $P(A \cup B)$

(b) $P(\bar{A})$

(c) $P(\bar{A} \cup B)$

(d) $P(\bar{A} \cap \bar{B})$

Solution

(a) $A \cup B$ is the disjoint union of $A \cap \bar{B}$, $A \cap B$ and $\bar{A} \cap B$ (*disjoint* means that no two of these three sets have any elements in common). Hence

$$P(A \cup B) = P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B).$$

To find $P(A \cap \bar{B})$, note $P(A) = P(A \cap \bar{B}) + P(A \cap B)$. So

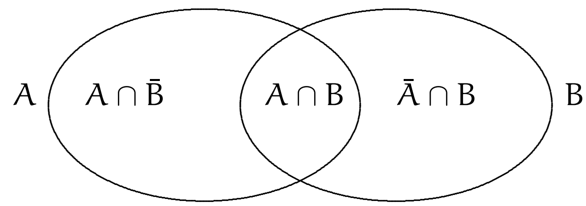
$$P(A \cap \bar{B}) = P(A) - P(A \cap B) \text{ for any events } A \text{ and } B. \quad \text{(Property 4)}$$

Thus in our example $P(A \cap \bar{B}) = 0.2 - 0.1 = 0.1$.
Similarly $P(\bar{A} \cap B) = P(B \cap \bar{A}) = P(B) - P(B \cap A) = 0.6 - 0.1 = 0.5$. So

$$P(A \cup B) = 0.1 + 0.1 + 0.5 = 0.7.$$

Note the *general fact* :

$$\begin{aligned} P(A \cup B) &= P(A \cap \bar{B}) + P(A \cap B) + P(\bar{A} \cap B) \\ &= P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B) \\ \Longrightarrow P(A \cup B) &= P(A) + P(B) - P(A \cap B) \quad \text{(Property 5)} \end{aligned}$$



Also,

$$P(A \cap B) = P(A) + P(B) - P(A \cup B) \quad \text{(Property 6)}$$

(b) $P(\bar{A}) = 1 - P(A) = 1 - 0.2 = 0.8$.

(c) $P(\bar{A} \cup B) = P(\bar{A}) + P(B) - P(\bar{A} \cap B) = 0.8 + 0.6 - 0.5 = 0.9$.

(d) $P(\bar{A} \cap \bar{B}) = 1 - P(A \cup B) = 1 - 0.7 = 0.3$.

Example 1.3.2 *Weather records over 1000 days show rain on 550 days and sunshine on 350 days. If 700 days had either rain or sunshine, find the probability that on a day selected at random there was*

(a) Both rain and sunshine.

(b) Rain but not sunshine.

(c) Neither rain nor sunshine.

SOLUTION: First set up some notation to describe the events of interest:
R : Rain on selected day.
S : Sunshine on selected day.

(a) $P(R \cap S) = P(R) + P(S) - P(R \cup S)$ by Property 6

So $P(R \cap S) = \dfrac{550}{1000} + \dfrac{350}{1000} - \dfrac{700}{1000} = \dfrac{1}{5}$.

(b) $P(R \cap \bar{S}) = P(R) - P(R \cap S)$ by Property 4

So $P(R \cap \bar{S}) = \dfrac{550}{1000} - \dfrac{1}{5} = \dfrac{7}{20}$.

(c) $P(\bar{R} \cap \bar{S}) = P(\overline{R \cup S}) = 1 - P(R \cup S)$ by Property 3

So $P(\bar{R} \cap \bar{S}) = 1 - \dfrac{700}{1000} = \dfrac{3}{10}$.

Problem 1.3.3 *A college has 400 2nd Science students, of whom 150 study mathematics, 50 study both mathematics and chemistry, and 220 study either mathematics or chemistry. Find the probability that a student selected at random studies*

(a) Chemistry

(b) Chemistry and not mathematics

(c) Neither chemistry nor mathematics.

1.4 Conditional Probability

Example 1.4.1 *Two fair dice are rolled, 1 red and 1 blue. The Sample Space is*

$$S = \{(1,1),(1,2),\ldots,(1,6),\ldots,(6,6)\}$$

-36 outcomes, all equally likely (here $(2,3)$ *denotes the outcome where the red die show 2 and the blue one shows 3).*

(a) Consider the following events:
A : Red die shows 6.
B : Blue die shows 6.

Find $P(A),\ P(B)$ and $P(A \cap B)$.

SOLUTION: $P(A) = \dfrac{1}{6},\ P(B) = \dfrac{1}{6}$
$A \cap B = \{(6,6)\} \Longrightarrow P(A \cap B) = \dfrac{1}{36}$.

NOTE: $\frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$ so $P(A \cap B) = P(A) \times P(B)$ for this example.

This is not surprising - we expect A to occur in $\frac{1}{6}$ of cases. In $\frac{1}{6}$ of *these* cases. i.e. in $\frac{1}{36}$ of *all* cases, we expect B to also occur.

(b) Consider the following events:
C : Total Score is 10.
D : Red die shows an even number.

Find $P(C),\ P(D)$ and $P(C \cap D)$.

SOLUTION:

$P(C) = P(\{(4,6),(5,5),(6,4)\}) = \dfrac{3}{36} = \dfrac{1}{12}$.
$P(D) = \dfrac{1}{2}$.
$C \cap D = \{(4,6),(6,4)\} \Longrightarrow P(C \cap D) = \dfrac{2}{36} = \dfrac{1}{18}$.
NOTE: $\frac{1}{18} \neq \frac{1}{2} \times \frac{1}{12}$ so $P(C \cap D) \neq P(C) \times P(D)$.
Why does multiplication not apply here as in part (a) ?

ANSWER: Suppose C occurs : so the outcome is either $(4,6),\ (5,5)$ or $(6,4)$. In *two* of these cases, namely $(4,6)$ and $(6,4)$, the event D also occurs. Thus

Although $P(D) = \frac{1}{2}$, the probability that D occurs *given that* C occurs is $\frac{2}{3}$.

We write $P(D|C) = \frac{2}{3}$, and call $P(D|C)$ the *conditional probability of* D *given* C.

NOTE: In the above example

$$P(C \cap D) = \frac{1}{18} = \frac{1}{12} \times \frac{2}{3} = P(C) \times P(D|C).$$

This is an example of a general rule.

Definition 1.4.2 *If* A *and* B *are events, then* $P(A|B)$*, the* conditional probability *of* A *given* B*, is defined by*

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

Problem 1.4.3 *Experiment : Roll a die.*

Events: A - score is a multiple of 3. $A = \{3, 6\}$.
B - score is at most 5. $B = \{1, 2, 3, 4, 5\}$.
Find $P(A)$, $P(B)$, $P(A \cap B)$, $P(A|B)$ and $P(B|A)$.
(Of course $P(A|B)$ and $P(B|A)$ will be different in general.)

Definition 1.4.4 *(Independence) Let* A *and* B *be events. They are called* independent *if*

$$P(A \cap B) = P(A) \times P(B).$$

Equivalently, A and B are independent if

$$P(A|B) = P(A) \text{ and } P(B|A) = P(B).$$

These means that the probability of one of these events occurring is unaffected by whether the other occurs or not.

NOTE: If the events A and B are mutually exclusive then $P(A|B) = 0$ and $P(B|A) = 0$.

Problem 1.4.5 *Experiment : Roll two dice, one red and one blue. Which of the following pairs of events are independent? In each case find* $P(A)$, $P(B)$, $P(A \cap B)$, $P(A|B)$ *and* $P(B|A)$.

(a)	A: Red die shows even score.	B: Blue die shows even score.
(b)	A: Red die scores 1.	B: Total score is 8.
(c)	A: Red die shows even score.	B: Total score is even.
(d)	A: Red die scores 4 or more.	B: Total score is 6 or more.

Example 1.4.6 *A factory has two machines* A *and* B *making 60% and 40% respectively of the total production. Machine A produces 3% defective items, and B produces 5% defective items. Find the probability that a given defective part came from A.*

SOLUTION: We consider the following events :

A: Selected item comes from A.
B: Selected item comes from B.
D: Selected item is defective.

We are looking for $P(A|D)$. We know :

$$P(A) = 0.6,\ P(B) = 0.4,\ P(D|A) = 0.03,\ P(D|B) = 0.05$$

Now $P(A|D) = \dfrac{P(A \cap D)}{P(D)}$. So we need $P(A \cap D)$ and $P(D)$.

- $P(A \cap D) = P(A)P(D|A) = 0.6 \times 0.03 = 0.018$
- $P(D) = P(D \cap B) + P(D \cap A)$ since D is the union of the mutually exclusive events $D \cap A$ and $D \cap B$ (the entire sample space is the union of the mutually exclusive events A and B).

$$\begin{aligned} P(D) &= P(D \cap B) + P(D \cap A) \\ &= P(B)P(D|B) + P(A)P(D|A) \\ &= 0.4 \times 0.05 + 0.6 \times 0.03 \\ &= 0.038. \end{aligned}$$

Finally then $P(A|D) = \dfrac{0.018}{0.038} = \dfrac{9}{19}$.
NOTE: In the above problem we used the relation

$$P(A|D) = \frac{P(D)P(D|A)}{P(D)P(D|A) + P(D)P(D|B)}$$

where A and B are disjoint sets whose union is the whole sample space. This is an application of a theorem known as Bayes's Theorem.

Definition 1.4.7 *A* partition *of a sample space S is a collection of events $E_1, E_2, \dots, E_n$ for which*

1. $E_1 \cup E_2 \cup \cdots \cup E_n = S$.
2. $E_i \cap E_j = \emptyset$ *for $i \neq j$: the E_i are pairwise mutually exclusive – no two of them can occur simultaneously.*

Theorem 1.4.8 *(Bayes's Theorem) Let $E_1, \dots, E_n$ be a partition of a sample space S. Let $A \subseteq S$ be any event, and choose $i \in \{1, \dots, n\}$. Then*

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^{n} P(E_j)P(A|E_j)}.$$

Proof:

1. Certainly $P(E_i|A) = \dfrac{P(A \cap E_i)}{P(A)}$.

2. $P(A \cap E_i) = P(E_i)P(A|E_i)$ by Definition 1.4.2.

3. To describe $P(A)$:

$$\begin{aligned} P(A) &= P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_n) \\ &= P(E_1)P(A|E_1) + P(E_2)P(A|E_2) + \cdots + P(E_n)P(A|E_n) \\ &= \sum_{j=1}^{n} P(E_j)P(A|E_j). \end{aligned}$$

Hence

$$P(E_i|A) = \frac{P(E_i)P(A|E_i)}{\sum_{j=1}^{n} P(E_j)P(A|E_j)}.$$

□

Example 1.4.9 *A test is developed to detect a certain medical condition. It is found that the population breaks into three categories.*

Category A	Afflicted individuals who give a positive result 97% of the time
Category D	Individuals with a different condition who give a positive result 10% of the time
Category H	Unaffected individuals who give a positive result 5% of the time.

The categories A, D and H represent $1\%, 3\%$ and 96% of the population respectively. What is the probability that someone who is selected at random and tests positive for the condition actually has the condition?
Let A be the event someone is afflicted with the disease and R the event that the test result is positive. We want $P(A|R)$.
By Bayes's theorem

$$P(A|R) = \frac{P(A)P(R|A)}{P(A)P(R|A) + P(D)P(R|D) + P(H)P(R|H)}$$

(Here $E_1 = A$, $E_2 = D$, $E_3 = H$ form a partition.)

$$\begin{aligned} &= \frac{(.01)(.97)}{(.01)(.97) + (.03)(0.1) + (.96)(.05)} \\ &= 0.16 \end{aligned}$$

So only 16% of positive results will have the disease. This is not a good test.

Problem 1.4.10 *Suppose the test is carried out twice and the test case tests positive both times. Assuming the two tests are independent, what then is the probability that the test case has the disease?*

Problem 1.4.11 *A machine consists of four components linked in parallel, so that the machine fails only if all four components fail. Assume component failures are independent of each other. If the components have probabilities 0.1, 0.2, 0.3 and 0.4 of failing when the machine is turned on, what is the probability that the machine will function when turned on?*

Problem 1.4.12 *A student takes a multiple choice exam in which each question has 5 possible answers, one correct. For 70% of the questions, the student knows the answer, and she guesses at random for the remaining 30%.*

(a) What is the probability that on a given question the student gets the correct answer?

(b) If the student answers a question correctly, what is the probability that she knows the answer?

1.5 Some Counting Principles

In discrete problems, estimating the probability of some event often amounts to counting the number of possible outcomes that have some property of interest, and expressing this as a proportion of the total number of outcomes. For example, determining the probability that a randomly dealt poker hand is a full house means counting the number of possible full houses and counting the number of total poker hands. "Discrete" means roughly that the outcomes of the experiment are somehow "spread out" instead of blended together in a continuous fashion. So there is a close connection between problems of discrete probability theory and problems of enumerative combinatorics (i.e. counting).
The following is a list of some basic but important counting principles.

1. Let n be a positive integer. Then "n factorial", denoted $n!$, is defined by

$$\boxed{n! = n \times (n-1) \times \cdots \times 2 \times 1.}$$

   $n!$ is the number of ways of arranging n distinct objects in order : the number of *permutations* of n objects.

   Example 1.5.1 *There are* $3! = 3 \times 2 \times 1 = 6$ *permutations of the 3 letters* a, b, c *:*

$$abc,\ acb,\ bac,\ bca,\ cab,\ cba.$$

2. Let k, n be positive integers with $k \leqslant n$. The number of ways of choosing k objects from a set of n distinct objects, and arranging them in order is

$$\boxed{n \times (n-1) \times \cdots \times (n-k+1) = \tfrac{n!}{(n-k)!}.}$$

   There are n choices for the first object, $n-1$ for the second, etc., finally $n-k+1$ for the kth object. This number is called ${}^nP_k$, the number of k-permutations of n distinct objects.

   Example 1.5.2 *In a race with 8 runners, the number of ways in which the gold, silver and bronze medals can be awarded is*

$${}^8P_3 = \frac{8!}{(8-3)!} = 8 \times 7 \times 6 = 336.$$

3. For n, k positive integers with $k \leqslant n$, the number of ways to select k objects (in no particular order) from a set of n distinct objects is $\frac{n!}{k!(n-k)!}$.

   This number is denoted ${}^nC_k$ or $\binom{n}{k}$, and called "n choose k".

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

Example 1.5.3 *If the 8-person race ends in a three-way dead heat, the number of possibilities for the three winners is*

$$\binom{8}{3} = \frac{8!}{3!(8-3)!} = \frac{8!}{3!5!} = 56.$$

Example 1.5.4 *Find the probability that a randomly dealt poker hand will be*

(a) A full house (3 cards of one rank and 2 of another).

(b) A (non-straight) flush (5 cards of the same suit, not of consecutive rank).

SOLUTION: There are $\binom{52}{5}$ $(= 2598960)$ different poker hands.
(a) To count the full houses : there are 13 choices for the rank of the 3-card set. Within the four cards of this rank, the number of ways to choose three is $\binom{4}{3} = 4$. Having chosen these, 12 choices remain for the rank of the pair. Within this rank there are $\binom{4}{2} = 6$ ways to choose 2 cards. Thus the number of full houses is

$$13 \times \binom{4}{3} \times 12 \times \binom{4}{2} = 13 \times 4 \times 12 \times 6 = 3744.$$

Thus the probability of a full house being dealt is

$$\frac{3744}{\binom{52}{5}} \sim 0.0014.$$

(b) Number of Flushes : There are four choices for a suit, and having chosen a suit the number of ways to choose 5 cards belonging to it is $\binom{13}{5}$. Thus the number of flushes in a 52-card deck is

$$4 \times \binom{13}{5} = 5148.$$

From this count we need to exclude the straight flushes of which there are 40 : there are 10 in each suit, since a straight may have any of Ace,2,3,...,10 as the lowest rank card (according to my recent poker researches, the sequences A,2,3,4,5 and 10,J,Q,K,A both qualify as straights). Thus the number of straight flushes is 40, and the number of non-straight flushes is $5148 - 40 = 5108$. The probability of a non-straight flush is

$$\frac{5108}{\binom{52}{5}} \sim 0.002.$$

So a non-straight flush is more likely than a full house.

Problem 1.5.5 *Find the probability that a randomly dealt poker hand is*

(a) *A straight flush.*

(b) *A straight (5 cards of consecutive ranks, not all of the same suit).*

(c) *3 of a kind (3 cards of the same rank, 2 more of different ranks).*

Problem 1.5.6 *A bridge hand contains 13 cards. Find the probability that a randomly dealt bridge hand*

(a) *Contains cards from all four suits.*

(b) *Contains exactly five clubs.*

Example 1.5.7 *(The "Birthday" Problem) Twenty people are assemble in a room. Find the probability that two of them have the same birthday. (Assume there are 365 days all equally likely to be a birthday).*

SOLUTION: Let $(x_1, \dots, x_{20})$ be the list of birthdays (so x_i is the birthday of the ith person). The number of possibilities for the list is $(365)^{20}$. The number with the x_i all different is ${}^{365}P_{20}$ (ordered lists of 20 from 365).

$$ {}^{365}P_{20} = \frac{(365)!}{(345)!} = 365 \times 364 \times 363 \times \cdots \times 347 \times 346. $$

Thus the probability that at least two birthdays are the same is

$$ 1 - \frac{365 \times 364 \times \cdots \times 346}{(365)^{20}} \sim 0.411. $$

Problem 1.5.8 *Give a formula for the probability that among n people, two have the same birthday. The following table indicates the (maybe surprising?) values of this probability for various values of n:*

$n = 23$	~ 0.51
$n = 30$	~ 0.71
$n = 37$	~ 0.85
$n = 50$	~ 0.97
$n = 100$	> 0.999999

This concludes Chapter 1, Probability Spaces. A "Probability Space" is just a sample space equipped with a probability measure.

Chapter 2

Discrete Random Variables

2.1 Probability Distributions

Example 2.1.1 *A coin is biased so that it shows heads with probability* 0.6.
Experiment : Toss this coin 5 times.

Sample Space S : Sequences of H and T with 5 terms ($|S| = 2^5 = 32$).
Let X be the number of heads in an outcome, e.g. in the outcome $THHTH$, $X = 3$.
We write

$$X(THHTH) = 3, \; X(HTHTT) = 2, \; X(HHTHH) = 4, \text{ etc.}$$

Possible values of X are $0, 1, 2, 3, 4, 5$.
Using the fact the the results of different tosses are independent, we can calculate the probabilities of each of these

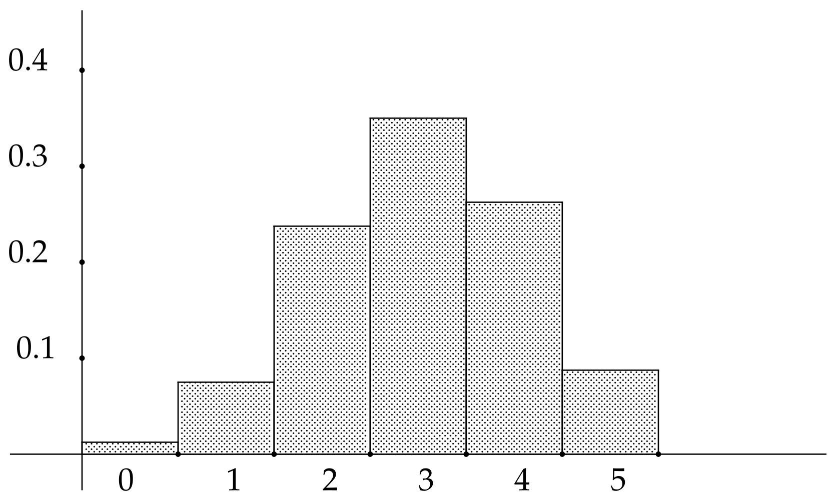
$$\begin{aligned}
P(X = 0) &= (0.4)(0.4)\dots(0.4) = (0.4)^5 = 0.01024 \\
P(X = 1) &= \binom{5}{1}(0.6)^1(0.4)^4 = 0.0768 \\
&\quad \text{—we can choose one of the 5 tosses for the } H. \\
P(X = 2) &= \binom{5}{2}(0.6)^2(0.4)^3 = 0.2304 \\
&\quad \text{—we can choose two of the 5 tosses for the } H \text{ in } \binom{5}{2} \; ways. \\
P(X = 3) &= \binom{5}{3}(0.6)^3(0.4)^2 = 0.3456 \\
&\quad \text{—we can choose 3 of the 5 tosses for the } H \text{ in } \binom{5}{3} \; ways.
\end{aligned}$$

$$P(X=4) = \binom{5}{4}(0.6)^4(0.4)^1 = 0.2592$$

—we can choose 4 of the 5 tosses for the H in $\binom{5}{4}$ *ways*.

$$P(X=5) = \binom{5}{5}(0.6)^5 = 0.07776$$

(Observe that the sum of these probabilities is 1).
This information can be displayed in a *histogram* (bar chart) as follows :



Probability Function for X

REMARKS

- X in this example is a *discrete random variable*. It associates a number (in this case the number of heads) to every outcome of an experiment. Officially

A random variable on a sample space S is a function $X : S \longrightarrow \mathbb{R}$.

- $X(S)$ denotes the set of possible values assumed by the random variable X (so in this example $X(S) = \{0,1,2,3,4,5\}$ - possible numbers of heads in 5 tosses).

- The *probability function* (or *density function* or *probability distribution*) associated to the random variable X is the function

  $$f_X : X(S) \longrightarrow [0,1]$$

  defined by $f_X(x) = P(X = x)$ for $x \in X(S)$.

  So in Example 2.1.1, $3 \in X(S)$ and $f_X(3) = P(X = 3) = 0.3456$.

THE BINOMIAL PROBABILITY FUNCTION

Definition 2.1.2 *A* Bernoulli Trial *is an experiment with two possible outcomes,* success *(s), occurring with probability* p*, and* failure *(f), occurring with probability* $q = 1 - p$.

e.g. Tossing a coin (if "Heads" is regarded as "success").

Definition 2.1.3 *If a Bernoulli trial (in which success has probability* p*) is performed* n *times, let* X *be the number of successes in the* n *trials. Then* X *is called a* binomial random variable *with parameters* n *and* p.
So the X *of Example 2.1.1 is a binomial random variable with parameters* $n = 5$ *and* $p = 0.6$.

The probability function for X *is given by*

$$\boxed{f_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \text{ for } k = 0, 1, \ldots, n}$$

REMARK: Since there are $n + 1$ possible values for X, we should expect

$$\begin{aligned} P(X = 0) + P(X = 1) + \cdots + P(X = n) &= 1 \\ \text{i.e. } \sum_{k=0}^{n} f_X(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} &= 1. \end{aligned}$$

That this is true follows from the binomial theorem applied to the expression $(p + (1-p))^n$. Recall that the binomial theorem says

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

for a positive integer n.

Problem 2.1.4 *If 2% of items produced by a factory are defective, find the probability that a (randomly packed) box of 20 items contains*

(a) At least 3 defective items.

(b) At most 3 defective items.

NOTE: The number of defective items in a box of 20 is a binomial random variable with parameters $n = 20$ and $p = 0.02$.

Example 2.1.5 *A lotto competition works as follows. An entry involves selecting four numbers from* $\{1, 2, \ldots, 20\}$*. After all entries have been completed, four winning numbers are selected from the set* $\{1, 2, \ldots, 20\}$*. A "match" occurs within an entry when that entry includes one of the four winning numbers - so the number of matches in an entry can be 0,1,2,3 or 4. If* X *is the number of "matches" in a random selection of four numbers, describe the probability function of* X*.*

SOLUTION: First note that the number of possible entries is $\binom{20}{4}$.

A. $X = 0$ if the entry contains none of the winning numbers, but four of the 16 losing numbers. This can happen in $\binom{16}{4}$ ways. So

$$P(X = 0) = f_X(0) = \frac{\binom{16}{4}}{\binom{20}{4}} \approx 0.3756.$$

B. $X = 1$ if the entry contains one of the 4 winning numbers, and three of the 16 losing numbers. This can happen in $\binom{4}{1}\binom{16}{3}$ ways. So

$$P(X = 1) = f_X(1) = \frac{\binom{4}{1}\binom{16}{3}}{\binom{20}{4}} \approx 0.4623.$$

C. $X = 2$ if the entry contains two of the 4 winning numbers, and two of the 16 losing numbers. This can happen in $\binom{4}{2}\binom{16}{2}$ ways. So

$$P(X = 2) = f_X(2) = \frac{\binom{4}{2}\binom{16}{2}}{\binom{20}{4}} \approx 0.1486.$$

D. $X = 3$ if the entry contains three of the 4 winning numbers, and one of the 16 losing numbers. This can happen in $\binom{4}{3}\binom{16}{1}$ ways. So

$$P(X = 3) = f_X(3) = \frac{\binom{4}{3}\binom{16}{1}}{\binom{20}{4}} \approx 0.0132.$$

E. $X = 4$ if the entry contains all four of the winning numbers. This can happen only in one way. So

$$P(X = 4) = f_X(4) = \frac{1}{\binom{20}{4}} \approx 0.0002.$$

NOTE: This problem can be approached in a slightly more systematic way as follows. Entering the lotto involves selecting 4 objects from 20, of which four are "good" and 16 are "bad". The random variable X is the number of "good" objects in a selection. For $k = 0, 1, 2, 3, 4$, the number of selections that include k "good" objects is

$$\binom{4}{k}\binom{16}{4-k}.$$

Thus the probability function for X is given by

$$P(X = k) = f_X(k) = \frac{\binom{4}{k}\binom{16}{4-k}}{\binom{20}{4}}.$$

This X is an example of a *hypergeometric random variable*.

Definition 2.1.6 *Suppose that a collection of n distinct objects includes m "good" objects and $n - m$ "bad" objects. Suppose $k \leqslant m$ and let X be the number of good objects in a random selection of k objects from the full collection. Then X is a* hypergeometric random variable *with parameters (n, m, k) and the probability distribution of X is given by*

$$P(X = r) = f_X(r) = \frac{\binom{m}{r}\binom{n-m}{k-r}}{\binom{n}{k}}, \text{ for } r = 0, 1, \ldots, k.$$

Problem 2.1.7 *Let X be the number of clubs in a randomly dealt poker hand. Write down a formula for the probability function of X.*

THE GEOMETRIC PROBABILITY FUNCTION

Example 2.1.8 *A fair die is rolled repeatedly until it shows a 6. Let X be the number of rolls required. Describe the probability function of X.*

SOLUTION: Each roll is a Bernoulli trial in which a 6 represents success (s), and occurs with probability 1/6.
Any other score is failure : $P(f) = 5/6$.
Sample Space : $S = \{s, fs, ffs, fffs, \ldots\}$
Random Variable X : $X(s) = 1,\ X(fs) = 2,\ X(fffffs) = 6$, etc.
$X(S) = 1, 2, 3, \ldots \sim \mathbb{N}$ - possible values of X.

Probability Function :

$$\begin{aligned}
f_X(1) = P(s) &= \frac{1}{6} \\
f_X(2) = P(fs) &= \frac{5}{6} \times \frac{1}{6} \\
\vdots \quad &\quad \vdots \\
f_X(k) = P(\underbrace{ff\ldots f}_{k-1} s) &= \left(\frac{5}{6}\right)^{k-1} \frac{1}{6} \text{ for } k \in \mathbb{N}.
\end{aligned}$$

NOTE: We should expect $\sum_{k=1}^{\infty} f_X(k) = 1$; this is

$$\frac{1}{6} + \left(\frac{5}{6} \times \frac{1}{6}\right) + \left(\left(\frac{5}{6}\right)^2 \times \frac{1}{6}\right) + \ldots$$

- a *geometric series* with first term $\frac{1}{6}$ and common ratio $\frac{5}{6}$. Its sum is

$$\frac{1/6}{1-(5/6)} = 1.$$

X is an example of a *geometric random variable*.

Definition 2.1.9 *Independent Bernoulli trials with probability of success* p *are carried out until a success is achieved. Let* X *be the number of trials. Then* X *is the* geometric random variable *with parameter* p*, and probability function given by*

$$f_X(k) = q^{k-1}p, \text{ for } k = 1, 2, \ldots \text{ and } q = 1 - p.$$

Problem 2.1.10 *Suppose* X *is a geometric random variable with parameter* $p = 0.8$*. Calculate the probabilities of the following events.*

(a) $X \leqslant 3$.

(b) $4 \leqslant X \leqslant 7$.

(c) $X > 6$.

NOTE*: Recall that in a geometric series with initial term* a *and common ratio* r*, the sum of the first* n *terms is* $a\dfrac{1-r^n}{1-r}$.

THE NEGATIVE BINOMIAL PROBABILITY FUNCTION

Example 2.1.11 *A researcher interviews people selected at random in order to find 12 smokers. Let X be the number of people interviewed. If smokers constitute 30% of the population, what is the probability function of X?*

SOLUTION: Selecting a person is a Bernoulli trial, with probability of success 0.3 (the person is a smoker).
Sample Space : Sequences of s, f ending on the 12th s.
Random Variable X : Number of people interviewed to get 12 successes.
Possible values of X : $12, 13, 14, \ldots$
Probability Function :
We need $f_X(k) = P(X = k)$ where $k \in \{12, 13, \ldots\}$.

$$\begin{aligned} f_X(k) &= \text{Probability of success in the } k\text{th trial } \textit{and} \\ &\qquad \text{11 successes in the first } k-1 \text{ trials} \\ &= (0.3)\binom{k-1}{11}(0.3)^{11}(0.7)^{(k-1)-11} \\ &= \binom{k-1}{11}(0.3)^{12}(0.7)^{k-12} \end{aligned}$$

Definition 2.1.12 *If independent Bernoulli trials with probability of success p are carried out, let X be the number of trials required to observe r successes. Then X is the* negative binomial random variable *with parameters r and p. The probability function of X is described by*

$$\boxed{f_X(k) = \tbinom{k-1}{r-1} p^r (1-p)^{k-r} \text{ for } k = r, r+1, \ldots}$$

REMARKS:

1. If $r = 1$, this is the geometric random variable.
2. *Binomial* Random Variable with parameters n, p - no. of *successes* in n trials.

   Negative Binomial Random Variable with parameter $r.p$ - no. of *trials* required to get r successes.

Problem 2.1.13 *In Example 2.1.11, find the probability that*

(a) The 6th person interviewed is the 3rd smoker.

(b) There will be at least 3 smokers among the first 6 interviewees.

2.2 Expectation and Variance of a Random Variable

Example 2.2.1 *Let X be the number of clubs in a poker hand;* $X(S) = \{0, 1, 2, 3, 4, 5\}$.

Then X is a hypergeometric random variable with parameters $(52, 13, 5)$. The probability distribution of X is given by

$$p_X(k) = \frac{\binom{13}{k}\binom{39}{5-k}}{\binom{52}{5}}.$$

$p_X(0)$	$p_X(1)$	$p_X(2)$	$p_X(3)$	$p_X(4)$	$p_X(5)$
≈ 0.2215	≈ 0.4114	≈ 0.2743	≈ 0.0816	≈ 0.0107	≈ 0.0005

So the most likely number of clubs in a poker hand is 1. Suppose that a large number (say 10,000) of poker hands is dealt. Amongst the 10,000 hands, we "expect" the following distribution of clubs :

No. of Clubs	0	1	2	3	4	5
No. of Hands	2215	4114	2743	816	107	5

So the total number of clubs in the 10,000 hands is "expected" to be

$$0 \times 2215 + 1 \times 4114 + 2 \times 2743 + 3 \times 816 + 4 \times 107 + 5 \times 5 = 12,500.$$

i.e. the "average" number of clubs per hand is $\frac{12500}{10000} = 1.25$. Of course this is not surprising since one-quarter of all cards are clubs. Note

$$1.25 = 0 \times p_X(0) + 1 \times p_X(1) + 2 \times p_X(2) + 3 \times p_X(3) + 4 \times p_X(4) + 5 \times p_X(5) = \sum_{k=0}^{5} k p_X(k).$$

1.25 is the *expectation*, *expected value* or *mean* of the random variable X, denoted by $E(X)$ (or sometimes μ or $\mu(X)$).
$E(X)$ is a *weighted average* of the possible values of X, the weight attached to a particular value being its probability.

Definition 2.2.2 *Let $X : S \longrightarrow \mathbb{R}$ be a discrete random variable. The* expectation *of* X, *denoted* $E(X)$, *is defined by*

$$E(X) = \sum_{k \in X(S)} k p_X(k)$$

Note: As Example 2.2.1 shows, $E(X)$ may not actually be a possible value of X but it is the value assumed "on average" by X.

Problem 2.2.3 *Let X be the score on a fair die. Calculate* $E(X)$.

NOTE: Suppose that X is a hypergeometric random variable with parameters (n, m, k) (i.e. X is the number of good objects chose when k objects are selected from a collection of n in which m are good and $n - m$ are bad). Then

$$\boxed{E(X) = \frac{m}{n}k}$$

JUSTIFICATION FOR THIS ASSERTION: The proportion of "good" objects in the full selection is $\frac{m}{n}$. So on average we would expect a selection of k objects to contain $\frac{m}{n}k$ good ones.

Example 2.2.4 *Suppose that X is the number of smokers in a random selection of five persons from the population. Suppose also that 20% of people smoke. What is the expected value of X?*

SOLUTION: Possible values of X are 0,1,2,3,4,5 and the probability distribution of X is given by

$$f_X(k) = P(X = k) = \binom{5}{k}(0.2)^k(0.8)^{5-k}$$

Thus

$$\begin{aligned}
E(X) &= 0 \times \binom{5}{0}(0.2)^0(0.8)^5 + 1 \times \binom{5}{1}(0.2)^1(0.8)^4 + 2 \times \binom{5}{2}(0.2)^2(0.8)^3 + \\
&\quad 3 \times \binom{5}{3}(0.2)^3(0.8)^2 + 4 \times \binom{5}{4}(0.2)^4(0.8)^1 + 5 \times \binom{5}{5}(0.2)^5(0.8)^0
\end{aligned}$$

Note that for $k = 1, 2, 3, 4, 5$ we have $k\binom{5}{k} = 5\binom{4}{k-1}$. Thus

$$\begin{aligned}
E(X) &= 0.2\left(5(0.8)^4 + 5\binom{4}{1}(0.2)(0.8)^3 + 5\binom{4}{2}(0.2)^2(0.8)^2 + 5\binom{4}{3}(0.2)^3(0.8) + 5(0.2)^4\right) \\
&= 0.2(5)\left((0.8)^4 + \binom{4}{1}(0.2)(0.8)^3 + \binom{4}{2}(0.2)^2(0.8)^2 + \binom{4}{3}(0.2)^3(0.8) + 5(0.2)^4\right) \\
&= 0.2(5)(0.8 + 0.2)^4 \\
&= 0.2(5) \\
&= 1.
\end{aligned}$$

This answer is not surprising. If the proportion of smokers in the population is 20%, we should expect that a random sample of five people would contain one smoker "on average".

NOTE: Suppose that X is a binomial random variable with parameters n and p. Then

$$\boxed{E(X) = np.}$$

JUSTIFICATION FOR THIS ASSERTION: The proportion of trials resulting in success ins p. Therefore among n trials, we would expect pn of them to result in success on average.

VARIANCE OF A RANDOM VARIABLE

Let X be a discrete random variable on a sample space S. Then $E(X)$ is a real number, and $X - E(X)$ is another discrete random variable on S; so is $(X - E(X))^2$.

Definition 2.2.5 1. *The* variance *of* X *is defined by*

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

Note $\mathrm{Var}(X) \geqslant 0$. $\mathrm{Var}(X)$ *is sometimes denoted by* σ^2.

2. *The* standard deviation σ *of* X *is defined by* $\sigma = \sqrt{\mathrm{Var}(X)}$.

Example 2.2.6 *Find the variance and standard deviation of the random variable* X *of Example 2.2.1.*

SOLUTION: From Example 2.2.1, $E(X) = 1.25$. Then

$$\begin{aligned}
\mathrm{Var}(X) &= E((X - E(X))^2) \\
&= E((X - 1.25)^2) \\
&= \sum_{k=0}^{5} (k - 1.25)^2 p_X(k) \\
&\approx 0.8639 \text{ (Check)}. \\
\sigma &= \sqrt{\mathrm{Var}(X)} \approx 0.9294
\end{aligned}$$

NOTE: In general if X is hypergeometric random variable with parrameters n (total number of items), m (number of "good" items) and k (number of items in a selection), then the variance of X is given by

$$\boxed{\mathrm{Var}(X) = k\frac{m}{n}\left(1 - \frac{m}{n}\right)\frac{n-k}{n-1}}$$

Problem 8: If X is the score on a fair die, calculate $\mathrm{Var}(X)$.

Variance and standard deviation measure the "spread" of probability of X and expectation measures its "centre". A low value for σ or $\mathrm{Var}(X)$ means that values of X are likely to be close to $E(X)$.

Example 2.2.7 *Suppose that X is the number of successes in a single Bernoulli trial with probability of success p. What is the variance of X?*

SOLUTION: We know that $E(X) = p$ and that X has only two possible values, 0 and 1, occuring with probabilities $1 - p$ and p respectively. Thus

$$\begin{aligned} \mathrm{Var}(X) &= E((X - E(X))^2) \\ &= (0 - p)^2 P(X = 0) + (1 - p)^2 P(X = 1) \\ &= p^2(1 - p) + (1 - p)^2 p \\ &= p(1 - p) \end{aligned}$$

Suppose that X is a binomial random variable with paramaters n and p. So X is the number of successes in n trials, each with probability p of success.
S : Sequences of s and f of length n.
For $i = 1, \dots, n$, define $X_i : S \longrightarrow \mathbb{R}$ by

$$X_i = \begin{cases} 1 & \text{if} \quad s \text{ in } i\text{th trial} \\ 0 & \text{if} \quad f \text{ in } i\text{th trial} \end{cases}$$

e.g. if $n = 3$, $X_1(ssf) = 1$, $X_2(ssf) = 1$, $X_3(ssf) = 0$, and $X(ssf) = 2 = 1 + 1 + 0$.
Then $X = X_1 + X_2 + \cdots + X_n$.
NOTE: The random variables $X_1, X_2, \dots, X_n$ are *mutually independent* : this means that the probabilty of one of them assuming a particular value is unaffected by the values assumed by others in an instance of the experiment.
Mutually independent random variables have the following property : *the variance of their sum is the sum of their variances*. Thus if X is a binomial random variable with parameters n and p, then

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

and

$$\mathrm{Var}(X) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2) + \cdots + \mathrm{Var}(X_n) = np(1 - p).$$

Example 2.2.8 *Suppose that X is the number of university graduates in a random selection of 12 people from the population, and that university graduates account for 18% of the population. Find the expectation, variance and standard deviation of X.*

SOLUTION: X is a binomial random variable with parameters $n = 12$ and $p = 0.18$. Thus

- $E(X) = np = 12 * 0.18 = 2.16$.
- $\mathrm{Var}(X) = \sigma^2 = np(1 - p) = 12 * 0.18 * 0.82 = 1.772$

- $\sigma = \sqrt{\sigma^2} \approx 1.331.$

OTHER "STANDARD" DISCRETE RANDOM VARIABLES

1. X: Geometric Random Variable with parameter p
   X is the number of trials needed to get a 1st success.
   Probability Distribution : $p_X(k) = q^{k-1}p, \text{ for } k = 1, 2, 3, \ldots.$

$$\boxed{E(X) = \frac{1}{p} \quad \text{and} \quad Var(X) = \frac{1-p}{p^2}}$$

2. X: Negative Binomial Random Variable with parameters r and p
   X is the number of trials to get the rth success.

$$\boxed{E(X) = \frac{r}{p} \quad \text{and} \quad Var(X) = \frac{r(1-p)}{p^2}}$$

Chapter 3

Continuous Random Variables

3.1 Cumulative Distribution and Density Functions

Definition 3.1.1 *A random variable is called* continuous *if it can take values anywhere in an interval of the real numbers.*

Continuous random variables do not have probability functions in the sense discussed in Chapter 2. To discuss probability theory for continuous random variables, we need the idea of a *cumulative distribution function*.
Let S be a sample space and let $X : S \longrightarrow \mathbb{R}$ be a random variable on S.

Definition 3.1.2 *The* cumulative distribution function *(cdf) of* X *is the function* $F_X : \mathbb{R} \longrightarrow [0, 1]$ *defined by*

$$F_X(x) = P(X \leqslant x), \text{ for } x \in \mathbb{R}.$$

$F_X(x)$ *is the probability that* X *will assume a value less than or equal to* x.

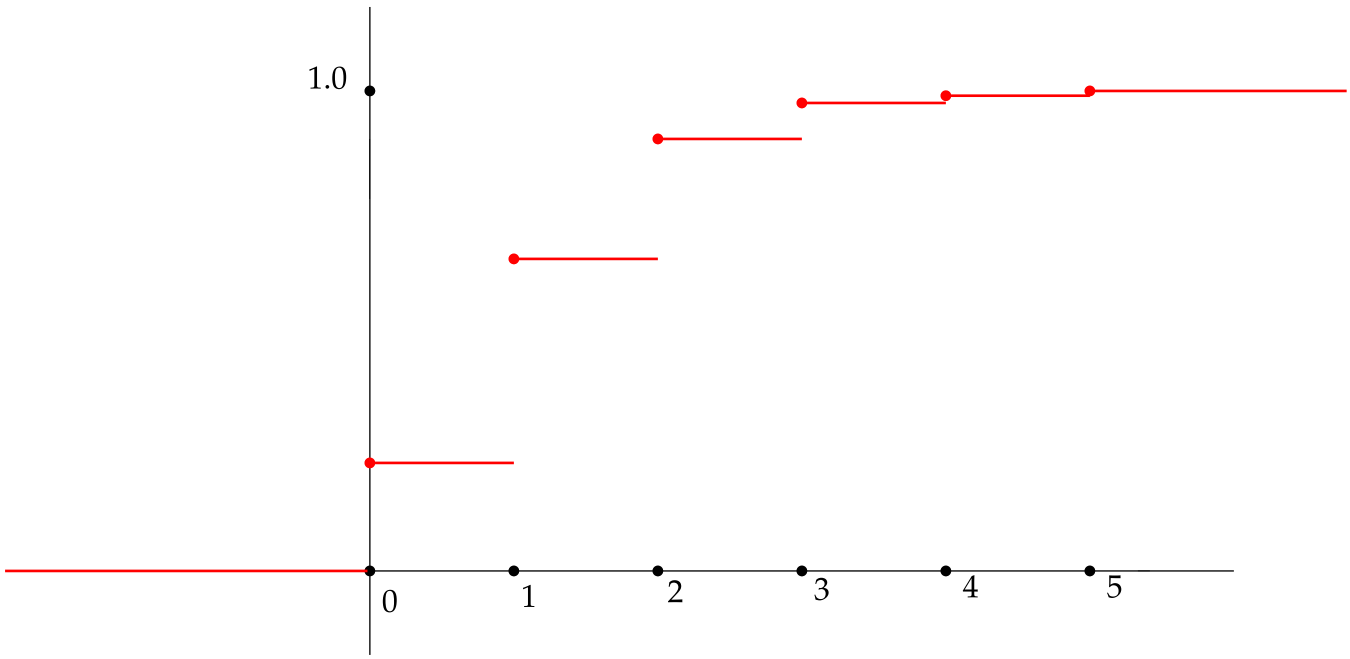
Example 3.1.3 *(Recall Example 2.2.1)*
X *: No. of clubs in a poker hand.*

k	0	1	2	3	4	5
$P(X = k)$	0.2215	0.4114.	0.2743	0.0816	0.0107	0.0005

The cdf for X is given by the following description :

$$F_X(k) = \begin{cases} 0 & \text{for } k < 0 \\ 0.2215 & \text{for } 0 \leqslant k < 1 \\ 0.6329 & \text{for } 1 \leqslant k < 2 \\ 0.9072 & \text{for } 2 \leqslant k < 3 \\ 0.9888 & \text{for } 3 \leqslant k < 4 \\ 0.9995 & \text{for } 4 \leqslant k < 5 \\ 1 & \text{for } k \geqslant 5 \end{cases}$$

Graph of F_X:



<u>Notes</u>:

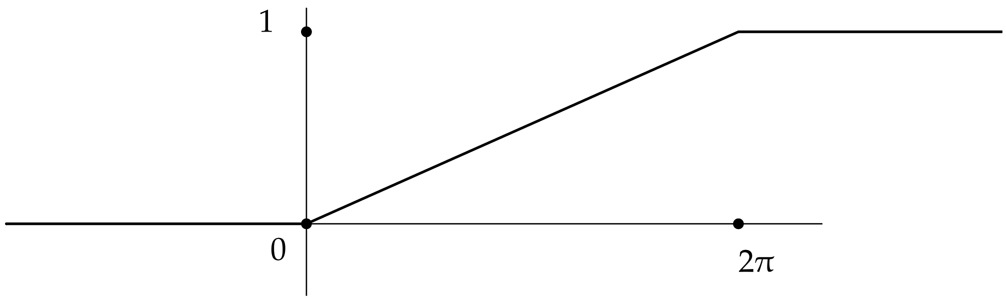
1. $F_X(x) = 0$ for $x < 0$: X cannot be negative.
   In general, $\lim_{t \to -\infty} = 0$ for any cdf F_X.

2. $F_X(x) = 1$ for $x \geqslant 5$: X will always be $\leqslant 5$.
   In general $\lim_{t \to \infty} = 1$ for any cdf F_X.

3. F_X is a non-decreasing function on $\mathbb{R}$. This is a general property of cumulative distribution functions.

4. F_X is discontinuous at $X = 0, 1, 2, 3, 4, 5$. This is typical of the cdf of a *discrete* random variable. A random variable X is referred to as *continuous* if its cumulative distribution function is continuous on $\mathbb{R}$.

Example 3.1.4 *A vertically mounted wheel is spun and allowed to come to rest. Let* X *be the angle (measured counterclockwise) between its initial position and final position. Then* $0 \leqslant X < 2\pi$ *and the cdf for* X *is given by:*



X is equally likely to assume any value in the interval $[0, 2\pi)$, so $F_X(x)$ increases uniformly from 0 to 1 on $[0, 2\pi]$; the graph of F_X on $[0, 2\pi]$ is a *line*.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2\pi}x & 0 \leqslant x \leqslant 2\pi \\ 1 & x > 2\pi \end{cases}$$

X is an example of a *uniform* continuous random variable.

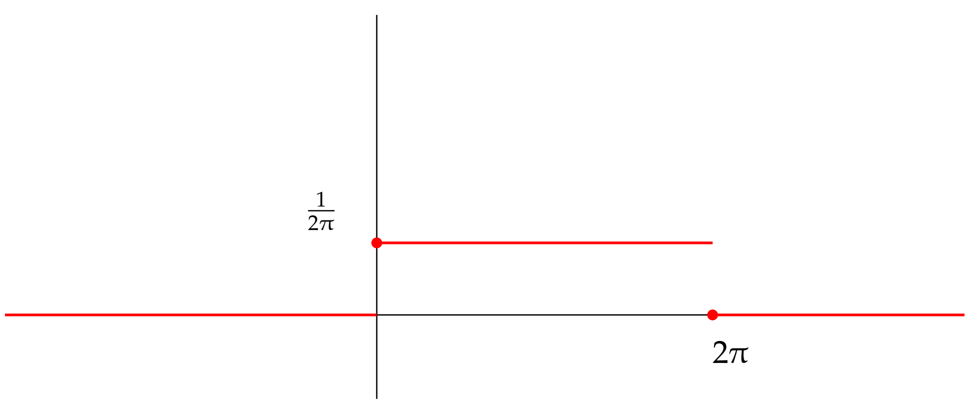
NOTE: Define a function $f : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$f(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{2\pi} & 0 \leqslant x < 2\pi \\ 0 & x \geqslant 2\pi \end{cases}$$

Note that for every $x \in \mathbb{R}$

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

The function f is called a *probability density function* (pdf) for X. Its graph, which is shown below, reflects the fact that X always assumes a value in the interval $[0, 2\pi)$ and that all values in this interval are equally likely.



Problem 3.1.5 *Let A, B be real numbers with $A < B$. Suppose X is a continuous random variable which assumes a value in the interval $[A, B]$, all values in this interval being equally likely. Write down a formula and graph for the cdf of X (X has the continuous uniform distribution).*

Definition 3.1.6 *Let X be a continuous random variable with cdf F_X. A* probability density function *(pdf) for X is a function $f : \mathbb{R} \longrightarrow [0, \infty)$ for which*

$$F_X(x) = \int_{-\infty}^{x} f(t)dt, \text{ for } x \in \mathbb{R}.$$

i.e. for every $x \in \mathbb{R}$, $P(X \leqslant x)$ is the area under the graph of $f(t)$, to the left of x.

NOTES:

1. If f is a pdf for X, then
$$\int_{-\infty}^{\infty} f(t)dt = 1.$$
(This follows from $\lim_{x\to\infty} F_X(x) = 1$).

2. If F_X is differentiable for some random variable X, then its derivative F'_X is a pdf for X (this follows from the Fundamental Theorem of Calculus).

3. Suppose $a, b \in \mathbb{R}$, $a < b$. Then
$$\begin{aligned} P(a \leqslant X \leqslant b) &= P(X \leqslant b) - P(X \leqslant a) \\ &= F_X(b) - F_X(a) \\ &= \int_{\infty}^{b} f(f)dt - \int_{-\infty}^{a} f(t)dt \\ &= \int_{a}^{b} f(t)dt. \end{aligned}$$
$$P(a \leqslant X \leqslant b) = \int_{a}^{b} f(t)dt.$$

Expectation and Variance of a Continuous Random Variable

Definition 3.1.7 *Let X be a continuous random variable with pdf f. Then the expectation of X is defined by*

$$E(X) = \int_{-\infty}^{\infty} tf(t)dt.$$

Problem 3.1.8 *If X is as in Example 3.1.4, show that $E(X) = \pi$.*

Definition 3.1.9 $\mathrm{Var}(X) = E((X - E(X))^2)$*: variance of* X*. This is also given by* $E(X^2) - (E(X))^2$ *(this is true also in the discrete case).*

Fact: $E(X^2) = \int_{-\infty}^{\infty} t^2 f(t)\,dt$. So

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} t^2 f(t)\,dt - (E(X))^2 = \int_{-\infty}^{\infty} t^2 f(t)\,dt - \left(\int_{-\infty}^{\infty} t f(t)\,dt\right)^2.$$

The *standard deviation* σ of X is defined by $\sigma = \sqrt{\mathrm{Var}(X)}$.

Problem 3.1.10 *Find the variance and standard deviation of the random variable* X *of Example 3.1.4.*

3.2 The Normal Distribution

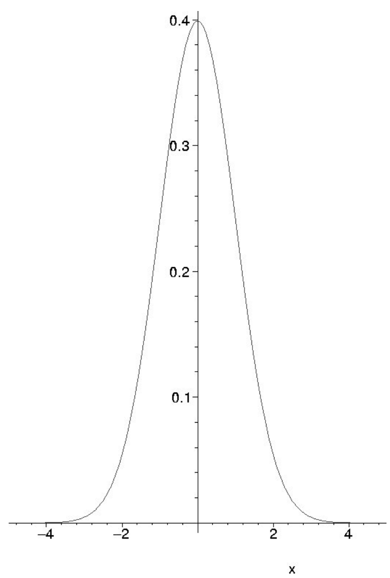
Consider the function $f : \mathbb{R} \longrightarrow \mathbb{R}$ defined by

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

This function has the following properties:

- It is *even* (i.e. $f(-x) = f(x)$ for $x \in \mathbb{R}$)
- $\lim_{x \to \infty} f(x) = 0.$
- $\lim_{x \to -\infty} f(x) = 0.$

The graph of f looks like :



REMARK: It is possible using polar coordinates to show that

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

Definition 3.2.1 *A continuous random variable having the above* f *as a pdf is said to have the* standard normal distribution. *Such a random variable is said to have distribution* $N(0, 1)$.

Problem 3.2.2 *If* X *has distribution* $N(0, 1)$, *show that* $E(X) = 0$ *and* $\mathrm{Var}(X) = 1$.

Outline of Solution:

1. $\mathsf{E}(X) = \int_{-\infty}^{\infty} t \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \right) dt.$
   Note that the integrand here is an *odd* function, hence the integral is zero (provided that the improper integral converges which is not difficult to check). Thus $\mathsf{E}(X) = 0$.

2. $\text{Var}(X) = \int_{-\infty}^{\infty} t^2 \left(\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} \right) dt - (\mathsf{E}(X))^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^2 e^{-\frac{1}{2}t^2} dt.$
   Integrating by parts with $u = t$, $v' = te^{-t^2}$ gives

$$\text{Var}(X) = -\frac{1}{\sqrt{2\pi}} te^{-\frac{1}{2}t^2} \Big|_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}t^2} dt = 0 + 1 = 1.$$

Example 3.2.3 *Suppose that* X *is a random variable having the distribution* $\mathsf{N}(0, 1)$. *Find*

(a) $\mathsf{P}(X \leqslant 2.0)$

(b) $\mathsf{P}(X > 2.5)$

(c) $\mathsf{P}(X < -0.5)$

(d) $\mathsf{P}(-1.0 \leqslant X \leqslant 2.0)$

Solution: Let $\Phi : \mathbb{R} \longrightarrow [0, 1]$ be the cdf of X. (The symbol Φ is the upper case of the Greek letter *phi*). For part (a), we know from the definition of a pdf that

$$\mathsf{P}(X \leqslant 2.0) = \Phi(2.0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{2} e^{-\frac{t^2}{2}} dt.$$

PROBLEM: The expression $e^{-\frac{t^2}{2}}$ does not have an antiderivative expressible in terms of elementary functions. So to estimate the value of this integral we need to use a numerical approximation. A table of values of $1 - \Phi(x)$ for $0 \leqslant x < 3$, in increments of 0.01, is provided for this purpose.

(a) From the table, $\mathsf{P}(X \leqslant 2.0) = 1 - \mathsf{P}(X > 2.0) \approx 1 - 0.0228$ (from looking up $z = 2$ in the table. So

$$\mathsf{P}(X \leqslant 2) \approx 1 - 0.0228 = 0.9772.$$

(b) $\mathsf{P}(X > 2.5) \approx 0.0062$: directly from the table.

(c) By symmetry, $P(X < -0.5) = P(X > 0.5)$. This can be read from the table at $z = 0.5$. So

$$P(X < -0.5) = 0.3085.$$

(d)

$$\begin{aligned}
P(-1.0 \leqslant X \leqslant 2.0) &= P(X \leqslant 2) - P(X < 1) \\
&= (1 - P(X \geqslant 2)) - P(X > 1) \\
&\approx (1 - 0.0228) - (0.1587) \\
&= 0.8185.
\end{aligned}$$

Problem 3.2.4 *If X has the distribution $N(0, 1)$, find the following :*

(a) $P(X \leqslant -0.7)$

(b) $P(-0.6 \leqslant X \leqslant 0.8)$

(c) $P(X \leqslant -0.052)$

INVERSE PROBLEM: Suppose X has distribution $N(0, 1)$. Find that real number x for which $P(X \leqslant x) = 0.8$ (i.e. find that x for which we expect X to be $\leqslant x$ in 80% of cases).

SOLUTION: We want that value of x for which the area under the graph to the left of x is 0.8, and the area to the right is 0.2. So look in the right hand side column of the table for 0.2000. We have 0.2005 when $z = 0.84$, and 0.1977 when $z = 0.85$. We can estimate the value x corresponding to 0.2 as follows :

$$\begin{aligned}
x &\approx 0.84 + \frac{0.2005 - 0.2000}{0.2005 - 0.1997}(0.85 - 0.84) \\
&\approx 0.8418.
\end{aligned}$$

Problem 3.2.5 *Find that real number x for which $P(|X| \leqslant x) = 0.8$, where X has the standard normal distribution.*

Note: $|X| \leqslant x$ means $-x \leqslant X \leqslant x$: so we need $P(-x \leqslant X \leqslant x) = 0.8$, meaning that by symmetry we must have $P(X < -x) = P(X > x) = 0.1$. We want $X > x$ in 10% of cases.

Other Normal Distributions

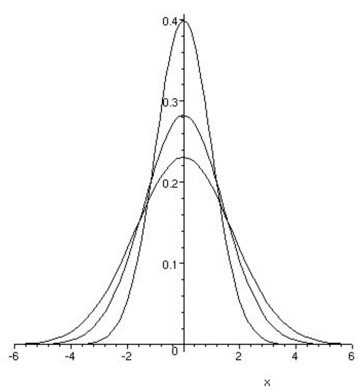
Random variables having normal distributions arise frequently in practice. Of course, not all have expected value 0 and variance 1.

Definition 3.2.6 *A random variable* X *is said to have the distribution* $N(\mu, \sigma^2)$ *if* X *has pdf* f_X *given by*

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

where $\mu \in \mathbb{R}$ *and* $\sigma > 0$ *are fixed.*

REMARK: If $\sigma = 1$, then $f_X(x) = f(x - \mu)$, where f is the pdf of the the $N(0, 1)$ distribution. So the graph of $f_X(x)$ in this case is just that of $f(x)$ moved μ units to the right.
So if X has distribution $N(\mu, \sigma^2)$ then X has expected value μ and variance σ^2. For fixed μ and σ, the graph of $f_X(x)$ is a "bell-shaped" curve, in which the value of $f_X(x)$ is negligible for X more than about 3σ away from μ. The graph is symmetric about $x = \mu$, and its "flatness" is controlled by σ, larger values of σ giving a wider, flatter curve. This makes sense, since a large value of σ means a high standard deviation. The following diagram shows the graphs for $\mu = 0$ corresponding to $\sigma = 1, 2$ and 3.



Example 3.2.7 *The height (in metres) of a person selected at random is known to be normally distributed with a mean (i.e. expected value) of* 1.65 *and a standard deviation of* 0.12.

(a) What is the probability that a person's height will exceed 1.75?

(b) What is the probability that a person's height will be between 1.60 and 1.75?

(c) Above what height can we expect to find the tallest 5% of the population?

Problem: We only have tables for $N(0, 1)$. What can we do with $N(1.65, (0.12)^2)$?

Theorem 3.2.8 *Suppose* X *has the distribution* $N(\mu, \sigma^2)$. *Define* $Y = \dfrac{X-\mu}{\sigma}$. *Then* Y *has the distribution* $N(0,1)$.

So: we work with $Y = \dfrac{X-\mu}{\sigma}$ instead of X.

Back to the Example: Let X be the height of a randomly selected person. So $X \sim N(1.65, (0.12)^2)$. Then $Y = \dfrac{X-1.65}{0.12}$ has distribution $N(0,1)$.

(a) $X > 1.75 \implies X - 1.65 > 0.1 \implies \dfrac{X-1.65}{0.12} > \dfrac{0.1}{0.12} \implies Y > 0.8333$. So we need $P(N(0,1) > 0.83$. From the tables we can interpolate

$$P(X > 1.75) \approx 0.2033 - \frac{1}{3}(0.2033 - 0.2005) \approx 0.2024.$$

(b) $$\begin{aligned} P(1.6 \leqslant X \leqslant 1.75) &= P(-0.05 \leqslant X - 1.65 \leqslant 0.1) \\ &= P(-0.4166 \leqslant Y \leqslant 0.8333) \end{aligned}$$

Linearly interpolating again, we estimate

$$P(1.6 \leqslant X \leqslant 1.75) \approx 0.4592.$$

(c) From the table we can read that for the tallest 5% we have $Y \geqslant 1.645$. Then

$$\frac{X-1.65}{0.12} \geqslant 1.6449 \implies X - 1.65 \geqslant 0.1974 \implies X \geqslant 1.85$$

The tallest 5% of the population will be taller than 1.85m.

Problem 3.2.9 *Assume that the flight time (from takeoff to landing) from Dublin Airport to London Heathrow is normally distributed with a mean of 50 minutes and a standard deviation of 5 minutes.*

(a) What is the probability that the flight time will exceed one hour?

(b) What is the probability that the flight time will be between 45 and 55 minutes?

(c) Below what duration can we expect the fastest 10% of flights?

REMARK/EXERCISE (cf (b) above) : If X has distribution $N(\mu, \sigma^2)$ then the probability that X will be between $\mu - \sigma$ and $\mu + \sigma$ (i.e. within one standard deviation of the mean) is

$$P(-1 \leqslant N(0,1) \leqslant 1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 0.6826.$$

i.e. we expect 68% of observed values to be within one standard deviation of the mean. Similarly,

$$P(\mu - 2\sigma \leqslant X \leqslant \mu + 2\sigma) = P(-2 \leqslant N(0,1,) \leqslant 2) \approx 0.944.$$

For a normal distribution we expect 94.4% of observed values to be within two standard deviations of the mean.

Chapter 4

The Central Limit Theorem

4.1 Random Samples

Recall from the first lecture of the course that *Statistics* is about trying to infer information about a population from knowledge of a sample from that population. A typical example is opinion polling : suppose a referendum is taking place on the Lisbon Treaty. An opinion poll involves interviewing 100 people and asking them how they plan to vote. Assuming that the sample of 100 voters was genuinely random, what does the result of the poll tell us about the likely outcome of the referendum? This is a mathematical question.
The theme of this section is the *Central Limit Theorem*, which explains why the normal distribution is so important.

Definition 4.1.1 *(Random Sample) Suppose* X *is a random variable associated to some experiment. If the experiment is performed* n *times (under identical conditions) we have independent random variables* $X_1, X_2, \dots, X_n$ *associated to repetitions 1 to* n*, each having the same distribution as* X*. Then* $(X_1, \dots, X_n)$ *is called a* random sample *of* X.

The random variable $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ is called the *sample mean*.

Example 4.1.2 *If* X *is the height of a randomly selected person, let* X_i *be the height of the* i*th person selected, for* $i = 1, \dots, 10$*. Then* $(X_1, \dots, X_{10})$ *is a random sample of* X *and* $\bar{X}$ *is the average height of the first 10 people selected.*

Theorem 4.1.3 *(The Central Limit Theorem) Let* X *be a random variable with mean* μ *and variance* σ^2*. Let* $X_1, \dots, X_n$ *be a random sample of* X *and let* $\bar{X}(n) = \frac{1}{n}(X_1 + \cdots + X_n)$ *be the sample mean. The as* $n \longrightarrow \infty$*, the cdf of* $\bar{X}$ *approaches that of a random variable with distribution* $N(\mu, \sigma^2/n)$.

i.e. for large n, $\bar{X}$ has approximately the distribution $N(\mu, \sigma^2/n)$, so $\dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}}$ has approximately the distribution $N(0, 1)$.

REMARKS:

1. It is not really clear from the above what "large n" means or what "approximately" means. The answer depends on the distribution of the underlying variable X. If X is itself a *normal* random variable with mean μ and variance σ^2, then $\bar{X}$ has the distribution $N(\mu, \sigma^2/n)$ regardless of the value of n. If X is not itself normally distributed, then $n \geqslant 30$ is generally considered a large enough sample size to model $\bar{X}$ as a normal random variable with distribution $N(\mu, \sigma^2/n)$ in practice.

2. The Central Limit Theorem says that for large samples the sample mean $\bar{X}$ is approximately a normal random variable with mean μ (the population mean) and variance σ^2/n. That the variance decreases as n gets larger is not surprising; we would expect that the average of a large sample should more likely to be close to the population mean than that of a small sample.

Example 4.1.4 *Suppose that the volume X of water in a bottle is a normal random variable with mean 1 litre and standard deviation 10ml. Find the probability that in a random sample of 16 bottles the average volume will exceed 1005ml.*

SOLUTION: $\bar{X}$ has distribution $N(1000, 100/16)$, so

$$Y = \frac{\bar{X} - 1000}{10/4} \sim N(0, 1).$$

Thus $P(\bar{X} > 1005) = P(\bar{X} - 1000 > 5) = P(Y > 2) = 1 - \Phi(2) = 0.0228$.

QUESTION: Let X be a random variable with mean/expected value μ (which is unknown) and variance σ^2 (which is known). Let $(X_1, \dots, X_n)$ be a random sample of X. What does the value of the sample mean $\bar{X}$ tell us about the value of μ?
ANSWER: Nothing with certainty, since the values in our sample might be nowhere near μ. However, since we know that $\bar{X}$ has approximately a normal distribution with mean μ and variance depending on the population variance σ^2 and the sample size n, we can estimate the probability that μ will be within a fixed distance of $\bar{X}$.

Definition 4.1.5 *In the above context, an interval $[A, B]$ of the real line is a* $q\%$ confidence interval *for μ if*

$$P(A \leqslant \mu \leqslant B) = \frac{q}{100}.$$

i.e. if μ lies in the interval $[A, B]$ with probability $\frac{q}{100}$.

Example 4.1.6 *Let X be the level of carbon monoxide (CO) emissions (in grammes/mile) of a randomly selected car of a certain make. It is known that the variance of X is 0.09. A sample of carbon monoxide emissions (in grammes/mile) from 8 cars gave the following results :*

$$17.3,\ 17.8,\ 18.0, 17.7,\ 18.2,\ 17.4,\ 17.6,\ 18.1.$$

1. *Based on this sample, find a 95% confidence interval for the mean CO emission.*
2. *Find the sample size required for a 95% confidence interval of less than 0.2g/mile for μ.*

SOLUTION:

(a) We know $\bar{X}$ is (approximately) normally distributed with mean μ and variance σ^2/n. Then

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

From the tables

$$\begin{aligned} P\left(-1.96 \leqslant \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leqslant 1.96\right) &= 0.95 \\ \Longrightarrow P(-1.96\sigma/\sqrt{n} \leqslant \bar{X}-\mu \leqslant 1.96\sigma/\sqrt{n}) &= 0.95 \\ \Longrightarrow P(|\bar{X}-\mu| \leqslant 1.96\sigma/\sqrt{n}) &= 0.95. \end{aligned}$$

Thus $\bar{X}$ is within $1.96\sigma/\sqrt{n}$ of μ with probability 0.95. Our 95% confidence interval for μ is given by

$$[\bar{X}-1.96\sigma/\sqrt{n}, \bar{X}+1.96\sigma/\sqrt{n}].$$

In our example, $\sigma = \sqrt{0.09} = 0.3$, $n = 8$ (sample size), and $\bar{X} = 17.76$ (average of the eight observations).
95% confidence interval for μ :

$$[17.76 - 1.96 \times (0.3)/\sqrt{8}, 17.76 + 1.96 \times (0.3)/\sqrt{8}] = [17.55, 17.97].$$

(b) For a sample of size n, the length of our 95% confidence interval for μ is $2 \times 1.96\sigma/\sqrt{n}$. So for an interval of less than 0.2 we require

$$\begin{aligned} 2\left(1.96\sigma/\sqrt{n}\right) &< 0.2 \\ \Longrightarrow \left(1.96 \times 0.3/\sqrt{n}\right) &< 0.1 \\ \Longrightarrow (1.96 \times 0.3) &< \sqrt{n} \times 0.1 \\ \Longrightarrow \sqrt{n} &> 1.96 \times 0.3/0.1 = 5.88 \\ \Longrightarrow n &> 34.6. \end{aligned}$$

So a sample size of at least 35 is required.

Problem 4.1.7 *Repeat the three parts of the above problem for*

(a) a 99% confidence interval for μ.

(b) a 80% confidence interval for μ.