Probability and Statistics I

Introduction to probability

Probability is a branch of mathematics that deals with calculating the chance of a given event's occurrence.

It is expressed as a number between 0 and 1 where:

- a number close to 0 means not likely
- a number close to 1 means quite likely

If the probability of an event is exactly 0, then the event cannot occur. For example, if today is Monday the probability that tomorrow is Sunday is 0 - an uncertain occurrence.

If the probability of an event is exactly 1, then the event will definitely occur. If today is Monday the probability that tomorrow is Tuesday is 1 - a certain occurrence.

Hence we now know that probability is a number between 0 and 1.

Approaches to Probability

Our problem is then: how does an event get assigned a particular probability value?

There are three approaches to (interpretations of) probability namely:

1. The subjective approach
2. The relative frequency approach
3. The classical approach

The subjective approach

This approach is the simplest in practice, but therefore it also the least reliable. Think of it as the 'whatever it is to you' approach. Here are some examples:

- 'I think there is an 80% chance of rain today.'
- 'I think there is a 50% chance that the world's oil reserves will be depleted by the year 2100.'
- 'I think there is a 1% chance that the Harambee Stars will end up in the World Cup.'

Example 1

At which end of the probability scale would you put the probability that:

1. One day you will die?

2. You can swim around the world in 30 hours?

3. You will win the lottery some day?

4. A randomly selected student will get an A in this course?

5. You will get an A in this course?

The relative frequency approach

Here probability is interpreted to mean the relative frequency with which the outcome will be obtained if the process were repeated a large number of times under similar conditions.
The relative frequency approach involves taking the following three steps in order to determine $p(A)$, the probability of an event A:

1. Perform an experiment a large number of times, n, say.

2. Count the number of times the event A of interest occurs, call the number $N(A)$, say.

3. Then, the probability of event A equals:

$$p(A) = \frac{N(A)}{n}$$

Arguments against the relative frequency approach

- ✓ There is no definite indication of an actual number that would be considered large enough.

- ✓ If you toss your coin 1,000,000 times, do you get 500,000 heads? (No limit is specified for the permissible variation from 0.5).

- ✓ Besides, if a process is repeated under similar conditions – you should expect the same outcome

- ✓ Not all processes can be repeated a large number of times e.g. Tom and Mary are having an affair; what is the probability that they will get married in two years?

Example 2

Some trees in a forest were showing signs of disease. A random sample of 200 trees of various sizes was examined yielding the following results:

Type	Disease free	Doubtful	Diseased	Total
Large	35	18	15	68
Medium	46	32	14	92
Small	24	8	8	40
Total	105	58	37	200

a) What is the probability that one tree selected at random is large?

b) What is the probability that one tree selected at random is diseased?

c) What is the probability that one tree selected at random is both small and diseased?

d) What is the probability that one tree selected at random is either small or disease-free?

e) What is the probability that one tree selected at random from the population of medium trees is doubtful of disease?

The classical approach

Here the concept of probability is based on the concept of equally likely outcomes e.g. when a coin is tossed, there are two (2) possible outcomes – a Head and a Tail.
If the two outcomes are equally likely, then they must have the same probability of occurrence. Since the total probability is 1 then

$$p(\text{Head}) = p(\text{Tail}) = \frac{1}{2}$$

As long as the possible outcomes are equally likely (!!!), the probability of event A is:

$$p(A) = \frac{p(N(A))}{p(N(S))}$$

where

$N(A)$ - the number of elements in the event A, and

$N(S)$ - the number of all possible elements.

Arguments against the classical approach

✓ What if the outcomes are not equally likely, how do we assign probabilities? e.g. a student sits for an exam, what are the chances that they pass? – Passing and Failing are not equally likely outcomes. We therefore require other methods to assign probability.

✓ Besides the concept of equally likely outcomes is essentially based on the concept of probability that we are trying to define – the statement that the two possible outcomes are equally likely is the same as the statement that the two outcomes have the same probability.

Example 2

Suppose you draw one card at random from a standard deck of 52 cards. Assume the cards were manufactured to ensure that each outcome is equally likely.

☐ Let A be the event that the card drawn is a 2, 3, or 7.

☐ Let B be the event that the card is a 2 of hearts (H), 3 of diamonds (D), 8 of spades (S) or king of clubs (C).

a) What is the probability that a 2, 3, or 7 is drawn?

b) What is the probability that the card is a 2 of hearts, 3 of diamonds, 8 of spades or king of clubs?

c) What is the probability that the card is either a 2, 3, or 7 or a 2 of hearts, 3 of diamonds, 8 of spades or king of clubs?

d) What is the probability that event A and B occurs?

Note however that:

✓ probability will be the way we quantify how likely something is to occur.

✓ the theory of probability does not depend on the interpretation or approach to probability.

The Mathematical Theory of Probability

Once probabilities have been assigned to the outcome of a process, the mathematical theory of probability provides the appropriate methodology for the further studies of these probabilities namely:

1. Methods of determining the probabilities of certain events from the specified probability of each possible outcome of an experiment e.g. given $p(A)$ and $p(B)$, find $p(A \quad \text{and} \quad B)$ and
2. Methods of revising the probabilities of events when additional relevant information is obtained.

The problem is twofold.

Basic Concepts in Probability Theory

To understand the language of probability, we begin by defining basic concepts used in probability theory.

Definition of Terms

Experiment: A repeatable process which gives rise to a number of known possible outcomes e.g.

- ✓ Tossing a coin is an experiment whose outcomes are Head (H) or Tail (T);
- ✓ Tossing two coins – HH, HT, TH, TT;
- ✓ Throwing a die – 1, 2, 3, 4, 5, 6;
- ✓ Throwing two die – $(1,1)$, $(1,2)$, $(1,3)$, ..., $(6,6)$; 36 possible outcomes;
- ✓ Throwing a die and tossing a coin – $1H$, $2H$, $3H$, $4H$, $5H$, $6H$, $1T$, $\cdots$, $6T$; 12 possible outcomes;
- ✓ Picking a student from a MAS103 class – Male or Female.

Random Experiment: In probability theory we start off with an experiment whose result cannot be determined beforehand but we know all the possible outcome.
An experiment is said to be random if its result cannot be determined beforehand.
Sample Space: The sample space is denoted by S or Ω. It refers to the set (or a collection) of all the possible outcomes of an experiment e.g.

- ✓ Tossing a coin, $S = \{\text{Head}, \text{Tail}\}$ or $S = \{H, T\}$;
- ✓ Tossing two coins $S = \{HH, HT, TH, TT\}$;

Each possible outcome of a random experiment is called a **sample point**.

Since we have random experiments of many types, the set Ω may consist of:

i) A finite number of elements

ii) Countably infinite number of elements

iii) Uncountably many elements

The elements could be numerical or non-numerical.

Example 1

Consider the following random experiments denoted by E. Find the sample space and state whether they have numerical/non-numerical, finite/countably infinite/uncountably many elements.

✓ E: Tossing a coin;

✓ E: Rolling a dice;

✓ E: Number of on-going calls in a particular telephone exchange;

✓ E: Temperature of a given city;

An Event is denoted with capital letters A, B, C ... is a subset of the sample space S. That is, $\mathsf{A} \subset \mathsf{S}$, where '$\subset$' denotes 'is a subset of' – meaning is contained in.

Consider the experiment of throwing two dice. The sample space is:

	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

An event A - getting a sum of 7; or the event B - getting a sum of 5; or the event C - getting a sum of 2;

Simple and Compound Events: An event is said to be a simple event if it consists of one outcome otherwise it is a compound event e.g. when tossing two die event C – getting a sum of two is a simple event while event A above is a compound event.

NB: The probability of an event A occurring is denoted $p(A)$

Equally Likely Events: These are events that have the same theoretical probability (or likelihood) of occurring. For example: Each numeral on an unbiased die is equally likely to occur when the die is tossed. Sample space of throwing a die: $\{1, 2, 3, 4, 5, 6\}$.

Note:

- ✓ We say that a particular event A has occurred if the outcome of the experiment is a member of (i.e. contained in) A.
- ✓ The sample space of an experiment is an event which always occurs when the experiment is performed. This implies $p(S) = 1$.

ALGEBRA OF SETS

Since events and sample spaces are just sets, let us review the algebra of sets to enable us apply them when calculating the probability of events.

What is a set?

A set is a well-defined collection of objects called the elements or members of the set. They are denoted using capital letters just like events.

The algebra of sets defines the properties and laws of sets, the set-theoretic operations of union, intersection, and complementation and the relations of set equality and set inclusion,(*Wikipedia*).

The following laws will help us answer probability questions.

1. ϕ denotes the empty or null set - the set that has no elements.
2. $C \cup D$ read C union D defines the elements in C or D or both; It is the event consisting of all outcomes that are in C or in D or in both.
3. $A \cap B$ read A intersection B defines the elements in A and B; It is the event consisting of all outcomes that are in both A and B.
4. If $A \cap B = \phi$, then A and B are called mutually exclusive events (or disjoint events) – A and B have no outcomes in common.
5. The complement, $D^{'}$ of D also denoted D^c is the event consisting of all outcomes that are not in D but are in the sample space.

6. If $E \cap F \cap G \cap ... = S$, then E, F, G, and so on are called exhaustive events.

7. The difference $A–B$ is defined as $A \cap B^c$ - the event that both A and B^c occur.

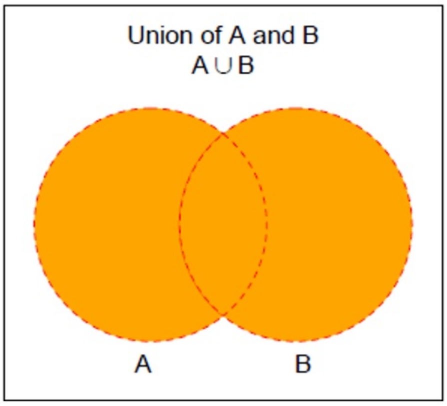
8. A is a subset of B, denoted $A \subset B$, if $e \in A$ implies $e \in B$.

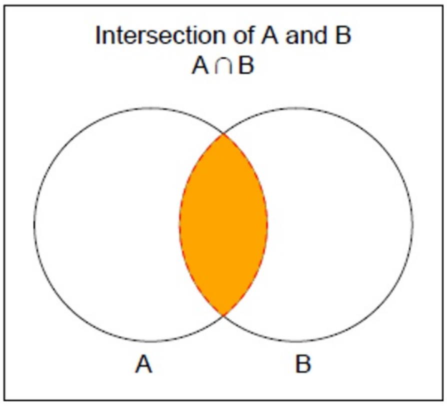
9. Two sets are equal, $A = B$, if $A \subset B$ and $B \subset A$.

VENN DIAGRAMS

Venn Diagrams provided a pictorial representation of sets hence events.

For Union and Intersection

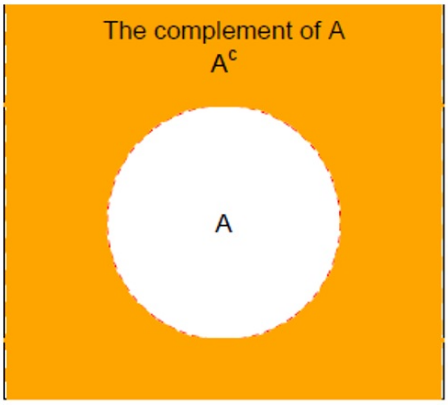


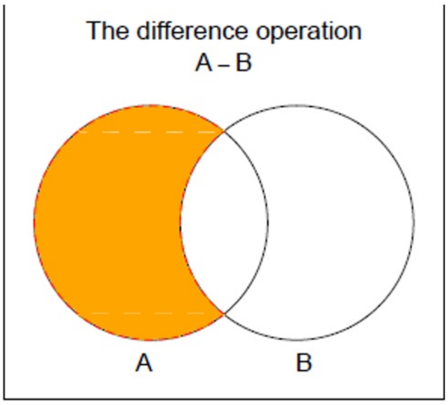


The event $A \cup B$ denotes the event that either A or B or both occur.

The event $A \cap B$ denotes the event that both A and B occur.

For complement and difference

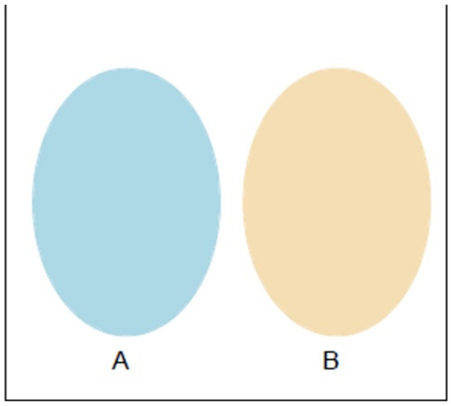


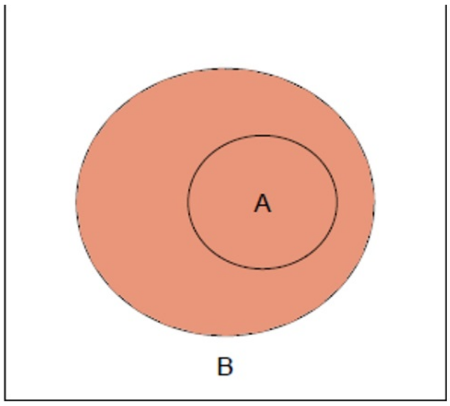


The event A^c denotes the event that A does not occur.

The event $A - B$ denotes the event $A \cap B^c$, the event that both A and B^c occur.

For A, B disjoint and A subset of B





For A and B disjoint, $A \cap B = \phi$ meaning both events cannot occur at the same time. These are referred to as mutually exclusive events.

Class Exercise

1. A card is selected at random from a pack of 52 playing cards. Let A be the event that the card is an ace and D the event that it is a diamond.
   Find

   (a) $p(A \cap D)$

   (b) $p(A \cup D)$

   (c) $p(A')$

   (d) $p(A' \cap D)$

2. A group of 275 people at a music festival were asked if they play guitar, piano or drums:

   - 35 play drums only
   - 20 play guitar only
   - 15 play piano only
   - 30 play guitar and drums
   - 10 play piano and drums
   - 65 people play guitar and piano
   - 1 person plays all three instruments

(a) Draw a Venn diagram to represent the information

(b) A festival goer is chosen at random from the group. Find the probability that the person chosen

i) plays piano.

ii) plays at least two of guitar, piano or drums.

iii) plays exactly one of the instruments.

iv) plays none of the instruments.

Using formulae to solve probability questions

Instead of always drawing Venn diagrams to answer probability questions (in any case they break down when the events are more than three), we can derive formulas to enable us solve probability questions.

Let $p(A) = a$, $p(B) = b$ and $p(A \cap B) = i$ then $p(A \cup B) = (a - i) + i + (b - i) = a + b - i$.

Hence

$$p(A \cup B) = p(A) + p(B) - p(A \cap B) - \text{the Addition Rule.}$$

Rewriting the Addition Rule gives

$$p(A \cap B) = p(A) + p(B) - p(A \cup B).$$

Class Exercise

Use formulae and what you have learned about the algebra of sets to answer the following questions.

1) Let A and B be two events such that $p(A) = 0.6$, $p(B) = 0.7$ and $p(A \cup B) = 0.9$. Determine

(a) $p(A \cap B)$;

(b) $p(A^{'})$;

(c) $p(A^{'} \cap B)$;

(d) $p(A^{'} \cup B)$.

2) Consider two events T and Q where $p(T) = p(Q) = 3p(T \cap Q)$ and $p(T \cup Q) = 0.75$. Determine

(a) $p(T \cap Q)$;

(b) $p(T)$;

(c) $p(Q')$;

(d) $p(T' \cap Q')$;

(e) $p(T' \cap Q)$.

COUNTING TECHNIQUES

The classical approach to probability defines the probability that an event A occurs as

$$p(A) = \frac{\text{Number of required outcomes}}{\text{Total number of possible outcomes}} = \frac{n(A)}{n(S)} \tag{1}$$

It therefore requires you to be able to count the total number of outcomes in the event and in the sample space i.e. we should be able to list down all the possible outcomes of an experiment. There are many situations in which it would be too difficult and/or too tedious to list all of the possible outcomes in a sample space.
In your Basic Mathematics Course MMA 100, you learned some counting techniques that will enable you count the number of elements in a sample space without actually having to identify the specific outcomes. Reminders of some of the specific counting techniques we will explore include the multiplication rule, permutations and combinations.

The Multiplication Principle

If there are

- n_1 outcomes of a random experiment E_1
- n_2 outcomes of a random experiment E_2
- $\cdots$ and $\cdots$
- n_m outcomes of a random experiment E_m

then there are $n_1 \times n_2 \times ... \times n_m$ outcomes of the composite experiment $E_1E_2...E_m$.
NB: Always take care of whether replications (repetitions) are allowed or if there are any restrictions.

Class Exercise 1

a) How many possible outcomes are there when we toss a coin and roll a die?

b) How many possible license plates could be stamped if each license plate were required to have exactly 3 letters and 4 numbers?

c) How many possible license plates could be stamped if each license plate were required to have 3 unique letters and 4 unique numbers?

Permutations

How many ways can four people fill four executive positions?

$4 \times 3 \times 2 \times 1 = 24$

The counting of the number of permutations can be considered as a generalization of the Multiplication Principle.

A Generalization of the Multiplication Principle

Suppose there are n positions to be filled with n different objects, in which there are:

- n choices for the 1^{st} position
- $n-1$ choices for the 2^{nd} position
- $n-2$ choices for the 3^{rd} position
- $\cdots$ and $\cdots$
- 1 choice for the last position

The Multiplication Principle tells us there are then in general:

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

ways of filling the n positions.

The symbol $n!$ is read as 'n - factorial', and by definition 0! equals 1.

A permutation of n objects is an ordered arrangement of the n objects. We often call such a permutation a 'permutation of n objects taken n at a time', and denote it as ${}^{n}P_{n}$.

That is: ${}^{n}P_{n} = n \times (n-1) \times (n-2) \times ... \times 1 = n!$

NB: The first superscripted n represents the number of objects you want to arrange, while the second subscripted n represents the number of positions you have for the objects to fill.

Class Exercise 2 (sampling with or without replacement)

With 6 names in a bag, randomly select a name.

a) How many ways can the 6 names be assigned to 6 job assignments if we assume that each person can only be assigned to one job? (sampled without replacement)

b) What if the 6 names were sampled with replacement?

Another Generalization of the Multiplication Principle

Suppose there are r positions to be filled with n different objects, in which there are:

- n choices for the 1^{st} position
- $n - 1$ choices for the 2^{nd} position
- $n - 2$ choices for the 3^{rd} position
- $\cdots$ and $\cdots$
- $n - (r - 1)$ choice for the last position

The Multiplication Principle tells us there are in general:

$$n \times (n - 1) \times (n - 2) \times ... \times n - (r - 1)$$

ways of filling the r positions. We can show that in general, this quantity equals

$$\frac{n!}{(n - r)!}$$

A permutation of n objects taken r at a time is an ordered arrangement of n different objects in r positions. The number of such permutations is:

$$^{n}P_{r} = \frac{n!}{(n - r)!} \tag{2}$$

The superscripted n represents the number of objects you want to arrange, while the subscripted r represents the number of positions you have for the objects to fill.

Worked Example

An artist has 9 paintings. How many ways can he hang 4 paintings side-by-side on a gallery wall?

Solution

$$^{9}\mathsf{P}_4 = \frac{9!}{(9-4)!} = 3024 \quad \text{arrangements.}$$

Combinations

Example

Maria has three tickets for a concert. She would like to use one of the tickets herself. She could then offer the other two tickets to any of four friends (Ann, Beth, Chris, Dave). How many ways can 2 people be selected from 4 to go to the concert?

Definition: The number of unordered subsets, called a combination of n objects taken r at a time is,

$$^{n}\mathsf{C}_r = \frac{n!}{r!(n-r)!} \tag{3}$$

We say 'n choose r'. Hence

$$^{4}\mathsf{C}_2 = \frac{4!}{2!(4-2)!} = 12$$

Class Exercise 3

1. Twelve (12) patients are available for use in a research study. Only seven (7) should be assigned to receive the study treatment. How many different subsets of seven patients can be selected?

2. Consider a standard deck of cards containing 13 face values (Ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack, Queen, and King) and 4 different suits (Clubs, Diamonds, Hearts, and Spades) to play five-card poker.

   i) If you are dealt five cards, what is the probability of getting a 'full-house' hand containing three kings and two aces (KKKAA)?

   ii) If you are dealt five cards, what is the probability of getting any full-house hand?

AXIOMS OF PROBABILITY

Previously, we defined probability informally. We now consider a formal definition using the axioms of probability.

An axiom is simply a rule that has to be satisfied.

Definition

Probability is a (real-valued) set function p that assigns to each event A in the sample space S a number $p(A)$, called the probability of the event A which must satisfy the following three axioms:

1. The probability of any event A must be nonnegative, that is $p(A) \geqslant 0$.

2. The probability of the sample space is 1, that is, $p(S) = 1$ implying $0 \leqslant p(A) \leqslant 1$.

3. Given mutually exclusive (disjoint) events $A_1, A_2, A_3, \dots$ that is where $A_i \cap A_j = \phi, \text{for} i \neq j$,

   A. The probability of a finite union of the events is the sum of the probabilities of the individual events, that is:

$$\begin{aligned} p(A_1 \cup A_2 \cup \cdots \cup A_k) &= p(A_1) + p(A_2) + p(A_3) + \cdots + p(A_k) \\ &= \sum_{i=1}^{k} p(A_i) \end{aligned}$$

   B. The probability of a countably infinite union of the events is the sum of the probabilities of the individual events, that is:

$$\begin{aligned} p(A_1 \cup A_2 \cup \cdots) &= p(A_1) + p(A_2) + p(A_3) + \cdots \\ &= \sum_{i=1}^{\infty} p(A_i) \end{aligned}$$

These conditions are known as the axioms of the theory of probability.

1. The first axiom states that all the probabilities are nonnegative real numbers.

2. The second axiom attributes a probability of unity to the universal event S, thus providing a normalization of the probability measure.

3. The third axiom states that the probability function must be additive, consistently with the intuitive idea of how probabilities behave.

All probabilistic results are based directly or indirectly on the axioms and only the axioms. Any assignment of probability to an event must satisfy the THREE axioms stated above regardless of your interpretation of probability.

Class Exercise

Suppose that the MMA105 class has 43 students, such that 1 is Physics, 4 are Mathematics, 20 are Business, 9 are Computer, and 9 are Geography major students. Randomly select a student from the class. Define the following events:

1. P - the event that a Physics major is selected.

2. M - the event that a Mathematics major is selected.

3. B - the event that a Business major is selected.

4. C - the event that a Computer major is selected.

5. G - the event that a Geography major is selected.

The sample space is $S = \{P, M, B, C, G\}$.
Using the relative frequency approach to assigning probability to events, assign probabilities to the events $p(P)$, $p(M)$,$p(B)$,$p(C)$ and $p(G)$. Show that the three axioms of probability are satisfied.

Probability rules derived from the axioms of probability

Theorem 1: $p(A) = 1 - p(A^{'})$
Theorem 2: $p(\phi) = 0$
Theorem 3: If events A and B are such that $A \subseteq B$, then $p(A) \leqslant p(B)$.
Theorem 4: $p(A) \leqslant 1$.
Theorem 5: For any two events A and B, $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.

Class Exercise

1. Prove the five theorems derived from the three axioms of probability.

2. A company has bid on two large construction projects. The company president believes that the probability of winning the first contract is 0.6, the probability of winning the second contract is 0.4, and the probability of winning both contracts is 0.2. What is the probability that the company wins:

   a) at least one contract?

   b) the first contract but not the second contract?

c) neither contract?

d) exactly one contract?

3. If it is known that $A \subseteq B$, what can be definitively said about $p(A \cap B)$?

4. If 7% of the population smokes cigars, 28% of the population smokes cigarettes, and 5% of the population smokes both, what percentage of the population smokes neither cigars nor cigarettes?

Conditional Probability

A conditional probability is a probability of an event given that another event has occurred. For example, rather than being interested in knowing the probability that a randomly selected students fails the probability exam, we might instead be interested in knowing the probability that a randomly selected student fails the probability exam given that the student is an education major.

We are finding the probability of event B occurring given that event A has occurred denoted $p(B/A)$.

Event A is termed the prior event while event B is the subsequent event.

Conditional probabilities have the effect of shrinking the sample space to the prior event.

Example 1

A researcher is interested in evaluating how well a diagnostic test works for detecting renal disease in patients with high blood pressure. She performs the diagnostic test on 137 patients, 67 with known renal disease and 70 who are known to be healthy. The diagnostic test comes back either positive (the patient has renal disease) or negative (the patient does not have renal disease). Here are the results of her experiment:

Test Result

Truth	Positive	Negative	Total
Renal disease	44	23	67
Healthy	10	60	70
Total	54	83	137

a) What is the probability that the patient tests positive?

b) What is the probability that a patient has renal disease?

c) What is the probability that a patient tests positive given that they have renal disease?

Solution

Let T denote the event that a patient test positive and D the event that a patient has renal disease

a) $p(T) = \dfrac{n(T)}{n(S)} = \dfrac{54}{137}$.

b) $p(D) = \dfrac{n(D)}{n(S)} = \dfrac{67}{137}$.

c) $p(T/D) = \dfrac{n(T \cap D)}{n(D)} = \dfrac{44}{67}$. Here you are considering only the proportion of those who are diseased (sample space shrinks); how many of them test positive.

Therefore, from part c) in the example we can write down a formula for conditional probability as

$$p(A/B) = \frac{p(A \cap B)}{p(B)}; \quad p(B) > 0$$

Multiplying both sides by $p(B)$ we get

$$p(A \cap B) = p(B) \times p(A/B) - \text{the Multiplication Rule}$$

or

$$p(B/A) = \frac{p(A \cap B)}{p(A)}; \quad p(A) > 0$$

Multiplying both sides by $p(A)$ we get

$$p(A \cap B) = p(A) \times p(B/A) - \text{the Multiplication Rule}$$

Class Exercise 1

1. Two fair spinners each have faces numbered 1 – 4. The two spinners are thrown together and the sum of the faces shown on the spinners is recorded. Given that at least one spinner lands on a 3, find the probability of the spinners indicating a sum of exactly 5.

2. C and D are events such that $p(C) = 0.2$, $p(D) = 0.1$ and $p(C/D) = 0.3$. Find $p(D/C)$, $p(C' \cap D')$ and $p(C' \cap D)$.

3. The turnout of spectators at a motor rally is dependent on the weather. On a rainy day, the probability of a big turnout is 0.4 but if it does not rain the probability of a big turn increases to 0.9. The weather forecast gives a probability of 0.75 that it will rain on the day of the race.

   a) Draw a tree diagram to represent this information.

   b) Find the probability that

      i) there is a big turnout and it rains.

      ii) there is a big turnout.

Properties of Conditional Probability

Because conditional probability is just a probability, it satisfies the three axioms of probability. That is, as long as $p(B) > 0$:

1. $p(A/B) \geqslant 0$.

2. $p(B/B) = 1$.

3. If $A_1, A_2, \ldots, A_k$ are mutually exclusive events, then $p(A_1 \cup A_2 \cup ... \cup A_k/B) = p(A_1/B) + p(A_2/B) + ... + P(A_k/B)$ and likewise for infinite unions.

Class Exercise 2

Show that conditional probabilities satisfy the three axioms of probability.

MUTUALLY EXCLUSIVE AND INDEPENDENT EVENTS

Mutually Exclusive Events

Two or more events are said to be mutually exclusive if they cannot occur at the same time i.e. events that have no outcomes in common are said to be mutually exclusive events or disjoint events.
If A and B are mutually exclusive then $A \cap B = \phi$. This implies $p(A \cap B) = 0$. Recall: for any two events A and B

$$p(A \cup B) = p(A) + p(B) - p(A \cap B)$$

If A and B are mutually exclusive then

$$p(A \cup B) = p(A) + p(B)$$

Independent Events

Two events are independent, if the occurrence of one does not affect the occurrence of the other. Therefore if A and B are independent, the probability of A happening is the same whether or not B has happened i.e. $p(A/B) = p(A)$; similarly $p(B/A) = p(B)$.
Recall for any two events A and B

$$p(A \cap B) = p(A) \times p(B/A)$$

If A and B are independent then

$$p(A \cap B) = p(A) \times p(B)$$

Example 1

Events A and B are mutually exclusive and $p(A) = 0.2$ and $p(B) = 0.4$. Determine $p(A \cup B)$, $p(A \cap B^{'})$ and $p(A^{'} \cap B^{'})$.

Solution Example 1

$p(A \cup B) = p(A) + p(B) = 0.2 + 0.4 = 0.6$
$p(A \cap B^{'}) = p(A) + p(B^{'}) - p(A \cup B^{'}) = 0.2 + 0.6 - 0.6 = 0.2$
$(A^{'} \cap B^{'}) = p(A^{'}) + p(B^{'}) - p(A^{'} \cup B^{'}) = 0.8 + 0.6 - 1 = 0.4$

Tree Diagrams to answer probability questions

When the experiment of interest consists of a sequence of several stages, it is convenient to represent these with a tree diagram. Tree diagrams are used to display the sample space. The tiers in a tree diagram represent events while the branches represent the outcomes of the event. Once we have an appropriate tree diagram, probabilities and conditional probabilities can be entered on the various branches; this will make repeated use of the multiplication rule quite straightforward.

Exercise 1

1. A chain of video stores sells three different brands of DVD players. Of its DVD player sales, 50% are brand 1 (the least expensive), 30% are brand 2, and 20% are brand 3. Each manufacturer offers a 1-year warranty on parts and labor. It is known that 25% of brand 1's DVD players require warranty repair work, whereas the corresponding percentages for brands 2 and 3 are 20% and 10%, respectively.

a) What is the probability that a randomly selected purchaser has bought a brand 1 DVD player that will need repair while under warranty?

b) What is the probability that a randomly selected purchaser has a DVD player that will need repair while under warranty?

c) If a customer returns to the store with a DVD player that needs warranty repair work, what is the probability that it is a brand 1 DVD player? A brand 2 DVD player? A brand 3 DVD player?

2. It is known that 30% of a certain company's washing machines require service while under warranty, whereas only 10% of its dryers need such service. If someone purchases both a washer and a dryer made by this company, what is the probability that

   a) both machines need warranty service?

   b) neither machine needs service?

3. Each day, Monday through Friday, a batch of components sent by a first supplier arrives at a certain inspection facility. Two days a week, a batch also arrives from a second supplier. Eighty percent of all supplier 1's batches pass inspection, and 90% of supplier 2's do likewise. What is the probability that, on a randomly selected day, two batches pass inspection? Assume that on days when two batches are tested, whether the first batch passes is independent of whether the second batch does so.

Bayes Theorem

The Bayes' theorem enables us compute a posterior probability $p(A_j/B)$ from given prior probabilities $p(A_i)$ and conditional probabilities $p(B/A_i)$. In other words, if a subsequent event has occurred, we use the Bayes' theorem to find the probability that a given prior event occurred.

To state the Bayes' theorem, we first need another result, the Law of Total Probability. Recall that events $A_1, ..., A_k$ are mutually exclusive if no two have any common outcomes.

The events are exhaustive if one A_i must occur, so that $A_1 \cup A_2 \cup \cdots \cup A_k = S$.

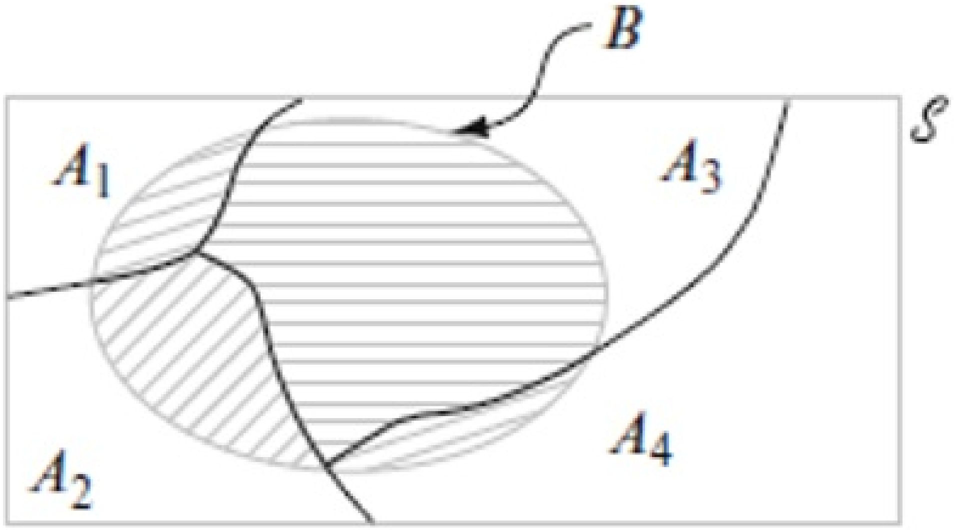
Law of Total Probability

Let $A_1, ..., A_k$ be mutually exclusive and exhaustive events. Then for any other event B,

$$p(B) = p(B/A_1)p(A_1) + \cdots + p(B/A_k)p(A_k) = \sum_{i=1}^{k} p(B/A_i)p(A_i)$$

Proof

Because the $A_i's$ are mutually exclusive and exhaustive, if B occurs it must be in conjunction with exactly one of the $A_i's$. That is, $B = (A_1 \quad \text{and} \quad B)$ or $\cdots$ or $(A_k \quad \text{and} \quad B) = (A_1 \cap B) \cup \cdots \cup (A_k \cap B)$, where the events $(A_i \cap B)$ are mutually exclusive. This 'partitioning of B' is illustrated in the figure below.



Thus

$$p(B) = p(B/A_1)p(A_1) + \cdots + p(B/A_k)p(A_k) = \sum_{i=1}^{k} p(B/A_i)p(A_i)$$

as desired.

The Bayes' Theorem

Let $A_1, ..., A_k$ be a collection of mutually exclusive and exhaustive events with $p(A_i) > 0$ for $i = 1, \cdots, k$. Then for any other event B, for which $p(B) > 0$

$$p(A_j/B) = \frac{p(A_j \cap B)}{p(B)} = \frac{p(B/A_j)p(A_j)}{\sum_{i=1}^{k} p(B/A_i).p(A_i)}, j = 1, \cdots, k.$$

The transition from the second to the third expression in the Bayes' Theorem rests on using the multiplication rule in the numerator and the law of total probability in the denominator. The proliferation of events and subscripts in the Bayes' Theorem can be a bit intimidating to probability newcomers. As long as there are relatively few events in the partition, a tree diagram can be used as a basis for calculating posterior probabilities without ever referring explicitly to Bayes' theorem.

Example 2

In a factory machines A, B and C produce electronic components. Machine A produces 16%, Machine B produces 50% and Machine C produces 34%. Some of the components are defective. Machine A produces 4% defective, Machine B produces 3% defective and Machine C produces 7% defective.

a) Draw a tree diagram to represent this information.

b) Find the probability that a randomly selected component is

   i) produced by Machine A and is defective.

   ii) defective.

c) Given that a randomly selected component is defective, find the probability that it was produced by Machine B.

THE CONCEPT OF A RANDOM VARIABLE

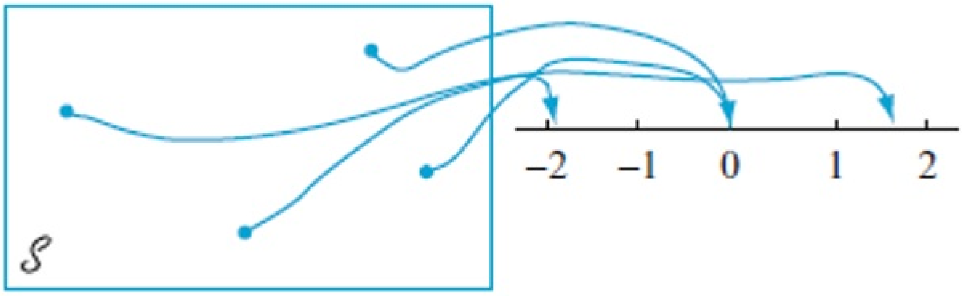
Introduction

Whether an experiment yields qualitative or quantitative outcomes, methods of statistical analysis require that we focus on certain numerical aspects of the data (such as a sample proportion $\frac{x}{n}$, mean $\bar{x}$, or standard deviation s). The concept of a random variable allows us to pass from the experimental outcomes themselves to a numerical function of the outcomes. There are two fundamentally different types of random variables — discrete random variables and continuous random variables.

Random Variable

Each outcome of an experiment can be associated with a number by specifying a rule of association. Such a rule of association is called a random variable – a variable because different numerical values are possible and random because the observed value depends on which of the possible experimental outcomes results. The figure below depicts a random variable.

A random variable is a way of mapping outcomes of random processes to numbers (quantify



outcomes) e.g. Let X be the outcome when you toss a coin,

$$X = \begin{cases} 1, & \text{if heads} \\ 0, & \text{if tails} \end{cases}$$

or let Y be the sum of the uppermost faces when two die are rolled.

When you quantify outcomes you can do more mathematics on the outcomes and equally more mathematical notations on the outcome. The probability that the sum of the uppermost faces is less than or equal to 12 is denoted as $p(Y \leqslant 12)$.

Capital letters, X, are used to denote the random variables, while small letters x, denote a particular value (or realized value) of the random variable.

For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S. In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers. Random variables are denoted using capital letters e.g. X while the realized value (or a particular value) of the random variable is denoted using small letters e.g. x.

Notation

- $X(s) = x$ means that x is the value associated with the outcome s by the random variable X.
- $p(x) = p(X = x)$ means the probability that the random variable X is equal to the particular value x.

Example 1

When a student attempts to connect to a university computer system, either there is a failure (F), or there is a success (S). With $S = \{S, F\}$ define an random variable X by $X(S) = 1$, $X(F) = 0$. The random variable X indicates whether (1) or not (0) the student can connect.

Sometimes the possible values are many and would be tedious to list; hence a description as in Example 2 can be used.

Example 2

Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer and define a random variable Y by

$$\mathsf{Y} = \begin{cases} 1, & \text{if the selected number is unlisted} \\ 0, & \text{if the selected number is listed in a directory} \end{cases}$$

For example, if 5282966 appears in the telephone directory, then $\mathsf{Y}(5282966) = 0$, whereas $\mathsf{Y}(7727350) = 1$ tells us that the number 7727350 is unlisted. A word description of this sort is more economical than a complete listing, so we will use such a description whenever possible.
In Examples 1 and 2, the only possible values of the random variable were 0 and 1. Such a random variable arises frequently enough to be given a special name, after the individual who first studied it.

Definition

Any random variable whose only possible values are 0 and 1 is called a Bernoulli random variable.

Class Exercise

Three fair coins are tossed. Let X = the number of heads counted. List each outcome in the sample space along with the associated value of X.

TWO TYPES OF RANDOM VARIABLES

In your MAS 101: Descriptive Statistics Course you distinguished between data resulting from observations on a counting (discrete) variable and data obtained by observing values of a measurement (continuous) variable. A slightly more formal distinction characterizes two different types of random variables.
A discrete random variable is a random variable whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on.
A random variable is continuous if both of the following apply:

i) Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g. from $-\infty$ to $+\infty$) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).

ii) No possible value of the variable has positive probability, that is, $p(X = c) = 0$ for any possible value c.

NB

1. To study basic properties of discrete random variables, only the tools of discrete mathematics - summation and differences - are required.
2. The study of continuous variables requires the continuous mathematics of the calculus - integrals and derivatives.

Class Exercise

For each random variable defined here, describe the set of possible values for the variable, and state whether the variable is discrete or continuous.

a) $X =$ the number of unbroken eggs in a randomly chosen standard egg carton.

b) $Y =$ the number of students on a class list for a particular course who are absent on the first day of classes.

c) $U =$ the number of times a duffer has to swing at a golf ball before hitting it.

d) $X =$ the length of a randomly selected rattlesnake.

e) $Z =$ the amount of royalties earned from the sale of a first edition of 10,000 textbooks.

f) $Y =$ the pH of a randomly chosen soil sample.

g) $X =$ the tension (psi) at which a randomly selected tennis racket has been strung.

h) $X =$ the total number of coin tosses required for three individuals to obtain a match (HHH or TTT).

PROBABILITY FUNCTIONS

A DISCRETE RANDOM VARIABLE

The probability distribution of a random variable X says how the total probability of 1 is distributed among (allocated to) the various possible values of the random variable X.

Example 1

Six lots of components are ready to be shipped by a supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of Defectives	0	2	0	1	2	0

Determine the probability distribution of X.

Solution

Let X be the number of defectives in the selected lot. The three possible X values are 0, 1, and 2. Of the six equally likely simple events, three result in $X = 0$, one in $X = 1$ and the other two in $X = 2$. Let $p(x) = p(X = x)$ then $p(0) = \frac{1}{2}$,$p(1) = \frac{1}{6}$ and $p(2) = \frac{2}{6}$. The values of X along with their probabilities collectively specify the probability distribution or probability mass function of X.

The probability mass function can be specified in a table:

x	0	1	2
$p(x) = p(X = x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{2}{6}$

or as a function

$$p(x) = p(X = x) = \begin{cases} \frac{1}{2}, & x = 0 \\ \frac{1}{6}, & x = 1 \\ \frac{2}{6}, & x = 2 \end{cases}$$

This means that if this experiment were repeated over and over again, in the long run $X = 0$ would occur one-half of the time, $X = 1$ one-sixth of the time, and $X = 2$ one-third of the time.

Properties of probability mass functions

A function is a probability mass function if it satisfies the following conditions

1. $p(x) \geqslant 0)$; probability exists

2. $\sum_{\forall x} p(x) = 1$; total probability $= 1$

Class Exercise 1

1) A tetrahedron die has the numbers 1, 2, 3 and 4 on its faces. The die is biased in a way that the probability of the die landing on any number x is $\frac{k}{x}$ where k is a constant. Find the probability mass function of X, the number the die lands on after a single roll.

2) A discrete random variable X has probability mass function

x	0	1	2
$p(x) = p(X = x)$	$\frac{1}{4} - a$	a	$\frac{7}{12} + a$

Determine a.

3) The random variable Y has probability function

$$p(y) = \begin{cases} ky, & y = 1, 3 \\ k(y-1), & y = 2, 4 \\ 0, & \text{otherwise} \end{cases}$$

Find k and the probability mass function of Y.

4) A discrete random variable X has probability mass function

x	1	2	3	4	5	6
$p(x)$	0.1	0.2	0.3	0.25	0.1	0.05

Find $p(1 < X < 5)$; $p(2 \leqslant X \leqslant 4)$; $p(3 < X \leqslant 6)$ and $p(X \leqslant 3)$.

You must have realized that for a discrete random variable

$$p(X \leqslant a) \neq p(X < a) \quad \text{unless} \quad p(X = a) = 0$$

Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ of a discrete random variable X with probability mass function $p(x)$ is defined for every number x by

$$\begin{aligned} F(x) &= p(X \leqslant x) \\ &= \sum_{y : y \leqslant x} p(y) \end{aligned}$$

For any number x, $F(x)$ is the probability that the observed value of X will be at most x.

Example 2

A store carries flash drives with either 1, 2, 4, 8, or 16 GB of memory. The accompanying table gives the distribution of Y = the amount of memory in a purchased drive:

y	1	2	4	8	16
$p(y)$	0.05	0.10	0.35	0.40	0.10

Determine the cumulative distribution function, $F(y)$, for the five possible values of Y.

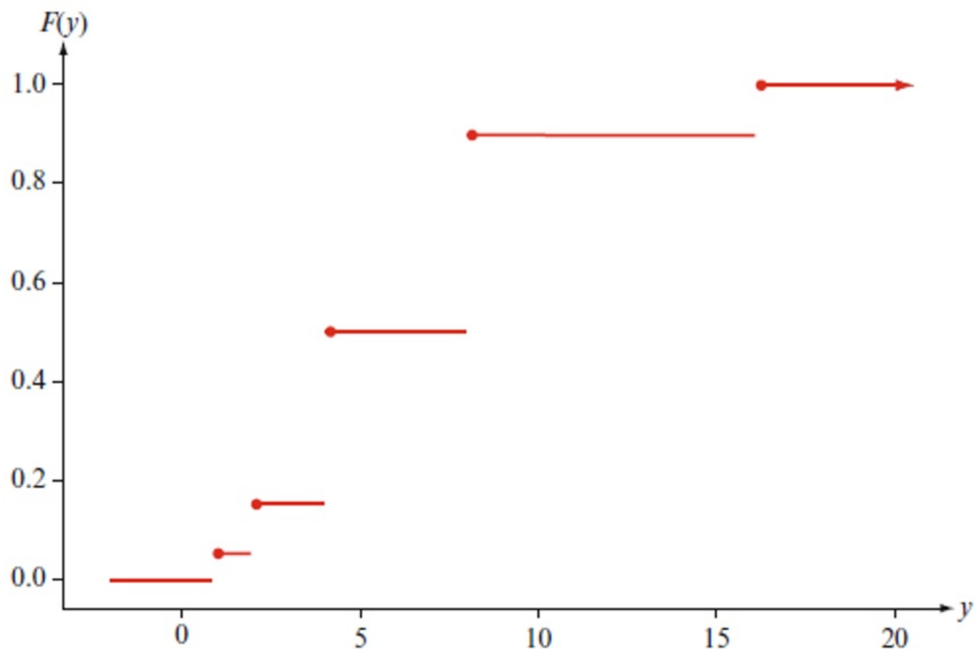
Solution

y	1	2	4	8	16
$p(y)$	0.05	0.10	0.35	0.40	0.10
$F(y)$	0.05	0.15	0.5	0.9	1

or expressed as a function

$$F(y) = \begin{cases} 0, & y < 1 \\ 0.05, & 1 \leqslant y < 2 \\ 0.15, & 2 \leqslant y < 4 \\ 0.5, & 4 \leqslant y < 8 \\ 0.9, & 8 \leqslant y < 16 \\ 1, & 16 \leqslant y \end{cases}$$

The graph of the cumulative distribution function is given by: For X a discrete random variable,



the graph of $F(x)$ will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a step function.

For any other number y, $F(y)$ will equal the value of F at the closest possible value of Y to the left of y.
For example,
$F(2.7) = p(Y \leqslant 2.7) = p(Y \leqslant 2) = F(2) = 0.15.$
$F(7.999) = p(Y \leqslant 7.999) = p(Y \leqslant 4) = F(4) = 0.5.$
$F(25) = p(Y \leqslant 25) = p(Y \leqslant 16) = F(16) = 1.$

Class Exercise 2

1) Any positive integer is a possible X value, and the probability mass function is

$$p(x) = \begin{cases} (1-p)^{x-1}, & x = 1, 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases}$$

Determine the cumulative distribution function of X.

2) The discrete random variable X has cumulative distribution function $F(x)$ defined by

$$F(x) = \begin{cases} \dfrac{x+k}{8}, & x = 1, 2, 3. \\ 0, & \text{otherwise} \end{cases}$$

i) Find the value of k.

ii) Draw the distribution table for the cumulative distribution function of X.

iii) Find $F(2.6)$.

iv) Find $p(x)$ the probability distribution of X.

3) The discrete random variable X has cumulative distribution function $F(x)$ defined by

$$F(x) = \begin{cases} 0, & x = 0. \\ \dfrac{1+x}{6}, & x = 1, 2, 3, 4, 5. \\ 1, & x > 5 \end{cases}$$

i) Find $F(4)$.

ii) Show that $p(4) = \frac{1}{6}$.

iii) Find $p(x)$.

PROPERTIES OF CDFs (for a discrete random variable)

Suppose X represents the numbers of defective components in a shipment consisting of six components, so that possible X values are $0, 1, ..., 6$. Then

$$\begin{aligned} p(3) = p(X=3) &= [p(0)+p(1)+p(2)+p(3)] - [p(0)+p(1)+p(2)] \\ &= p(X \leqslant 3) - p(X \leqslant 2) \\ &= F(3) - F(2) \end{aligned}$$

More generally, the probability that X falls in a specified interval is easily obtained from the cumulative distribution function. For example,

$$\begin{aligned} p(2 \leqslant X \leqslant 4) &= p(2)+p(3)+p(4) \\ &= [p(0)+p(1)+p(2)+p(3)+p(4)] - [p(0)+p(1)] \\ &= p(X \leqslant 4) - p(X \leqslant 1) \\ &= F(4) - F(1) \end{aligned}$$

Notice that $p(2 \leqslant X \leqslant 4) \neq F(4) - F(2)$. This is because the X value 2 is included in $p(2 \leqslant X \leqslant 4)$, so we do not want to subtract out its probability. However,

$$p(2 < X \leqslant 4) = F(4) - F(2)$$

because $X = 2$ is not included in the interval $(2 < X \leqslant 4)$.

For any two numbers a and b with $a \leqslant b$,

$$p(a \leqslant X \leqslant b) = F(b) - F(a-)$$

$F(a-)$ represents the maximum of $F(x)$ values to the left of a.

Equivalently, if a is the limit of values of x approaching from the left, then $F(a-)$ is the limiting value of $F(x)$. In particular, if the only possible values are integers and if a and b are integers, then

$$\begin{aligned} p(a \leqslant X \leqslant b) &= p(X = a \quad \text{or} \quad a+1 \quad \text{or} \ldots \text{or} \quad b) \\ &= F(b) - F(a-1) \end{aligned} \tag{4}$$

Taking $a = b$ in (12) yields $p(X = a) = F(a) - F(a-1)$ in this case. The reason for subtracting $F(a-)$ rather than $F(a)$ is that we want to include $p(X = a)$; $F(b) - F(a)$ gives $p(a < X \leqslant b)$. This proposition will be used extensively when computing binomial and Poisson probabilities.

PROBABILITY FUNCTIONS

A CONTINUOUS RANDOM VARIABLE

As mentioned earlier, the two important types of random variables are discrete and continuous. In this sub-topic, we study the continuous random variable that arises in many applied problems.

A random variable X is continuous if

1. possible values comprise either a single interval on the number line (for some $A < B$, any number x between A and B is a possible value) or a union of disjoint intervals, and
2. $p(X = c) = 0$ for any number c that is a possible value of X.

Example 1

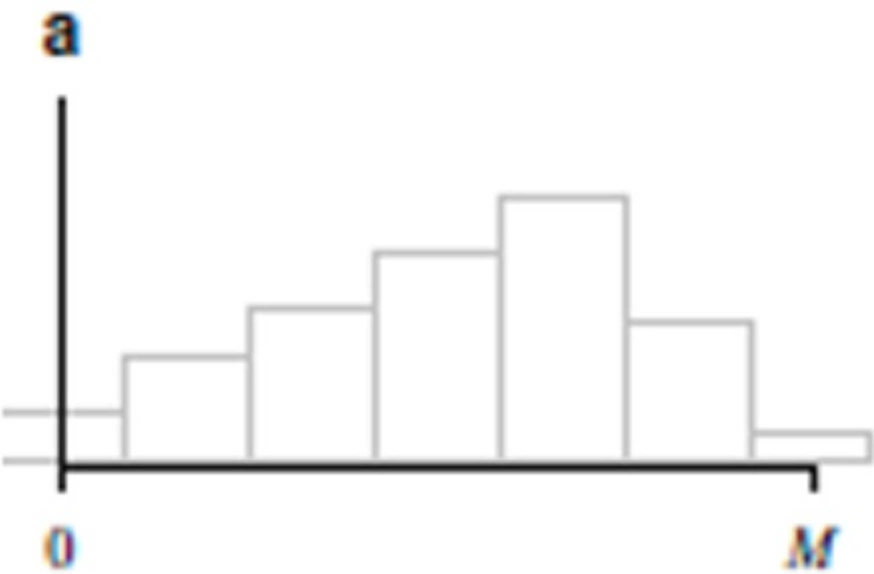
If in the study of the ecology of a lake, we make depth measurements at randomly chosen locations, then $X =$ the depth at such a location is a continuous random variable. Here A is the minimum depth in the region being sampled, and B is the maximum depth.

Example 2

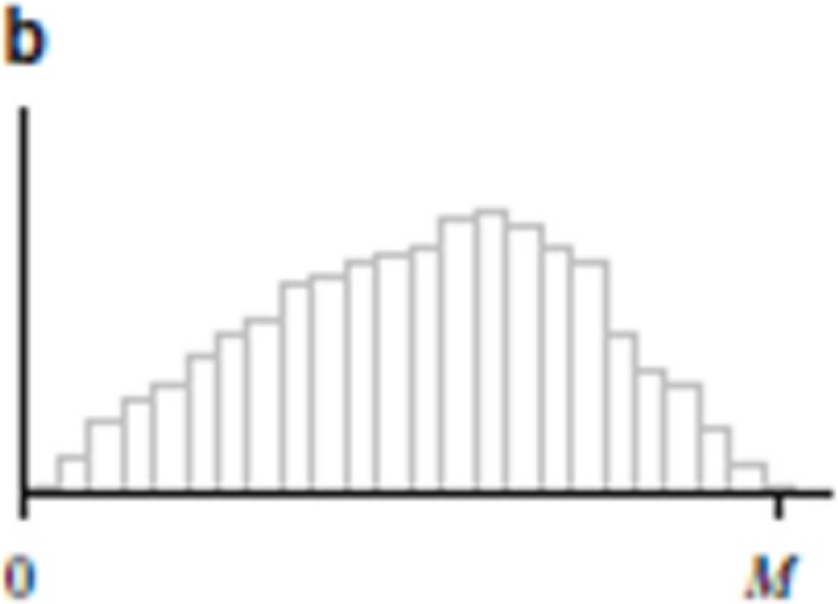
If a chemical compound is randomly selected and its pH X is determined, then X is a continuous random variable because any pH value between 0 and 14 is possible. If more is known about the compound selected for analysis, then the set of possible values might be a sub-interval of [0,14], such as $5.5 \leqslant X \leqslant 6.5$, but X would still be continuous.

PROBABILITY DENSITY FUNCTIONS

Suppose the variable X of interest is the depth of a lake at a randomly chosen point on the surface. Let $M =$ the maximum depth (in meters), so that any number in the interval [0, M] is a possible value of X. If we 'discretize' X by measuring depth to the nearest meter, then possible values are non-negative integers less than or equal to M. The resulting discrete distribution of depth can be pictured using a probability histogram. If we draw the histogram so that the area of the rectangle above any possible integer k is the proportion of the lake whose depth is (to the nearest meter) k, then the total area of all rectangles is 1. A possible histogram appears in the figures below.



If depth is measured much more accurately and the same measurement axis as in the figure above is used, each rectangle in the resulting probability histogram is much narrower, although the total area of all rectangles is still 1. A possible histogram is pictured in the next figure; it has a much smoother appearance than the first histogram.

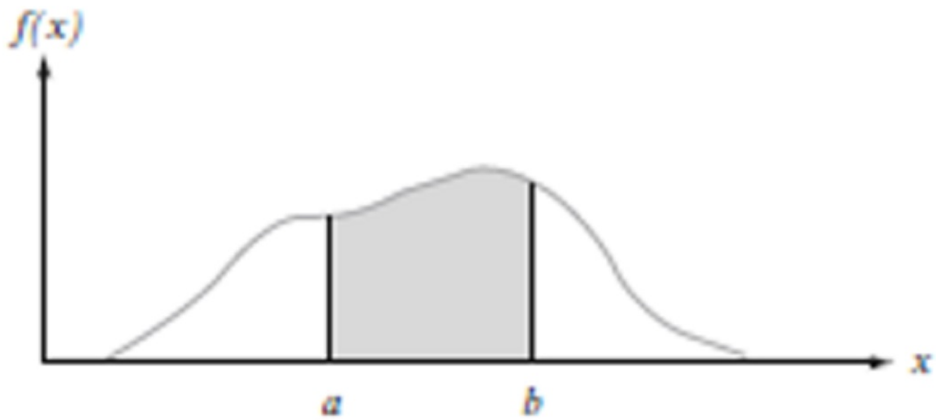


If we continue in this way to measure depth more and more finely, the resulting sequence of histograms approaches a smooth curve, as pictured in our last figure below. Because for each histogram the total area of all rectangles equals 1, the total area under the smooth curve is also 1. The probability that the depth at a randomly chosen point is between a and b is just the area under the smooth curve between a and b. It is exactly a smooth curve of the type pictured in the following figure that specifies a continuous probability distribution.

Let X be a continuous random variable. Then a probability distribution or probability density function of X is a function $f(x)$ such that for any two numbers a and b with $a \leqslant b$,

$$p(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

That is, the probability that X takes on a value in the interval $[a, b]$ is the area above this interval and under the graph of the density function, as illustrated in the figure below. The graph of $f(x)$ is often referred to as the density curve.



For $f(x)$ to be a legitimate probability density function it must satisfy two conditions namely:

1. $f(x) \geqslant 0$; probability exists.

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

Also always remember that for a continuous random variable, $p(X = x) = 0$ hence $p(a \leqslant X \leqslant b) = p(a \leqslant X < b) = p(a < X \leqslant b) = p(a < X < b)$.

Recall: How to integrate

1. $\int x^n dx = \dfrac{x^{n+1}}{n+1} + c$.

2. $\int \dfrac{1}{x} dx = \ln(x) + c$.

3. $\int ax^n dx = a\dfrac{x^{n+1}}{n+1} + c$.

4. $\int e^x dx = e^x + c$.

5. $\int e^{ax} dx = \dfrac{1}{a} e^{ax} + c$.

Class Exercise 1

1. The random variable X has probability density function

$$f(x) = \begin{cases} kx(4-x), & 2 \leqslant x \leqslant 4 \\ 0, & \text{otherwise} \end{cases}$$

Determine the value of k and sketch the probability density function of X.

2. The random variable Y has probability density function

$$f(y) = \begin{cases} k, & 1 < y < 2 \\ k(y-1), & 2 \leqslant y \leqslant 4 \\ 0, & \text{elsewhere} \end{cases}$$

Determine the value of k and sketch the probability density function of Y.

3. The random variable X has probability density function

$$f(x) = \begin{cases} kx^3, & 1 \leqslant x \leqslant 4 \\ 0, & \text{elsewhere} \end{cases}$$

Determine the value of k.

4. The random variable Z has probability density function

$$f(z) = \begin{cases} k, & 0 \leqslant z < 2 \\ k(2z-3), & 2 \leqslant z \leqslant 3 \\ 0, & \text{elsewhere} \end{cases}$$

Determine the value of k and sketch the probability density function of Z for all values of Z.

5. The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is, in general, subject to uncertainty. Consider the reference line connecting the valve stem on a tire to the center point, and let X be the angle measured clockwise to the location of an imperfection. One possible probability density function for X is

$$f(x) = \begin{cases} \frac{1}{300}, & 0 \leqslant z \leqslant 360 \\ 0, & \text{elsewhere} \end{cases}$$

Determine the probability that the angle is between 90 and 180 i.e. $p(90 \leqslant X \leqslant 180)$.

Cumulative Distribution Function

The cumulative distribution function $F(x)$ for a continuous random variable X is defined for every number x by

$$F(x) = p(X \leqslant x) = \int_{-\infty}^{x} f(x)dx$$

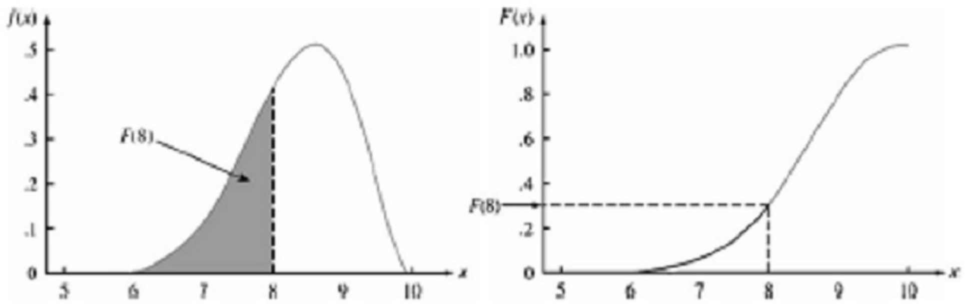
For each x, $F(x)$ is the area under the density curve to the left of x as illustrated in the figure below. $F(x)$ increases smoothly as x increases.

Example

The random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{4}x, & 1 \leqslant x \leqslant 3 \\ 0, & \text{elsewhere} \end{cases}$$

Determine $F(x)$



Solution

For $x < 1$, $F(x) = \int_{-\infty}^{x} 0dx = 0$

For $1 \leqslant x \leqslant 3$, $F(x) = \int_{-\infty}^{1} 0dx + \int_{1}^{x} \frac{1}{4}xdx = \frac{x^2-1}{8}$

For $x > 3$, $F(x) = \int_{-\infty}^{1} 0dx + \int_{1}^{3} \frac{1}{4}xdx + \int_{3}^{x} 0dx = 1$

Hence

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{x^2-1}{8}, & 1 \leqslant x \leqslant 3 \\ 1, & x > 3 \end{cases}$$

Properties of CDFs for continuous random variables

$p(a < X \leqslant b) = p(X \leqslant b) - p(X \leqslant a) = F(b) - F(a)$

$p(a \leqslant X \leqslant b) = p(a < X \leqslant b) + p(X = a) = F(b) - F(a) + f(a)$

$p(a < X < b) = p(a < X \leqslant b) - p(X = b) = F(b) - F(a) - f(b)$

However for a continuous random variable $p(X \leqslant x) = 0$ and

$$f(x) = \frac{d}{dx}F(x)$$

Class Exercise 2

1. The random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{5}, & 1 < x < 2 \\ \frac{1}{5}(x-1), & 2 \leqslant x \leqslant 4 \\ 0, & \text{elsewhere} \end{cases}$$

   Find $F(x)$.

2. The random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{5}x + \frac{3}{20}x^2, & 0 \leqslant x \leqslant 2 \\ 1, & x > 2 \end{cases}$$

Find

i) $p(X \leqslant 1.5)$.

ii) $p(0.5 \leqslant X \leqslant 1.5)$.

iii) $p(X = 1)$.

iv) $f(x)$.

3. Suppose that the time in minutes that a person has to wait at a certain station for a train is seen to be a random phenomenon and has a probability function specified by the cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \leqslant x < 1 \\ \frac{1}{2}, & 1 \leqslant x < 2 \\ \frac{x}{4}, & 2 \leqslant x < 4 \\ 1, & x > 4 \end{cases}$$

What is the probability that a person will have to wait

a) more than three minutes?

b) less than three minutes?

c) between one and three minutes?

SOME IMPORTANT RESULTS

This section will provide nine results which you will use henceforth without prove. You will meet and prove most of them in your Mathematics courses.

1. Summation and its properties, $\sum$ read 'sigma', is the summation notation.

a) $x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$

b) If a and b are fixed numbers then

$\sum_{i=1}^{n} bx_i = b\sum_{i=1}^{n} x_i$

$\sum_{i=1}^{n}(bx_i + a) = b\sum_{i=1}^{n} x_i + an$

2. $(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n}a^0 b^n$

3. $e^x = 1 + \dfrac{x}{1!} + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \dfrac{x^k}{k!}$

4. $\sum_{l=0}^{\infty} ar^l = a + ar + ar^2 + \cdots = \dfrac{a}{(1-r)}, |r| < 1$

5. $1^2 + 2^2 + 3^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$

6. $\sum_{k=0}^{\infty} kx^{(k-1)} = 1 + 2x + 3x^2 + 4x^3 + \cdots = \dfrac{1}{(1-x)^2}, |x| < 1$

7. $\sum_{k=0}^{\infty} k^2x^{(k-1)} = 1 + 4x + 9x^2 + 16x^3 + \cdots = \dfrac{(1+x)}{(1-x)^3}, |x| < 1$

8. $1 + rx + \dfrac{r(r+1)}{2!}x^2 + \dfrac{r(r+1)(r+2)}{3!}x^3 + \cdots = (1-x)^{-r}$

9. $\lim_{n\to\infty}\left(1 - \dfrac{a}{n}\right)^n = e^{-a}$

EXPECTATIONS AND MOMENTS OF RANDOM VARIABLES

Another view of probability mass functions

It is often helpful to think of a probability mass function as specifying a mathematical model for a discrete population. Consider selecting at random a student who is among the 15000 registered for the current semester at Maseno University. Let X be the number of courses for which the selected student, is registered for and suppose that X has the following probability mass function:

x	1	2	3	4	5	6	7
$p(x) = p(X = x)$	0.01	0.03	0.13	0.25	0.39	0.17	0.02

Since $p(1) = 0.01$, we know that $0.01 \times 15000 = 150$ of the students are registered for one course, and similarly for the other x values.

x	1	2	3	4	5	6	7
$p(x) = p(X = x)$	0.01	0.03	0.13	0.25	0.39	0.17	0.02
Number of Students	150	450	1950	3750	5850	2550	300

To compute the average number of courses per student, or the average value of X in the population, we should calculate the total number of courses and divide by the total number of students. Since each of 150 students is taking one course, these 150 contribute 150 courses to the total. Similarly, 450 students contribute 2(450) courses, and so on. The population average value of X is then

$$\frac{1(150) + 2(450) + 3(1950) + \cdots + 7(300)}{15000} = 4.57 \tag{5}$$

Since $\frac{150}{15000} = .01 = p(1)$, $\frac{450}{15000} = .03 = p(2)$, and so on, an alternative expression for the population average value of X is

$$1p(1) + 2p(2) + 3p(3) + \cdots + 7p(7) = \sum_{\forall x} xp(x) \tag{6}$$

This shows that to compute the population average value of X, we need only the possible values of X along with their probabilities (proportions).

In particular, the population size is irrelevant as long as the probability mass function is given. The average or mean value of X is then a weighted average of the possible values $1, \cdots, 7$, where the weights are the probabilities of those values.

NB: The mean value of a random variable X is called the **expected value** of (or **expectation** of) the random variable X and is denoted $E(X)$.

EXPECTATION OF A RANDOM VARIABLE

Let $g(x)$ denote any function of X; then we define the expected value of $g(x)$, $E[g(x)]$ as

$$E[g(x)] = \begin{cases} \sum_{\forall x} g(x)p(x), & \text{for } x \text{ discrete} \\ \int_{-\infty}^{\infty} g(x)f(x), & \text{for } x \text{ continuous} \end{cases}$$

Example 1

A coin is thrown 3 times. X denotes the number of heads recorded. Let $g(x) = 2x + 1$. Find $E[g(x)]$.

Solution

x	0	1	2	3
$p(x) = p(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$g(x) = 2x + 1$	1	3	5	7
$g(x)p(x)$	$\frac{1}{8}$	$\frac{9}{8}$	$\frac{15}{8}$	$\frac{7}{8}$

$E[g(x)] = \frac{1}{8} + \frac{9}{8} + \frac{15}{8} + \frac{7}{8} = 4$

Example 2

A random variable Y has probability function

$$f(y)] = \begin{cases} 2y, & 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Determine $E[2y + 3]$.

Solution

$$\begin{aligned} E[2y + 3] &= \int_0^1 (2y + 3)2y\,dy \\ &= \int_0^1 (4y^2 + 6y)\,dy \\ &= \left[\frac{4y^3}{3} + 3y^2\right]_0^1 \\ &= \left(\frac{4}{3} + 3\right) - (0 + 0) \\ &= \frac{13}{3} \end{aligned}$$

Special case

Let $g(x) = x$ then

$$E[x] = \begin{cases} \sum_{\forall x} xp(x), & \text{for } x \text{ discrete} \\ \int_{-\infty}^{\infty} xf(x), & \text{for } x \text{ continuous} \end{cases}$$

This special type of expectation is called the mean of the random variable and is denoted by μ. Hence $\mu = E[X]$. It is also referred to as the 1^{st} moment of the random variable X about the origin.

$E[X^r]$ is the r^{th} moment of the random variable X about the origin.

The mean of a random variable X is it's 1^{st} moment about the origin.

In Example 1, the mean of the random variable X is given by

$$\mu = E[X] = \sum_{\forall x} xp(x) = \left(0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8}\right) = 1.5$$

Properties of the Mean

Theorem 1

Let X be a random variable and k a real number then:

1.) $E[kX] = kE[X]$

2.) $E[X + k] = E[X] + k$

Class Exercise 1

Prove Theorem 1 for X discrete and continuous.

Special Type of Function

Let $g(x) = (x - \mu)^2$ then where μ is the mean of the random variable X then

$$E[(x-\mu)^2] = \begin{cases} \sum_{\forall x}(x-\mu)^2 p(x), & \text{for } x \text{ discrete} \\ \int_{-\infty}^{\infty}(x-\mu)^2 f(x), & \text{for } x \text{ continuous} \end{cases}$$

This special type of expectation is called the variance of the random variable and is denoted by $\sigma^2 = Var(X)$. It is also referred to as the 2^{nd} moment of the random variable X about the mean.

Class Exercise 2

Show that the 1^{st} moment of a random variable X about its mean is 0.

In general $E[(X - A)^r]$ refers to the r^{th} moment of the random variable X about A.

In Example 1, the variance of the random variable X is given by

$$\begin{aligned} \sigma^2 &= E[(X-\mu)^2] \\ &= E[(X-1.5)^2] \\ &= \sum_{\forall x}(X-1.5)^2 p(x) \\ &= \left[\left(\frac{9}{4}\times\frac{1}{8}\right)+\left(\frac{1}{4}\times\frac{3}{8}\right)+\left(\frac{1}{4}\times\frac{3}{8}\right)+\left(\frac{9}{4}\times\frac{1}{8}\right)\right] \\ &= 0.75 \end{aligned}$$

Theorem 2

Given that X is a random variable

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu Ex + E[\mu^2] \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

The variance is the 2^{nd} moment about the origin minus the 1^{st} moment about the origin squared.

Properties of the Variance

Theorem 3

Let X be a random variable and k a real number then:

1.) $Var[kX] = k^2 Var[X]$

2.) $Var[X + k] = Var[X]$

Class Exercise 3

Prove Theorem 3 for X discrete and continuous.

Example 3

A random variable Y has probability density function

$$f(y) = \begin{cases} \frac{y}{4}, & 1 \leqslant y \leqslant 3 \\ 0, & \text{elsewhere} \end{cases}$$

Find: $E(y)$, $E(2y-3)$, $Var(y)$ and $Var(2y-3)$.

Solution

$$
\begin{aligned}
\mathsf{E}(y) &= \int_1^3 \frac{y^2}{4} dy \\
&= \left[\frac{y^3}{12}\right]_1^3 \\
&= \frac{13}{6}
\end{aligned}
$$

$$\mathsf{E}(2y-3) = 2\mathsf{E}(y) - 3 = \frac{13}{3} - 3 = \frac{4}{3}$$

$$
\begin{aligned}
\mathsf{Var}(y) &= \mathsf{E}(y^2) - [\mathsf{E}(y)]^2 \\
&= \int_1^3 \frac{y^3}{4} dy - \frac{169}{36} \\
&= \left[\frac{y^4}{16}\right]_1^3 - \frac{169}{36} \\
&= \frac{80}{16} - \frac{169}{36} \\
&- \frac{11}{36}
\end{aligned}
$$

$$\mathsf{Var}(2y-3) = 2^2\mathsf{Var}(y) = 4 \times \frac{11}{36} = \frac{11}{9}.$$

Class Exercise 4

1. Telephone calls arriving at a company are referred immediately by the receptionist to other people working in the company. The time a call takes in minutes is modeled by the continuous random variable T having the probability density function

$$f(t) = \begin{cases} kt^2, & 0 \leqslant t \leqslant 10 \\ 0, & \text{elsewhere} \end{cases}$$

   a) Find k, $\mathsf{E}(t)$ and $\mathsf{Var}(t)$.

   b) Find the probability of a call lasting between 7 and 9 minutes.

   c) Sketch the probability density function of T.

2. Two fair ten shilling coins are tossed. The random variable X represents the total value of the coin that lands heads up.

   a) Find $\mathsf{E}(\mathsf{X})$ and $\mathsf{Var}(\mathsf{X})$.

   b) Random variables S and T are defined as follows: $\mathsf{S} = \mathsf{X} - 10$ and $\mathsf{T} = \frac{1}{2}\mathsf{X} - 5$. Show that $\mathsf{E}(\mathsf{S}) = \mathsf{E}(\mathsf{T})$.

   c) Find $\mathsf{Var}(\mathsf{S})$ and $\mathsf{Var}(\mathsf{T})$.

Moment Generating Functions

Definition: $E(X^r)$ defines the r^{th} moment of the random variable X about the origin.

Moment Generating Functions (mgfs) are used to generate moments of random variables about the origin.

Definition: The Moment Generating Function of the random variable X is defined as

$$M_X t = E(e^{tx}) = \begin{cases} \sum_{\forall x} e^{tx} p(x), & \text{for } x \text{ discrete} \\ \int_{-\infty}^{\infty} e^{tx} f(x), & \text{for } x \text{ continuous} \end{cases}$$

where t is a real number, if it exists.

NB: Moment generating functions do not exist for all random variables.

Recall: (from the special results provided earlier)

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + ... = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

Hence

$$e^{tx} = 1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + ... = \sum_{k=0}^{\infty} \frac{t^k x^k}{k!}$$

$$\begin{aligned} M_X(t) &= E(e^{tx}) \\ &= E\left[1 + tx + \frac{t^2x^2}{2!} + \frac{t^3x^3}{3!} + ...\right] \\ &= 1 + tE(x) + \frac{t^2}{2}E(x^2) + \frac{t^3}{6}E(x^3) + ... \end{aligned}$$

Let us define $M_X'(t)$ as the first derivative of $M_X(t)$ with respect to t.

$$M_X'(t) = E(x) + tE(x^2) + \frac{t^2}{2}E(x^3) + \textit{terms with higher powers ot } t \tag{7}$$

Let $t = 0$ in Equation (12)

$$M_X'(0) = E(x) \tag{8}$$

Let us define $M_X''(t)$ as the second derivative of $M_X(t)$ with respect to t. We get this by differentiating Equation (12) with respect to t to get

$$M_X''(t) = E(x^2) + tE(x^3) + \textit{terms with higher powers ot } t \tag{9}$$

Let $t = 0$ in Equation (20)

$$M_X''(0) = E(x^2) \tag{10}$$

In general, finding the n^{th} derivative of $M_X(t)$ with respect to t and letting $t = 0$ gives the n^{th} moment of the random variable X about the origin i.e.

$$M_X^{(n)}(0) = E(x^n) \tag{11}$$

Hence, we are able to use moment generating functions of random variables if they exist to find the mean and variance of a random variable since

$$\begin{aligned}
E(X) &= M_X^{'}(0) \\
Var(X) &= E(X^2) - [E(X)]^2 \\
&= M_X^{''}(0) - [M_X^{'}(0)]^2
\end{aligned}$$

Example 1

The probability mass function of a random variable X is given by

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0, 1, 2, ..., n \quad \text{where } p + q = 1 \\ 0, & \text{otherwise} \end{cases}$$

Determine the moment generating function of X. Hence determine $E(X)$ and $Var(X)$.

Solution

$$\begin{aligned}
M_X(t) &= E(e^t x) \\
&= \sum_{\forall x} e^{tx} p(x) \\
&= \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{n-x} \\
&= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} \\
&= (pe^t + q)^n
\end{aligned}$$

$$\begin{aligned}
M_X^{'}(t) &= npe^t(pe^t + q)^{n-1} \\
M_X^{''}(t) &= n(n-1)(pe^t)^2(pe^t + q)^{n-2}
\end{aligned}$$

$$\begin{aligned}
E(X) &= M'_X(0) \\
&= np \\
Var(X) &= M''_X(0) - [M'_X(0)]^2 \\
&= n(n-1)p^2 - (np)^2 \\
&= np(1-p)
\end{aligned}$$

Theorems on Moment Generating Functions

Theorem 1

$M_{cX}(t) = M_X(ct)$, c being a constant.

Proof By definition:

$L.H.S = M_{cX}(t) = E(e^{tcX})$

$R.H.S = M_X(ct) = E(e^{ctX}) = L.H.S$

Theorem 2

The moment generating function of the sum of a number of independent random variables is equal to the product of their respective moment generating functions. Symbolically, if $X_1, X_2, \ldots, X_n$ are independent random variables, then the moment generating function of their sum $X_1 + X_2 + \ldots + X_n$ is given by

$$M_{(X_1+X_2+\ldots+X_n)}(t) = M_{X_1}(t)M_{X_2}(t)\ldots M_{X_n}(t)$$

Proof By definition:

$$\begin{aligned}
M_{(X_1+X_2+\ldots+X_n)}(t) &= E[e^{t(X_1+X_2+\ldots+X_n)}] \\
&= E[e^{t(X_1)}e^{t(X_2)}...e^{t(X_n)}] \\
&= E[e^{t(X_1)}]E[e^{t(X_2)}]...E[e^{t(X_n)}] \quad \text{because the } X_i's \text{ are independent} \\
&= M_{X_1}(t)M_{X_2}(t)\ldots M_{X_n}(t)
\end{aligned}$$

Hence the theorem.

Theorem 3

Effect of change of origin and scale on the MGF. Let us transform X to the new variable U by changing both the origin and scale in X as follows:

$$U = \frac{X - a}{h}$$

where a and h are constants.

The moment generating function of U (about origin) is given by

$$\begin{aligned} M_U(t) &= E(e^{tU}) \\ &= E\left(e^{t\left(\frac{X-a}{h}\right)}\right) \\ &= E\left[e^{\frac{tX}{h}} \times e^{-\frac{at}{h}}\right] \\ &= e^{-\frac{at}{h}} E\left[e^{\frac{tX}{h}}\right] \\ &= e^{-\frac{at}{h}} M_X\left(\frac{t}{h}\right) \end{aligned}$$

where $M_X(t)$ is the moment generating function of X about the origin.

Theorem 4

The moment generating function of a distribution, if it exists, uniquely determines the distribution. This implies that corresponding to a given probability distribution, there is only one moment generating function (provided it exists) and corresponding to a given moment generating function, there is only one probability distribution. Hence $M_X(t) = M_Y(t) \Rightarrow X$ and Y are identically distributed.

Limitation of the Moment Generating Function

1. A random variable X may have no moments although its moment generation function exists.

2. A random variable X can have a moment generating function and some (or all) moments, yet the moment generation function does not generate the moments.

3. A random variable X can have all or some moments; but the moment generation function does not exist except perhaps at one point.

Class Exercise

1. Define the moment generating function of a random variable. Hence or otherwise find the moment generating function of

   a) $Y = aX + b$

   b) $Y = \dfrac{X - m}{\sigma}$

2. The random variable X takes the value n with probability $\dfrac{1}{2^n}, n = 1, 2, 3, \ldots$. Find the moment generating function of X and hence find the mean and variance of X.

3. Show that if $\bar{X}$ is the mean of n independent random variables, then

$$M_{\bar{X}}(t) = \left[M_X\left(\frac{t}{n}\right)\right]^n$$

4. Show that the moment generating function of the random variable X having the probability density function

$$f(x) = \begin{cases} \dfrac{1}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

   is

$$M_X(t) = \begin{cases} \dfrac{e^{2t} - e^{-t}}{3t}, & t \neq 0 \\ 1, & t = 0 \end{cases}$$

5. X is a random variable and $p(x) = ab^x$ where a and b are positive, $a + b = 1$ with x taking the values $0, 1, 2, \ldots$. Find the moment generating function of X. Hence show that $m_2 = m_1(2m_1 + 1)$, m_1 and m_2 being the first two moments.

SOME SPECIAL RANDOM VARIABLES

These are random variables which occur frequently in applied statistics or in practice. Some of them are discrete while others are continuous.

Discrete random variables

These include uniform, bernoulli, binomial, Poisson, negative binomial, geometric, hypergeometric, etc.

Continuous random variables

These include uniform (rectangular), beta, exponential, normal, gamma, chi-square etc.
We will study some of these distributions. At the end of it all, you should be able to:

a) state the properties of the random variables described by these distributions.

b) define the distributions by stating their probability functions, their cumulative density functions for continuous distributions, their mean, variances and moment generating functions if they exist.

c) solve probability questions involving these distributions.

The Discrete Uniform distribution

Definition

A random variable X is said to be a uniform discrete random variable if each of the outcomes of the random variable are equally likely; that is if its probability mass function is given by:

$$p(x) = p(X = x) = \begin{cases} \dfrac{1}{n}, & \text{where } n \text{ is the total number of sample points in the sample space} \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

Is Equation (12) a pmf?
$p(x) > 0$ and $\sum_{\forall x} \frac{1}{n} = n \times \frac{1}{n} = 1$. So yes, it is a pmf.
Consider the probability distribution for the score X on a single roll of a die. It is

$$p(x) = p(X = x) = \begin{cases} \dfrac{1}{6}, & x = 1, 2, 3, 4, 5, 6 \\ 0, & \text{otherwise} \end{cases}$$

This is an example of a discrete uniform distribution over the set of values $\{1, 2, 3, 4, 5, 6\}$. It is called discrete because the values are discrete and it is called uniform because all the probabilities are the same.
Conditions for a discrete uniform distribution are:

- ✓ A discrete random variable defined over a set of n distinct values.
- ✓ Each value is equally likely, i.e. $p(X = x) = \frac{1}{n}$ for each x.

Distributions have to be defined. Defining a distribution means getting its mean and variance.

Mean and Variance

In many cases X is defined over the set $\{1, 2, ..., n\}$, in such cases the mean (μ) and the variance (σ^2) are given by:

$$\begin{aligned}
\mu = E(X) &= \sum_{x=1}^{n} \frac{x}{n} \\
&= \frac{1}{n}(1 + 2 + ... + n) \\
&= \frac{1}{n} \times \frac{n(n+1)}{2} \\
&= \frac{n+1}{2}
\end{aligned}$$

$$\begin{aligned}
\sigma^2 = Var(X) &= E(X^2) - [E(X)]^2 \\
&= \sum_{x=1}^{n} \frac{x^2}{n} - \frac{(n+1)^2}{4} \\
&= \frac{1}{n}(1^2 + 2^2 + ... + n^2) - \frac{(n+1)^2}{4} \\
&= \frac{1}{n} \times \frac{n(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} \\
&= \frac{n^2-1}{12}
\end{aligned}$$

The moment generating function of a discrete uniform distribution does not exist.

The Bernoulli distribution

Bernoulli trail

A Bernoulli trial is a trail which has:

- ✓ Two outcomes technically termed 'success' and 'failure' and takes values 0 and 1.
- ✓ Probability of success denoted p is constant in each trail. This implies the probability of failure denoted $q = (1 - p)$.

For example when you toss a coin, you only have two outcomes; Head (H) or Tail (T). The process of tossing a coin i.e. getting a Head or a Tail is called a Bernoulli trial.

Definition

A random variable X is said to have a Bernoulli distribution if its probability mass function is given by

$$p(x) = p(X = x) = \begin{cases} p^x q^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

Is Equation (20) a pmf?

$p(x) > 0$ and $\sum_{x=0}^{1} p^x q^{1-x} = q + p = 1$. So yes, it is a pmf.

Mean and Variance

$$\mu = E(X) = \sum_{x=0}^{1} xp^x q^{1-x} \qquad\qquad \sigma^2 = Var(X) = E(X^2) - [E(X)]^2$$

$$= p \qquad\qquad = \sum_{x=0}^{1} x^2 p^x q^{1-x} - p^2$$

$$= p - p^2 = p(1-p)$$

$$= pq$$

Moment generating function

$$\begin{aligned} M_X(t) = E(e^{tx}) &= \sum_{\forall x} e^{tx} p(x) \\ &= \sum_{x=0}^{1} e^{tx} p^x q^{(1-x)} \\ &= (pe^t + q) \end{aligned}$$

Does it generate moments?

$M_X'(t) = pe^t$ and $M_X''(t) = pe^t$.

Hence $E(X) = M_X'(0) = p$ and $Var(X) = M_X''(0) - [M_X'(0)]^2 = p - p^2 = pq$.

The Binomial distribution

Using factorial notation to find the number of arrangements of some objects.

Example 1

Find all possible arrangements of

a) 3 objects where one is red, one is blue and one is green.

b) 4 objects where 2 are red and 2 are blue.

Solution

a) The first object can be chosen in 3 ways. The second object can be chosen in 2 ways. The third object can be chosen in 1 way. Hence there are $3 \times 2 \times 1 = 6$ possible ways.

b) If the red objects are labeled R_1 and R_2 and the blue objects B_1 and B_2, then you can treat the objects as 4 different ones and there are 4! ways of arranging them. However, arrangements with R_1R_2 are identical to the arrangements with R_2R_1 and so the total number needs to be divided by 2. A similar argument applies to the Bs and so the number of arrangements is $\frac{4!}{2! \times 2!} = 6$

So that

- ✓ n different objects can be arranged in $n! = n \times (n-1) \times \ldots \times 3 \times 2 \times 1$ ways.
- ✓ n objects with a of one type and $(n-a)$ of another can be arranged in $\binom{n}{a} = \frac{n!}{a!(n-a)!}$ ways.

Consider an experiment where we have:

- ✓ A fixed number of trials, n.
- ✓ Each trial has two possible outcomes – Success or Failure.
- ✓ The probability of success, p, in each trial is constant.
- ✓ The trials are independent.

If these four conditions are satisfied we say the random variable X which denotes the number of successes in a fixed number of trials has a binomial distribution and we write $X \sim B(n,p)$. A binomial experiment is a random experiment involving a sequence of independent and identical Bernoulli trials.

Suppose the outcome of n bernoulli trials is x successes and $(n-x)$ failures. The probability of this occurrence is

$$\begin{aligned} SSFSFFSS \cdots FFS &= SSS \ldots SFFF \ldots F \\ &= p(S)p(S) \cdots p(S)p(F)p(F) \cdots p(F) \\ &= p^x q^{n-x} \end{aligned}$$

Yet, the number of ways can we get x successes from n trials is $\binom{n}{x}$ ways. Hence the probability of getting x successes in n trials is

$$\binom{n}{x} p^x q^{n-x}$$

Definition

A random variable X is said to have a binomial distribution if its probability mass function is given by

$$p(x) = p(X = x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0, 1, ..., n \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Is Equation (21) a pmf?

$p(x) > 0$ and $\sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = \left(\binom{n}{0} p^0 q^n + \binom{n}{1} p^1 q^{n-1} + \cdots + \binom{n}{n} p^n q^0\right) = (p+q)^n = 1.$

So yes, it is a pmf.

Before we compute the mean and variance of the Binomial distribution, I want you to confirm that the following two results are true.

Result 1

$$\begin{aligned}\sum_{x=0}^{n} \binom{n}{x} p^x q^{(n-x)} &= \sum_{x=1}^{n} \binom{n-1}{x-1} p^{(x-1)} q^{(n-1-(x-1))}, \\ &= \sum_{x=2}^{n} \binom{n-2}{x-2} p^{(x-2)} q^{(n-2-(x-2))}, \\ &= 1.\end{aligned}$$

Result 2

$$\begin{aligned}\text{Recall} \quad Var(X) &= E(X^2) - [E(X)]^2, \\ \text{yet} \quad E[X(X-1)] &= E(X^2) - E(X), \\ E(X^2) &= E[X(X-1)] + E(X), \\ \text{hence} \quad Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2.\end{aligned}$$

Mean and Variance

$$\begin{aligned}\mu = E(X) &= \sum_{x=0}^{n} x \binom{n}{x} p^x q^{(n-x)} \\ &= \sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{(n-x)} \\ &= np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{(x-1)} q^{(n-1-(x-1))} \\ &= np\end{aligned}$$

$$\begin{aligned}\sigma^2 = Var(X) &= E[X(X-1)] + E(X) - [E(X)]^2 \\ &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^{(n-x)} + np - n^2p^2 \\ &= n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2} q^{(n-2-(x-2))} + np - n^2p^2 \\ &= n(n-1)p^2 + np - n^2p^2 \\ &= npq\end{aligned}$$

Moment generating function

$$M_X(t) = E(e^{tx}) = \sum_{\forall x} e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x q^{(n-x)} = \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{(n-x)} = (pe^t + q)^n.$$

Does it generate moments?

$$
\begin{aligned}
M_X(t) &= (pe^t + q)^n \\
M_X^{'}(t) &= npe^t(pe^t + q)^{n-1} \\
M_X^{''}(t) &= n(n-1)(pe^t)^2(pe^t + q)^{n-2} + npe^t(pe^t + q)^{n-1} \\
E(X) &= M_X^{'}(0) = np. \\
Var(X) &= M_X^{''}(0) - [M_X^{'}(0)]^2 = n(n-1)p^2 + np - n^2p^2 = np(1-p) = npq.
\end{aligned}
$$

Exercise

1. Let $X \sim B(n, p)$ with $E(X) = 2$ and $Var(X) = \frac{4}{3}$. Find $p(X > 4)$.

2. Assume that on average one telephone number out of 15 called between 2 p.m. and 3 p.m. on weekdays is busy. What is the probability that if 6 randomly selected telephone numbers are called

   i) Not more than 3

   ii) At least 3 of them are busy?

3. Consider an urn containing M balls k of which are red and the rest black. Let n balls be picked randomly with replacement. Let X be the number of red balls in a sample of size n. Show that $X \sim B(n, p)$.

4. Learn how to use the Cumulative Binomial tables to work out probabilities.

The Poisson distribution

Recall: The function e^x can be defined as a series

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \tag{15}$$

Let $x = \lambda$ in Equation (22) we get

$$e^\lambda = 1 + \frac{\lambda}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \tag{16}$$

dividing both sides of Equation (23) by e^λ gives

$$1 = \frac{\lambda^0 e^{-\lambda}}{0!} + \frac{\lambda^1 e^{-\lambda}}{1!} + \frac{\lambda^2 e^{-\lambda}}{2!} + \frac{\lambda^3 e^{-\lambda}}{3!} + \cdots = \sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} \tag{17}$$

Notice that the sum of the infinite series on the right hand side of Equation (24) equals 1 and you could use these values as probabilities to define a probability distribution.

Definition

A random variable X is said to have a Poisson distribution if its probability mass function is given by

$$p(x) = p(X = x) = \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, ... \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

From Equation (24), Equation (25) is a probability mass function. We say that X has a Poisson distribution with parameter λ and denote this as $X \sim P_o(\lambda)$.

Exercise

The random variable $X \sim P_o(1.2)$; find $p(X = 3)$; $p(X \geqslant 1)$; $p(3 < X \leqslant 5)$.

Mean and Variance

$$\begin{aligned} E(X) &= \sum_{\forall x} xp(x) \\ &= \sum_{x=0}^{\infty} x \frac{\lambda^x e^{-\lambda}}{x!} \\ &= \sum_{x=1}^{\infty} x \frac{\lambda^1 \lambda^{x-1} e^{-\lambda}}{x(x-1)!} \\ &= \lambda \sum_{x=1}^{\infty} \frac{\lambda^{x-1} e^{-\lambda}}{(x-1)!} \\ &= \lambda. \end{aligned}$$

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

The fact that the mean equals the variance is an important property of a Poisson distribution. The presence or absence of this property can be a useful indicator of whether or not a Poisson distribution is a suitable model for a particular situation.

Deciding whether or not a Poisson distribution is a suitable model

The random variable X represents the number of events that occur in an interval. The interval may be a fixed length in space or time.

If X is to have a Poisson distribution then the events must occur

- ✓ **singly** in space or time
- ✓ **independently** of each other
- ✓ at a **constant rate** in the sense that the mean number of occurrences in the interval is proportional to the length of the interval.

Such events are said to occur **randomly**. You can use the constant rate idea to adjust the parameter of a Poisson distribution.

The Poisson approximation to Binomial

In some cases you can approximate a binomial distribution with a Poisson distribution. Evaluating binomial probabilities when n is large can be quite difficult and in such circumstances it is sometimes simpler to use an approximation.
If $X \sim B(n, p)$ and

- n is large
- p is small

then X can be approximated by $P_o(np = \lambda)$.
If $np = \lambda$ then $p = \dfrac{\lambda}{n}$

$$\begin{aligned}
p(x) &= \binom{n}{x} p^x q^{n-x}, x = 0, 1, ... \\
&= \binom{n}{x}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x}, x = 0, 1, ... \\
&= \frac{n!}{x!(n-x)!}\left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&= \frac{n(n-1)(n-2) \times ... \times (n-(x+1))(n-x)!}{x!(n-x)!} \times \frac{\lambda^x}{n^x} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1-\frac{\lambda}{n})^x} \\
&= \frac{n(n-1)(n-2) \times ... \times (n-(x+1))}{n^x} \times \frac{\lambda^x}{x!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1-\frac{\lambda}{n})^x} \\
&= \frac{n}{n} \times \frac{n-1}{n} \times \frac{n-2}{n} \times ... \times \frac{n-(x+1)}{n} \times \frac{\lambda^x}{x!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1-\frac{\lambda}{n})^x} \\
&= 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times ... \times \left(1 - \frac{(x+1)}{n}\right) \times \frac{\lambda^x}{x!} \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1-\frac{\lambda}{n})^x} \\
\lim_{n\to\infty} p(x) &= \frac{\lambda^x}{x!} \lim_{n\to\infty} 1 \times \left(1 - \frac{1}{n}\right) \times \left(1 - \frac{2}{n}\right) \times ... \times \left(1 - \frac{(x+1)}{n}\right) \times \left(1 - \frac{\lambda}{n}\right)^n \times \frac{1}{(1-\frac{\lambda}{n})^x} \\
p(x) &= \frac{\lambda^x}{x!} \lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n \\
p(x) &= \frac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, ...
\end{aligned}$$

There is no clear rule as to what constitutes **large** n or **small** p but usually the value of np will be $\leqslant 10$ so that the Poisson table can be used. Generally the larger the value of n and the smaller the value of p the better the approximation will be.

Moment generating function

$$\begin{aligned}
M_X(t) = E(e^{tx}) &= \sum_{\forall x} e^{tx} p(x) \\
&= \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x e^{-\lambda}}{x!} \\
&= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(e^t\lambda)^x}{x!} \\
&= e^{-\lambda} \times e^{e^t\lambda} \\
&= e^{\lambda(e^t-1)}
\end{aligned}$$

Does it generate moments?

$$\begin{aligned}
M_X(t) &= e^{\lambda(e^t-1)} \\
M_X'(t) &= \lambda e^t e^{\lambda(e^t-1)} \\
M_X''(t) &= (\lambda e^t)^2 e^{\lambda(e^t-1)} + \lambda e^t e^{\lambda(e^t-1)} \\
E(X) &= M_X'(0) = \lambda. \\
Var(X) &= M_X''(0) - [M_X'(0)]^2 = \lambda(\lambda+1) - \lambda^2 = \lambda.
\end{aligned}$$

Exercise

1. A car hire firm has 2 cars which it hires out day by day. The number of demands for a car each day is distributed as a Poisson variable with mean 1.5. Calculate the proportion of days on which

   a) neither car is used.

   b) some demand is refused.

2. Look up the following distributions and study them like we have done for the above distributions

   a) Hypergeometric distribution

   b) Geometric distribution

   c) Negative-binomial distribution

THE RECTANGULAR DISTRIBUTION

Introduction

Models a random variable where probability is constant (or the same) over a given interval.

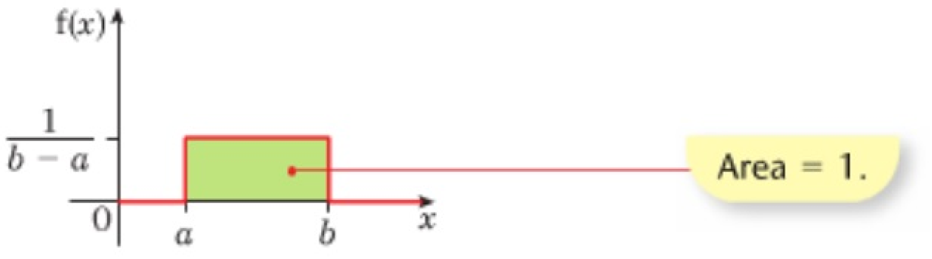
Definition

The continuous random variable X with probability density function is given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leqslant x \leqslant b \\ 0, & \text{elsewhere} \end{cases} \tag{19}$$

where a and b are constants is called a continuous uniform (or rectangular)distribution. It is denoted by $X \sim U(a, b)$.

A sketch of the p.d.f. is as follows.



Properties of the rectangular distribution

$$E(X) = \frac{a+b}{2}$$
$$Var(X) = \frac{(b-a)^2}{12}$$
$$F(x) = \frac{x-a}{b-a}$$

Class Exercise 1

Derive the properties of the rectangular distribution.

$$E(X) = \int_a^b x\frac{1}{b-a}dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{a+b}{2}$$

$$Var(X) = \int_a^b x^2\frac{1}{b-a} - \frac{(a+b)^2}{4} = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b - \frac{(a+b)^2}{4} = \frac{(b-a)^2}{12}$$

THE EXPONENTIAL DISTRIBUTION

Introduction

If an engineer is responsible for the quality of, say, copper wire for use in domestic wiring systems, he or she might be interested in knowing both the number of faults in a given length of wire and also the distances between such faults. While the number of faults may be analyzed by using the Poisson distribution, the distances between faults along the wire may be shown to give rise to the exponential distribution.

In probability theory and statistics, the exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate.

The exponential distribution describes the time for a continuous process to change state. (It models the waiting time until a Poisson happening).

- ✓ How long do we have to wait before a customer enters our shop?
- ✓ How long will it take before a call center receives the next call?

Illustration 1

Suppose X is the number of customers arriving at a bank in an interval of length 1.

X is a Poisson random variable. Let the mean number of events in the interval of length 1 be λ.

The waiting time until the 1^{st} customer arrives is a continuous random variable - $W = w$.

Can we determine $F(w) = p(W \leqslant w)$?

$$
\begin{aligned}
F(w) =& p(W \leqslant w)\\
=& 1 - p(W > w)\\
=& 1 - p(\text{ no customers arrive in } [0, w])\\
=& 1 - p(X = 0 \text{ with mean } \lambda w)\\
=& 1 - \frac{e^{-\lambda w}(\lambda w)^0}{0!}\\
=& 1 - e^{-\lambda w}, w > 0
\end{aligned}
$$

NB: If the mean number of events in an interval of length 1 is λ then the mean number of events in the interval of length w is λw.

Therefore

$$f(w) = F'(w) = -e^{-\lambda w}(-\lambda) = \lambda e^{-\lambda w}, 0 < w < \infty$$

Hence if λ is the mean number of events in an interval and θ is the mean waiting time until the first customer arrives then $\theta = \frac{1}{\lambda}$ and $\lambda = \frac{1}{\theta}$. The continuous random variable X follows and exponential distribution if its pdf is

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{1}{\theta} x}, & \theta > 0, x \geqslant 0 \\ 0, & \text{elsewhere} \end{cases}$$

Definition

A random variable X is said to have an exponential distribution if its probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqslant 0 \\ 0, & \text{elsewhere} \end{cases} \tag{20}$$

Here $\lambda > 0$ is the parameter of the distribution, often called the rate parameter. The distribution is supported on the interval $[0, \infty)$. If a random variable X has this distribution, we write $X \sim \mathrm{Exp}(\lambda)$.

NB: The parameterization involving the 'rate' parameter arises in the context of events arriving at a rate λ, when the time between events (which might be modeled using an exponential distribution) has a mean of $\beta = \lambda^{-1}$.

Is it a probability density function?

It should satisfy the properties

i) $f(x) \geqslant x$, i.e. probability exists and

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

$$
\begin{aligned}
\int_{-\infty}^{\infty} f(x)dx &= \int_{0}^{\infty} \lambda e^{-\lambda x} dx \\
&= -\frac{\lambda}{\lambda} \left[e^{-\lambda x}\right]_0^\infty \\
&= -1(0-1) = 1
\end{aligned}
$$

CDF of an exponential distribution

$$
\begin{aligned}
F(x) = p(X \leqslant x) = \int_{-\infty}^{x} f(x)dx &= \int_{-\infty}^{0} 0dx + \int_{0}^{x} \lambda e^{-\lambda x} dx \\
&= 0 + \left[-e^{\lambda x}\right]_0^x \\
&= 1 - e^{\lambda x}.
\end{aligned}
$$

MGF of an exponential distribution

$$
\begin{aligned}
M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)dx &= \int_{0}^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\
&= \lambda \int_{0}^{\infty} e^{-x(\lambda - t)} dx \\
&= -\frac{\lambda}{(\lambda - t)} \left[e^{-x(\lambda-t)}\right]_0^\infty \\
&= -\frac{\lambda}{(\lambda - t)} [0-1] \\
&= \frac{\lambda}{(\lambda - t)} \\
&= \lambda(\lambda - t)^{-1}.
\end{aligned}
$$

Mean and Variance of an exponential distribution

$$
\begin{aligned}
M_X(t) &= \lambda(\lambda - t)^{-1} \\
M'_X(t) &= \lambda(\lambda - t)^{-2} \\
M''_X(t) &= 2\lambda(\lambda - t)^{-3} \\
E(X) = M'_X(0) &= \frac{1}{\lambda} \\
Var(X) = M''_X(0) - [M'_X(0)]^2 &= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} \\
&= \frac{1}{\lambda^2}
\end{aligned}
$$

Class Exercise

1. Let be an exponential random variable with parameter $\lambda = \ln 3$. Compute $p(2 \leqslant X \leqslant 4)$.
2. What is the probability that a random variable is less than its expected value, if has an exponential distribution with parameter λ?
3. The time in hours an electric bulb takes to burn is a random phenomenon which obeys an exponential law with $\lambda = 0.001$. Find the probability that it will last more than 50 hours.

THE NORMAL DISTRIBUTION

Introduction

Mass-produced items should conform to a specification. Usually, a mean is aimed for but due to random errors in the production process a tolerance is set on deviations from the mean. For example if we produce piston rings which have a target mean internal diameter of 45 mm then realistically we expect the diameter to deviate only slightly from this value. The deviations from the mean value are often modeled very well by the normal distribution.

Definition

A random variable X is said to have a normal distribution with mean (μ) and variance (σ^2) if its probability density function is given by

$$
f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0 \\ 0, & \text{elsewhere} \end{cases} \tag{21}
$$

Many of the techniques in applied statistics are based on the normal distribution.

MGF of a normal distribution

$$M_X(t) = E[e^{tx}] = \int_{-\infty}^{\infty} e^{tx} f(x)dx = \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \tag{22}$$

Using the change of variable technique in Equation (22), let $y = \frac{x-\mu}{\sigma}$. This implies $x = \sigma y + \mu$ so that

$$\begin{aligned} \sigma y &= x - \mu \\ \sigma dy &= dx \end{aligned} \tag{23}$$

Substituting Equation (23) in Equation (22) and changing the limits gives

$$M_X(t) = \int_{-\infty}^{\infty} e^{t(\sigma y+\mu)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2}} \sigma dy = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\sigma t y - \frac{y^2}{2} + \mu t} dy \tag{24}$$

Consider the index in Equation (24)

$$\begin{aligned} \sigma t y - \frac{y^2}{2} + \mu t = -\frac{y^2}{2} + \sigma t y + \mu t &= -\frac{1}{2}(y^2 - 2t\sigma y - 2\mu t) \\ &= -\frac{1}{2}(y^2 - 2t\sigma y + t^2\sigma^2 - t^2\sigma^2 - 2\mu t) = -\frac{1}{2}[(y - t\sigma)^2 - (t^2\sigma^2 + 2\mu t)] \\ &= -\frac{1}{2}(y - t\sigma)^2 + \frac{1}{2}(t^2\sigma^2 + 2\mu t) \end{aligned} \tag{25}$$

Therefore

$$\begin{aligned} M_X(t) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t\sigma)^2 + \frac{1}{2}(t^2\sigma^2+2\mu t)} dy \\ &= e^{\frac{1}{2}(t^2\sigma^2+2\mu t)} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t\sigma)^2} \\ &= e^{\frac{1}{2}(t^2\sigma^2+2\mu t)} \end{aligned}$$

This is because

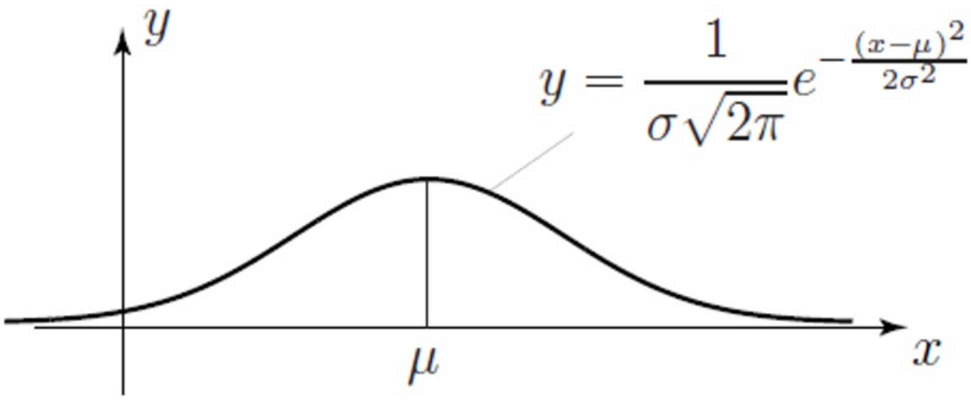
$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y-t\sigma)^2} = 1 \tag{26}$$

since $Y \sim N(t\sigma, 1)$.

Characteristics of the Normal Distribution

A normal distribution is completely defined by specifying its mean (the value of μ) and its variance (the value of σ^2). The normal distribution with mean μ and variance σ^2 is written $N(\mu, \sigma^2)$. Hence the distribution $N(20, 25)$ has a mean of 20 and a standard deviation of 5; remember that the second 'parameter' is the variance which is the square of the standard deviation.

✓ The curve is bell-shaped with the centre of the bell located at the value of μ.



✓ The height of the bell is controlled by the value of σ.

✓ As with all normal distribution curves it is symmetrical about the centre and decays exponentially as $x \longrightarrow \pm\infty$; it is asymptotic to the horizontal axis.

✓ As with any probability density function the area under the curve is equal to 1.

✓ The Empirical rule gives the proportion of the distribution within standard deviations of the mean so that

  ✓ $\mu \pm \sigma$ includes 68.2%.

  ✓ $\mu \pm 2\sigma$ includes 95.4%.

  ✓ $\mu \pm 3\sigma$ includes 99.97%.

CDF of a Normal distribution

$$F(x) = p(X \leqslant x) = \int_{-\infty}^{x} f(x)dx = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx$$

The cdf of a normal random variable is denoted $\Phi(x)$.

Standard Normal distribution

A standard normal random variable is denoted by Z. It is a special normal random variable with mean $\mu = 0$ and variance $\sigma^2 = 1$ i.e. $Z \sim N(0, 1)$. Hence

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}, -\infty \leqslant z \leqslant \infty$$

$$M_Z(t) = e^{-\frac{t^2}{2}}$$

Areas under the standard normal distribution curve are provided in standard normal tables. Every other normal distribution is converted to a standard normal; we say standardizing a normal distribution by

$$Z = \frac{X - \mu}{\sigma}$$

Class Exercise

Use of standard normal table

Example

Let X be a normal random variable with mean $\mu = 100$ and standard deviation $\sigma = 10$. Find $p(90 \leqslant X \leqslant 100)$.

Solution

$$\begin{aligned}
p(90 \leqslant X \leqslant 100) &= \int_{x=90}^{x=100} N(100, 10^2)\,dx \\
&= \int_{z=\frac{90-100}{10}}^{z=\frac{100-100}{10}} N(0,1)\,dz \\
&= \int_{z=-1}^{z=0} N(0,1)\,dz \\
&= \Phi(0) - \Phi(-1) = \Phi(1) - \Phi(0) \quad \text{due to symmetry of the normal distribution} \\
&= (0.5 - 0.1587) = (0.8413 - 0.5) \\
&= 0.3413
\end{aligned}$$

NORMAL APPROXIMATIONS

At the end of this Topic you should know how to

1. approximate a binomial distribution by a Poisson distribution

2. approximate a binomial distribution by a normal distribution

3. approximate a Poisson distribution by a normal distribution

4. apply a continuity correction

Introduction

Sometimes the calculations using a binomial or Poisson distribution can be quite cumbersome because of the numbers involved. In many cases the normal distribution provides a simple and accurate approximation to the probability required.

Using a continuity correction

The binomial and Poisson distributions are both discrete distributions but the Normal distribution is a continuous one. Recall that, for a discrete distribution $p(X \leqslant 5) \neq p(X < 5)$ while $p(X < 5) = p(X \leqslant 4)$. However for a continuous distribution such as the normal, $p(X \leqslant 5) = p(X < 5)$ since $p(X = 5) = 0$.

If you are approximating a discrete distribution, X, by a continuous distribution Y, you need to consider how to treat the decimal values of Y between the discrete values of X.

Recall that if $X = 5$ to the nearest integer, then $4.5 \leqslant X < 5.5$; we use this idea when approximating discrete distributions by normal (or any continuous) distribution and we call it a continuity correction.

Example 1

If X is a discrete distribution, apply a continuity correction to the following probabilities: $p(X \leqslant 6)$; $p(X > 10)$; $p(2 \leqslant X < 5)$.

Solution: $p(X \leqslant 6) \approx p(Y < 6.5)$; $p(X > 10) \approx p(Y \geqslant 10.5)$; $p(2 \leqslant X < 5) \approx p(1.5 \leqslant Y < 4.5)$.

The continuity correction can be summarized in the following simple rules

- ✓ First write your probability using $\leqslant$ or $\geqslant$.
- ✓ For p($X \leqslant n$) you simply approximate by $p(Y < n + 0.5)$.
- ✓ For $p(X \geqslant n)$ you simply approximate by $p(Y \geqslant n - 0.5)$.

You may find these rules helpful but the required continuity correction can always be found using a simple diagram.

Approximating a binomial distribution by a normal distribution

The random variable $X \sim B(n, p)$; Recall: $E[X] = np$ and $Var[X] = np(1 - p)$. If n is large calculating the binomial coefficients can be difficult and if n is very large most calculators will be unable to perform the calculation (try $\binom{500}{240}$ on your calculator).

If p is close to 0.5, then the binomial distribution will be fairly symmetric and so it is reasonable to assume that a normal distribution might provide a suitable approximation.

It is clearly sensible to choose a normal distribution which has the same mean and variance as the original binomial distribution and this leads to the following approximation.

If $X \sim B(n, p)$ and

- n is large
- p is close to 0.5

Then X can be approximated by $Y \sim N(np, np(1-p))$ i.e. $\sigma = \sqrt{(np(1-p))}$. There is no definitive answer to the question of how large n should be or how close to 0.5 p should be. The simple answer is that the larger n is and the closer p is to 0.5 the better. However the approximation does work well for relatively small values of n if p is close to 0.5 as the following example illustrates.

Example 2

The random variable $X \sim B(20, 0.4)$

a) Use tables to find $p(X \leqslant 6)$.

b) Use a normal approximation to estimate $p(X \leqslant 6)$.

Solution

a) $p(X \leqslant 6) = 0.2500$.

b) $X \sim B(20, 0.4)$ so $Y \sim N(20 \times 0.4, 20 \times 0.4 \times 0.6) = Y \sim N(8, 4.8)$; hence $p(X \leqslant 6) = p(Y < 6.5) = p\left(Z < \frac{6.5-8}{\sqrt{4.8}}\right) = p(Z < -0.6846) = 0.2483$.

NB: A calculator would give 0.24678... here both of these values are equal to 0.25 to 2 s.f. thus demonstrating that the approximation has worked well.

Approximating a Poisson distribution by a normal distribution

If the mean of a Poisson distribution is large then a normal approximation can be used. Just like for the binomial distribution we choose the normal distribution to have the same mean as the

original Poisson distribution mean and the same variance as the original Poisson distribution. Hence if λ is large, $X \sim P_o(\lambda)$ can be approximated by $Y \sim N(\lambda, \lambda)$.
As before there is no simple answer, other than the larger the better, to the question of how large λ should be before a Poisson distribution can be approximated to a normal distribution.

Example 3

The random variable $X \sim P_o(25)$ Use a normal approximation to estimate

a) $p(X > 30)$.

b) $p(18 \leqslant X < 35)$.

Solution
$X \sim P_o(25)$ so $Y \sim N(25, 25)$

a) $p(X > 30) \approx p(Y > 30.5) = p(Z > \frac{30.5-25}{5}) = p(Z > 1.1) = 0.1357$.

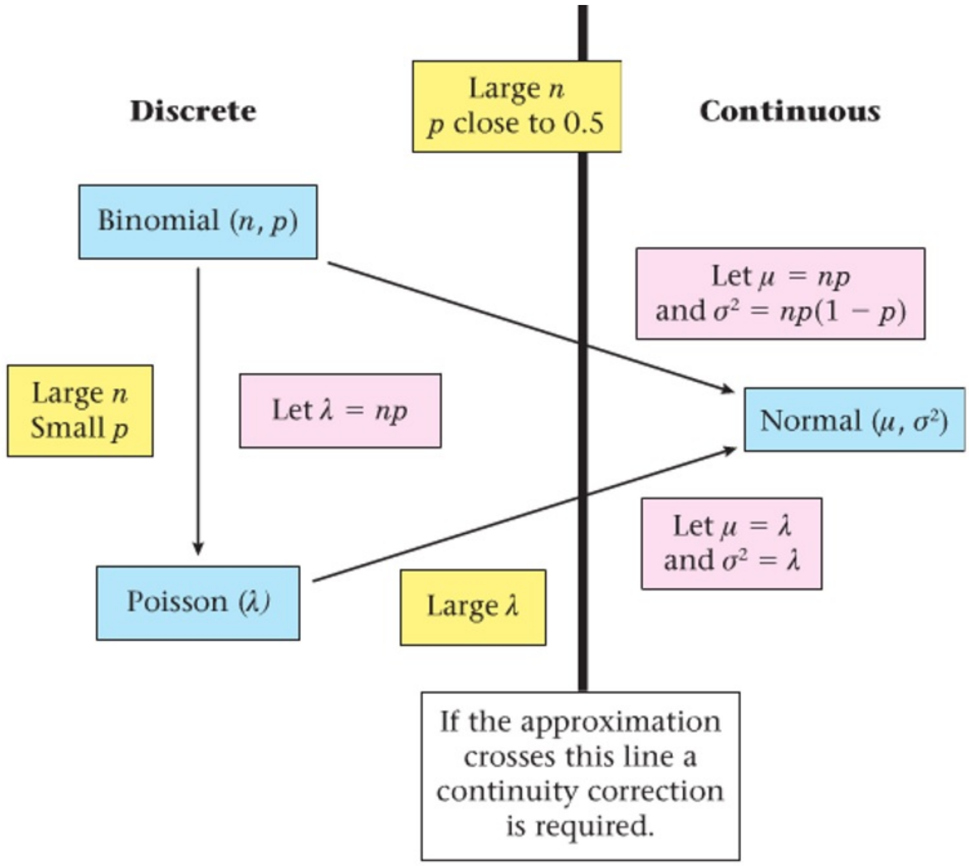
b) $p(18 \leqslant X < 35) = p(17.5 \leqslant X < 34.5) = p(\frac{17.5-25}{5} \leqslant Z < \frac{34.5-25}{5}) = p(-1.5 \leqslant Z < 1.9) = 0.9045$.

Class Exercise

A car hire firm has a large fleet of cars for hire by the day and it is found that the fleet suffers breakdowns at the rate of 21 per week. Assuming that the breakdowns occur at a constant rate, randomly in time and independently of one another, use a suitable approximation to estimate the probability that in any one week more than 27 breakdowns occur. (Answer = 0.0778).

Choosing the appropriate approximation

We have seen that a binomial distribution can sometimes be approximated by a Poisson distribution and sometimes it is approximated by a normal distribution. The approximations you need can be summarized by the following diagram.
If you are approximating a binomial distribution by a normal distribution you should always go directly to the normal distribution and not via a Poisson distribution as this involves one not two approximations and should therefore be more accurate.
For a binomial distribution, there are two possible approximations, depending on whether p lies close to 0.5 (in which case a normal distribution is used) or p is small in which case a Poisson distribution is used).



If in doubt over which approximation is appropriate a useful 'rule of thumb' is to calculate the mean np and, if this is less than or equal to 10, you should be able to use the Poisson tables so that approximation can be used. If the mean is more than 10 then a normal approximation is usually suitable.

Reference

1. Morris H DeGroot & Mark J Schervish (2012) Probability and Statistics; 4^{th} Edition, Pearson Education, Inc.

2. Robert V Hogg, Elliot A. Tanis & Dale L. Zimmerman (2015) Probability & Statistical Inference; 9^{th} Edition, Pearson Education, Inc.