Statistical Techniques for Data Science

Probability and Distributions

Prof.Gangaboraiah PhD (Stats)

Learning objectives of Probability

- At the end of the session on unit 2, the student should be able to
  - Distinguish between Deterministic and Probabilistic experiments
  - Understand different types of probability and compute and interpret the results obtained

Consider the equations

1. $v = u + at$,
   - u = Initial velocity
   - v = Final velocity
   - a = Acceleration
   - t = Time

2. $I = PRT/100$,
   - I = Simple Interest
   - P = Principal amount
   - R = Rate of interest
   - T = Duration of time

Deterministic

What is the difference between these statements?

Consider the example

1. Can it rain tomorrow?

2. The price of gold on 11.11.2016 was Rs.3105.00 per gram. Will it remain the same on 18.11.2016?

Probabilistic

What this process is called?

- Suppose you are an employee of a cell phone company and assigned the task to examine how many cell phones are defective in a lot. Experiment Random Experiment, why?
- You go to a gambling den and play the game of cards
- Suppose as a CEO of an organization, you are conducting an interview to select candidates for three vacant posts out of 15 candidates who have attended the interview

Experiment

- Any planned process of data collection. It consists of a number of trials (replications) under the similar identical condition How does this data put in one place?

Random experiment

- An experiment in which results are unpredictable

Trial

- Each time a random experiment is conducted is a trial

Outcome

- The result of a trial is called outcome

Outcome

- Example
  - Defective or non-defective cell phone
  - Getting an ace with diamond mark

Exhaustive outcome

- All possible occurrence of outcomes in a random experiment

Example

- If two cell phones are checked, the possibility of getting a defective cell phone is (NN, DN, ND, DD) where N=Non-defective, D=Defective
- The exhaustive outcome is also called Sample space denoted by Ω or S. $\therefore \Omega = \{NN, ND, DN, DD\}$

Mutually exclusive outcome

- Occurrence of an outcome prevents or precludes the occurrence of another outcome
- Example
  - Suppose a cell phone is selected at random, if it happens to be non-defective, then it prevents the occurrence of the selected cell phone being defective or vice-versa

Equally likely outcome

- If every outcome of a random experiment has equal chance of being occurred is called equally likely outcome
- Example
    - Suppose a cell phone is selected at random, occurrence of defective or non-defective cell phone is equally likely outcome.

Note that all random experiment do no yield equally likely outocmes.

Give an example yielding non-equally likely outcomes

Independent outcome

- Each outcome of a random experiment will be independent of each other

Example

- Occurrence of defectiveness or non-defectiveness in a cell phone is independent of each other. No one influence the occurrence of other.

Event

- An event is an outcome of an experiment usually denoted by a capital letter.
- Example
  - Suppose a cell phone is selected at random, if it happens to be defective, then it is called an event.

Favourable event

- An event in which one is interested in a random experiment.
- Example
  - Suppose in an accident, there are two possible outcomes viz., Survival or Death. If one is interested in Survival, then it will be a favourable event to that person.

Definition of Probability

- In a random experiment, out of ‘n’ exhaustive, mutually exclusive, equally likely, independent outcomes if ‘m’ of them are favorable to the occurrence of an event, say, ‘A’, then the probability of an event ‘A’, denoted by P(A) is

$$0 \leq \mathrm{P(A)} = \frac{\text{Events favourable to A}}{\text{Exhuastive outcomes}} = \frac{m}{n} \leq 1$$

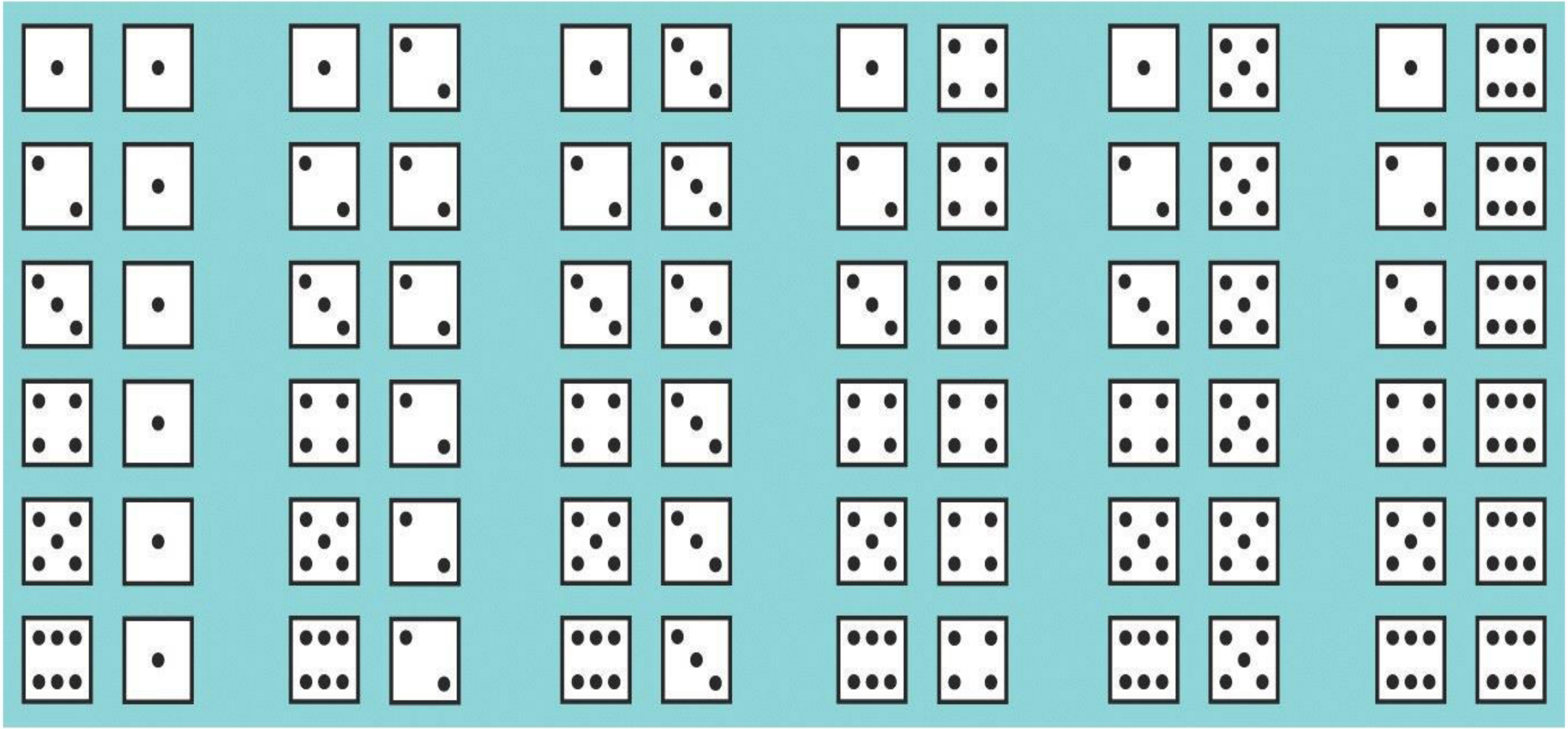
Example 1

- If three computer chips are selected at random, write the sample space $\Omega = \{\text{NNN}, \text{NND}, \text{NDN}, \text{DNN}, \text{NDD}, \text{DDN}, \text{DND}, \text{DDD}\}$

Example 2

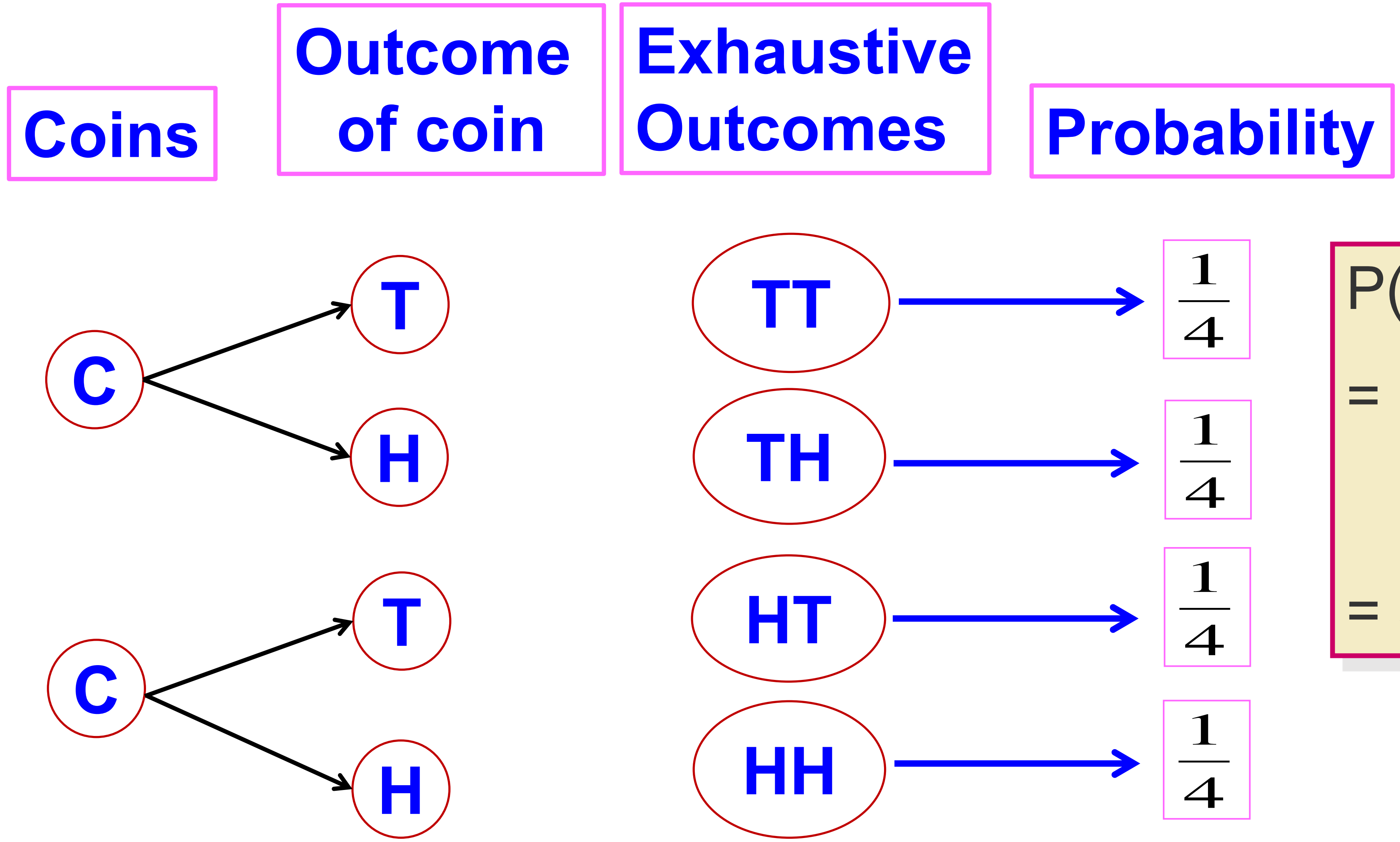
- If a candidate appear for an interview for a job, there are two possible outcomes viz., "selected" or "not-selected". If 'n' candidates appear for the interview, how many possible exhaustive outcomes are there? 2^n

Example 3: Write the sample space of throwing a pair of dice



1st Die	2nd Die					
	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

Example 3: A fair coin is tossed twice. What is the probability of getting at least one head?



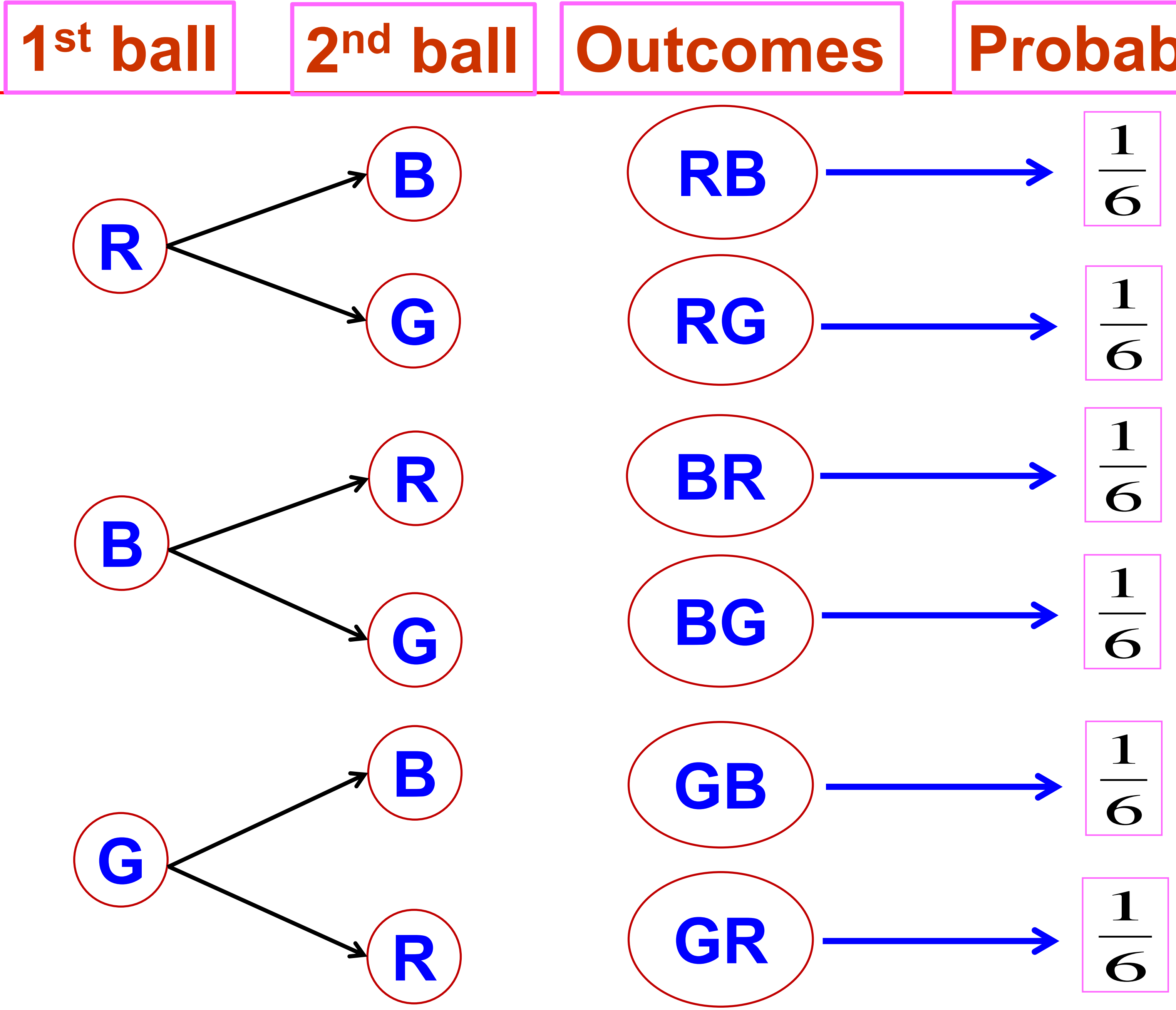
P(at least 1 head)

= P(TH) + P(HT)+ P(HH)

= 3/4

Example 4

- A bowl contains three balls: one red, one blue and one green. A child selects two balls at random. What is the probability that at least one red ball?



P(at least 1 red)

= P(RB) + P(BR)+

P(RG) + P(GR)

= 4/6

= 2/3

Complement of an event

If A is an occurrence of an event with probability $P(A)$, its non-occurrence is denoted by A^c with probability $P(A^c)$, then $P(A) + P(A^c) = 1$, so that $P(A^c) = 1 – P(A)$.

Example

- **Suppose a cell phone is selected at random, if A is the probability of defective cell phone, then its probability is given by $P(A) = ½$ and hence probability of non-defective cell phone is $P(A^c) = 1\text{-}P(A) = 1 - ½ = ½$**

Types of probability

There are three types of Probability

1. Theoretical probability: For theoretical reasons, we assume that all *n* possible outcomes of a particular experiment are equally likely, and we assign a probability of to each possible outcome. Example: The theoretical probability of rolling a 3 on a regular 6 sided die is 1/6.

Types of probability

2. Relative frequency interpretation of probability: Assuming that an experiment can be repeated many times and assuming that there are one or more outcomes that can result from each repetition. Then, the probability of a given outcome is the number of times that outcome occurs divided by the total number of repetitions.

$$\text{Probability of an event} = \frac{\text{How many times an event occurs}}{\text{How many trials}}$$

Example: A die is rolled 100 times. The number 3 is rolled 12 times. The relative frequency of rolling a 3 is 12/100.

Example: *Probability of Male versus Female Births*

Long-run relative frequency of males born in Karnataka is about 0.512 (512 boys born per 1000 births)

Table provides results of simulation: The proportion is far from 0.512 over the first few weeks but in the *long run* settles down around 0.512.

Relative Frequency of Male Births over Time

Weeks of Watching	Total Births	Total Boys	Proportion of Boys
1	30	19	.633
4	116	68	.586
13	317	172	.543
26	623	383	.615
39	919	483	.526
52	1237	639	.517

Types of probability

3. Personal or subjective probability: It is an estimate that reflects a person's opinion, or best guess about whether an outcome will occur. These are values (between 0 and 1 or 0 and 100%) assigned by individuals based on how likely they think events are to occur.

Example: The probability of candidate winning in an election is based on opinion poll is 60%.

The *mn* Rule

- If an experiment is performed in two stages, with ***m*** ways to accomplish the first stage and ***n*** ways to accomplish the second stage, then there are ***mn*** ways to accomplish the experiment.
- This rule is easily extended to ***k*** stages, with the number of ways equal to

$$n_1 \, n_2 \, n_3 \ldots n_k$$

Example: Toss two coins. The total number of simple events is: $2 \times 2 = 4$

Examples

Example: Toss three coins. The total number of simple events is: $2 \times 2 \times 2 = 8$

Example: Toss two dice. The total number of simple events is: $6 \times 6 = 36$

Example: Toss three dice. The total number of simple eve $6 \times 6 \times 6 = 216$

Permutations

- The number of ways you can arrange ***n*** distinct objects, taking them ***r*** at a time is

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

$$\text{where } n! = n(n-1)(n-2)\ldots(2)(1) \text{ and } 0! \equiv 1.$$

Example: How many 3-digit lock combinations can we make from the numbers 1, 2, 3, and 4?

The order of the choice is important!

$$^{4}P_{3} = \frac{4!}{1!} = 4(3)(2) = 24$$

Examples

Example: A lock consists of five parts and can be assembled in any order. A quality control engineer wants to test each order for efficiency of assembly. How many orders are there?

The order of the choice is important!

$$^{5}P_{5} = \frac{5!}{0!} = 5(4)(3)(2)(1) = 120$$

Combinations

- The number of distinct combinations of ***n*** distinct objects that can be formed, taking them ***r*** at a time is

$$ {}^{n}C_{r} = \frac{n!}{r!(n-r)!} $$

Example: Three members of a 5-person committee must be chosen to form a subcommittee. How many different subcommittees could be formed?

The order of the choice is not important!

$$ {}^{5}C_{3} = \frac{5!}{3!(5-3)!} = \frac{5(4)(3)(2)1}{3(2)(1)(2)1} = \frac{5(4)}{(2)1} = 10 $$

Example

- A box contains six balls, four red and two green. A child selects two balls at random. What is the probability that exactly one is red?

The order of the choice is not important!

$$^6C_2 = \frac{6!}{2!4!} = \frac{6(5)}{2(1)} = 15$$

ways to choose 2 balls.

$$^2C_1 = \frac{2!}{1!1!} = 2$$

ways to choose 1 green ball.

$$^4C_1 = \frac{4!}{1!3!} = 4$$

ways to choose 1 red ball.

4×2 =8 ways to choose 1 red and 1 green ball.

P(exactly one red) = 8/15

Example

1. Three light bulbs are selected at random from 15 bulbs of which 5 are defectives. Find the probability that

(i) None is defective

(ii) Exactly one is defective

(iii) At least one is defective

(iv) At most one is defective

Solution:

(i) - Exhaustive number of cases

$^{15}c_3$ = 455 ways

- Favourable cases

$^{10}c_3$ = 120 ways

Let A_1 be an event that none of the bulb chosen is defective, then $P(A_1) = 120/455 = 0.26$

Solution

(ii) - Exhaustive number of cases

$^{15}c_3 = 455$ ways

- Favourable cases

$^{5}c_1 \times {}^{10}c_2 = 5 \times 45 = 225$ ways

Let A_2 be an event that exactly one bulb chosen is defective, then $P(A_2) = 255/455 = 0.49$

(iii) Let A_3 be an event that at least one bulb chosen is defective, then

$$P(A_3) = 1 - P(A_1) = 1 - 0.26 = 0.74$$

(iv) Let A_4 be an event that at most one bulb chosen is defective, then

$$P(A_4)=P(A_1)+P(A_2)=0.26+0.49=0.75$$

Law of probability

Addition law for two events:

- If A and B are any two events with probabilities P(A) and P(B), then the occurrence of either A or B is

$$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Where

- $P(A \cup B)$ is the probability of occurrence of either the event A or the event B or both,

- $P(A \cap B)$ is the probability of occurrence of the event A and the event B

Addition law for more than two events:

- If $A_1, A_2, \ldots A_n$ are **any n events** with probabilities $P(A_1), P(A_2), \ldots, P(A_n)$, then the occurrence of either $A_1, A_2, \ldots, A_n$ is

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) - P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

Addition law for two events:

- If A and B are **two mutually exclusive events** with probabilities P(A) and P(B), then the occurrence of either A or B is

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

Addition law for more than two events:

- If A_1, A_2, …A_n are **n mutually exclusive events** with probabilities $P(A_1)$, $P(A_2)$, …,$P(A_n)$, then the occurrence of either A_1, A_2, …,A_n is

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$$

Example: Additive law

Suppose that there were 120 students in the classroom, and that they could be classified as follows:

	Brown hair	Not Brown hair
Male	20	40
Female	30	30

A: brown hair
$P(A) = 50/120$
B: female
$P(B) = 60/120$

$$
\begin{aligned}
P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\
&= 50/120 + 60/120 - 30/120 \\
&= 80/120 = 2/3
\end{aligned}
$$

When two events A and B are mutually exclusive,

$P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

	Brown hair	Not Brown hair
Male	20	40
Female	30	30

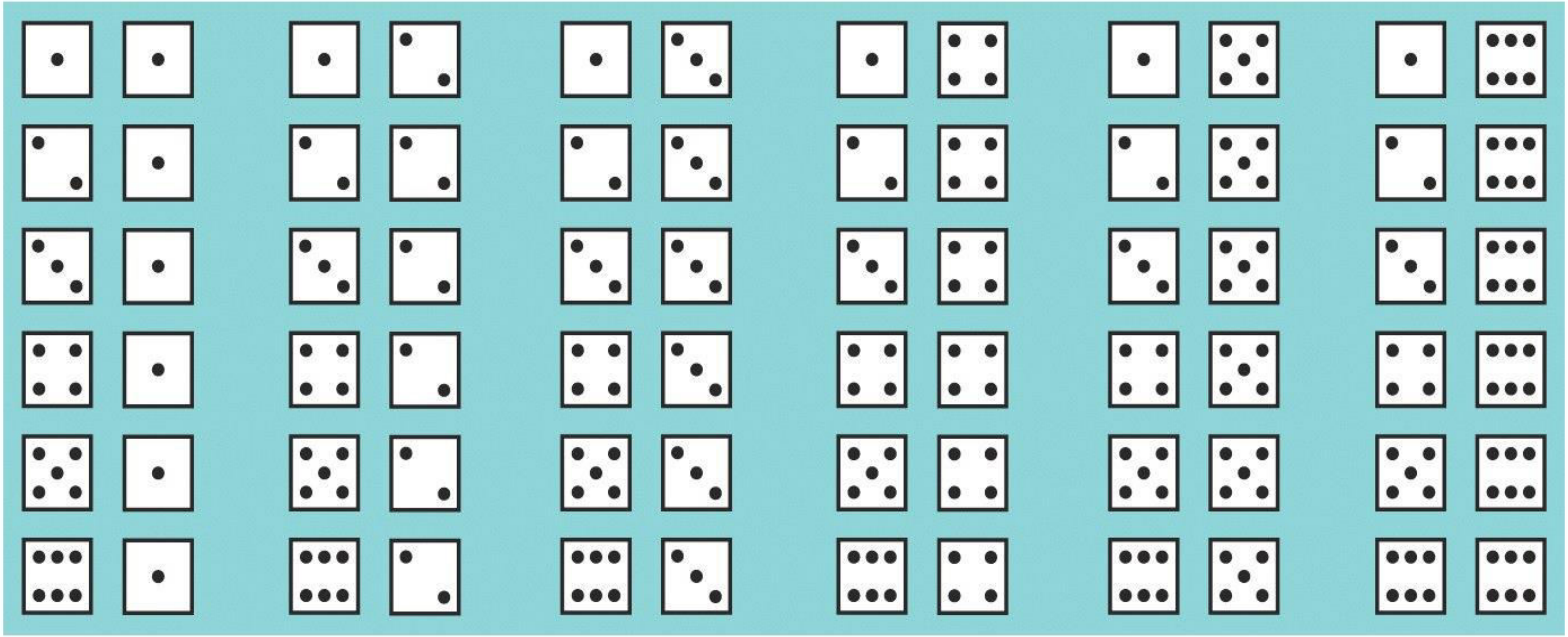
A: male with brown hair
$P(A) = 20/120$
B: female with brown hair
$P(B) = 30/120$

A and B are mutually exclusive, so that

$$P(A \cup B) = P(A) + P(B)$$
$$= 20/120 + 30/120$$
$$= 50/120$$

Example

- **Two unbiased dice are thrown. What is the probability of getting sum of their outcome as 7?**



Multiplication law for two events:

- If A and B are **two independent events** with probabilities P(A) and P(B), then the occurrence of A and B is

$$P(A \text{ and } B) = P(A \cap B) = P(A) \times P(B)$$

Multiplication law for more than two events:

- If $A_1, A_2, \dots A_n$ are **n independent events** with probabilities $P(A_1), P(A_2), \dots, P(A_n)$, then the occurrence of A_1 and A_2 and …and A_n is

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1) \times P(A_2) \times \dots \times P(A_n)$$

Probability Rule 1

- Any probability is a number between 0 and 1.
- A probability can be interpreted as the proportion of times that a certain event can be expected to occur.
- If the probability of an event is more than 1, then it will occur more than 100% of the time (Impossible!).

Probability Rule 2

- All possible outcomes together must have probability 1.
- Because some outcome must occur on every trial, the sum of the probabilities for all possible outcomes must be exactly one.
- If the sum of all of the probabilities is less than one or greater than one, then the resulting probability model will be incoherent.

Probability Rule 3

If two events have no outcomes in common, they are said to be ***disjoint***. The probability that one *or* the other of two *disjoint* events occurs is the *sum* of their individual probabilities.

- Age of woman at first child birth

  under 20: 25%

  20-24: 33%

  24 or younger: 58%

  25+: ? Rule 3 (or 2): 42%

Probability Rule 4

- The probability that an event does not occur is 1 minus the probability that the event does occur.
- As a jury member, you assess the probability that the defendant is guilty to be 0.80. Thus you must also believe the probability the defendant is not guilty is 0.20 in order to be coherent (consistent with yourself).
- If the probability that a flight will be on time is 0.70, then the probability it will be late is 0.30.

Probability Rules: Mathematical Notation

PROBABILITY RULES

Rule 1. The probability $P(A)$ of any event A satisfies $0 \leq P(A) \leq 1$.

Rule 2. If S is the sample space in a probability model, then $P(S) = 1$.

Rule 3. Two events A and B are **disjoint** if they have no outcomes in common and so can never occur together. If A and B are disjoint,

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

This is the **addition rule for disjoint events.**

Rule 4. For any event A,

$$P(A \text{ does not occur}) = 1 - P(A)$$

Marginal probabilities

Consider the example on Sex wise blood group distribution

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

What is the probability of a person selected randomly will have blood group A?

Marginal probabilities

Marginal probabilities appear on the "margins" of a probability table. It is probability of single outcome

Blood group	Male	Female	Total	Row probabilities
O	20	20	40	40/100=0.40
A	17	18	35	35/100=0.35
B	8	7	15	15/100=0.15
AB	5	5	10	10/100=0.10
Total	50	50	100	1
Column probabilities	50/100=0.50	50/100=0.5	1	

What is the probability that a person selected has blood group B given that he is male?

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

Probability to be computed

Given information

8/50 = 0.16 ???

How ???

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

$$\therefore P(M) = \frac{50}{100}$$

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

???

What is the probability that a person selected is male given that his blood group is B?

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

Given information

8/15 = 0.53 ???

Probability to be computed

How ???

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

$$\therefore P(M) = \frac{15}{100}$$

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

???

Conditional probabilities

Conditional probabilities is the probability of occurrence of an event on condition that another event has already occurred

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

8/50 = 0.16 ???

What is the probability that a person selected has blood group B given that he is male?

8/15 = 0.53 ???

What is the probability that a person selected is male given that his blood group is B?

Conditional probabilities

Mathematical definition

- Let A and B be any two events. The conditional probability of the event A given that the event B has already occurred is

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \; P(B) > 0$$

where

- $P(A \cap B)$ = Probability of occurrence of the events A and B
- $P(A \mid B)$ = Probability of the event A given that the event B has already occurred (Conditional probability)
- $P(B)$ = Probability of the of event B and should always be greater than 0

Conditional probabilities

- **Note that**

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

and

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

$$\therefore P(B|A)=P(A|B) \Rightarrow P(A)P(B|A)=P(B)P(A|B)$$

Conditional probabilities

Blood group	Male	Female	Total
O	**20**	20	40
A	**17**	18	35
B	**8**	7	15
AB	**5**	5	10
Total	**50**	50	100

- Let M be an event that the person selected is male

$$\therefore P(M) = \frac{50}{100}$$

- Let B be an event that the blood group is B

Conditional probabilities

Blood group	Male	Female	Total
O	20	20	40
A	17	18	35
B	8	7	15
AB	5	5	10
Total	50	50	100

$$\therefore P(B \mid M) = \frac{P(M \cap B)}{P(M)}$$

$$= \frac{\frac{8}{100}}{\frac{50}{100}} = \frac{8}{50} = \frac{4}{25} = 0.16$$

- The probability of combined occurrence M and B is

$$P(M \cap B) = \frac{8}{100}$$

Independent Event

If A and B are two independent events, then

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

- **We have**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B)P(A \mid B)$$

- **If A and B are independent then the conditional probability P(A|B) will become unconditional probability P(A), that is**

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

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(B)P(A \mid B) = P(B)P(A)$$

What this process is called?

- Suppose a card is drawn from a well shuffled deck of cards at random and the event is noted down, then the second card is drawn. Find the probability that

(i) the first card drawn is an ace given that the second card drawn is a diamond.

(ii) the first card drawn is an ace and replaced back to the deck then the second card is drawn which is a diamond

Example:

A manufacturer of airplane parts knows from the past experience that the probability is 0.80 that an order will be ready for shipment on time, and it is 0.72 that an order will be ready for shipment and will also be delivered on time. What is the probability that such an order will be delivered on time given that it was ready for shipment on time?

Solution:

Let R be the event that an order is ready for shipment on time and D be an event that it is delivered on time.

$\therefore\ P(R) = 0.80$ and $P(R \cap D) = 0.72$

$$\therefore P(D \mid R) = P(R \cap D)/P(R) = 0.72/0.80$$

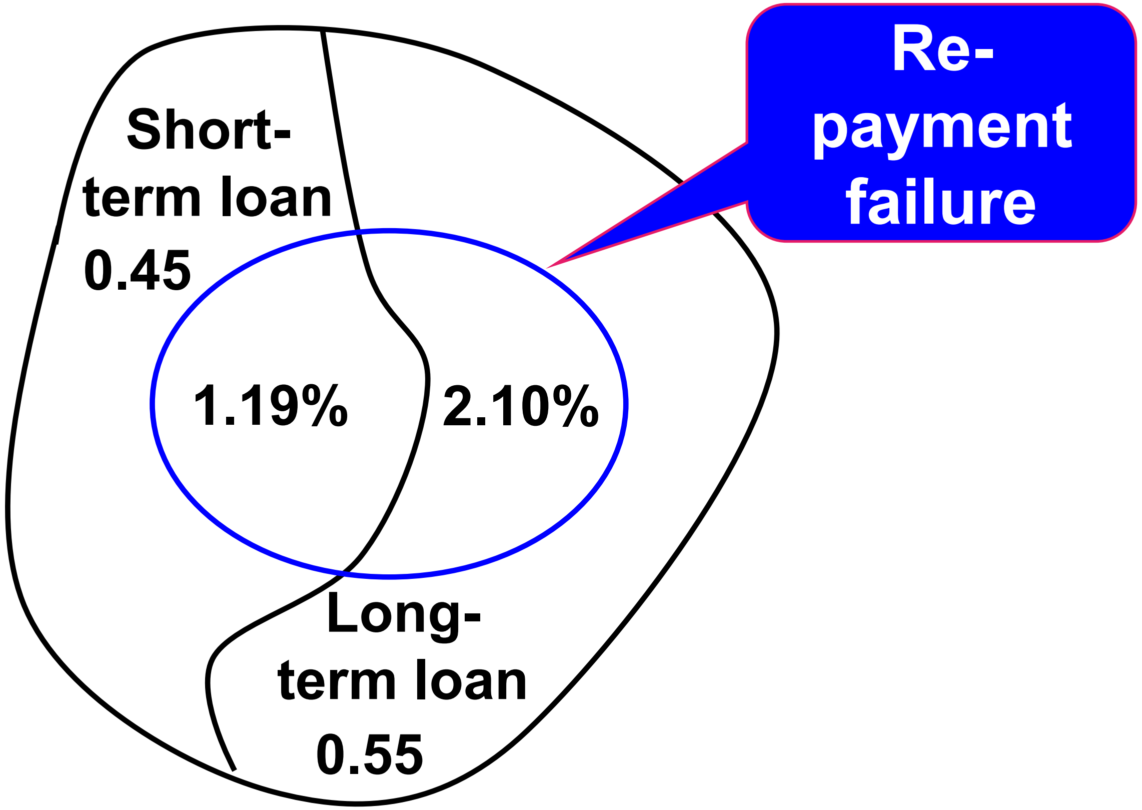
$$= 0.90$$

Thus, 90% of the shipments will be delivered on time provided they are shipped on time.

Note that $P(R|D)$, the probability that a shipment that is delivered on time was also ready for shipment on time, cannot be determined without further information; for this we would also have to know $P(D)$

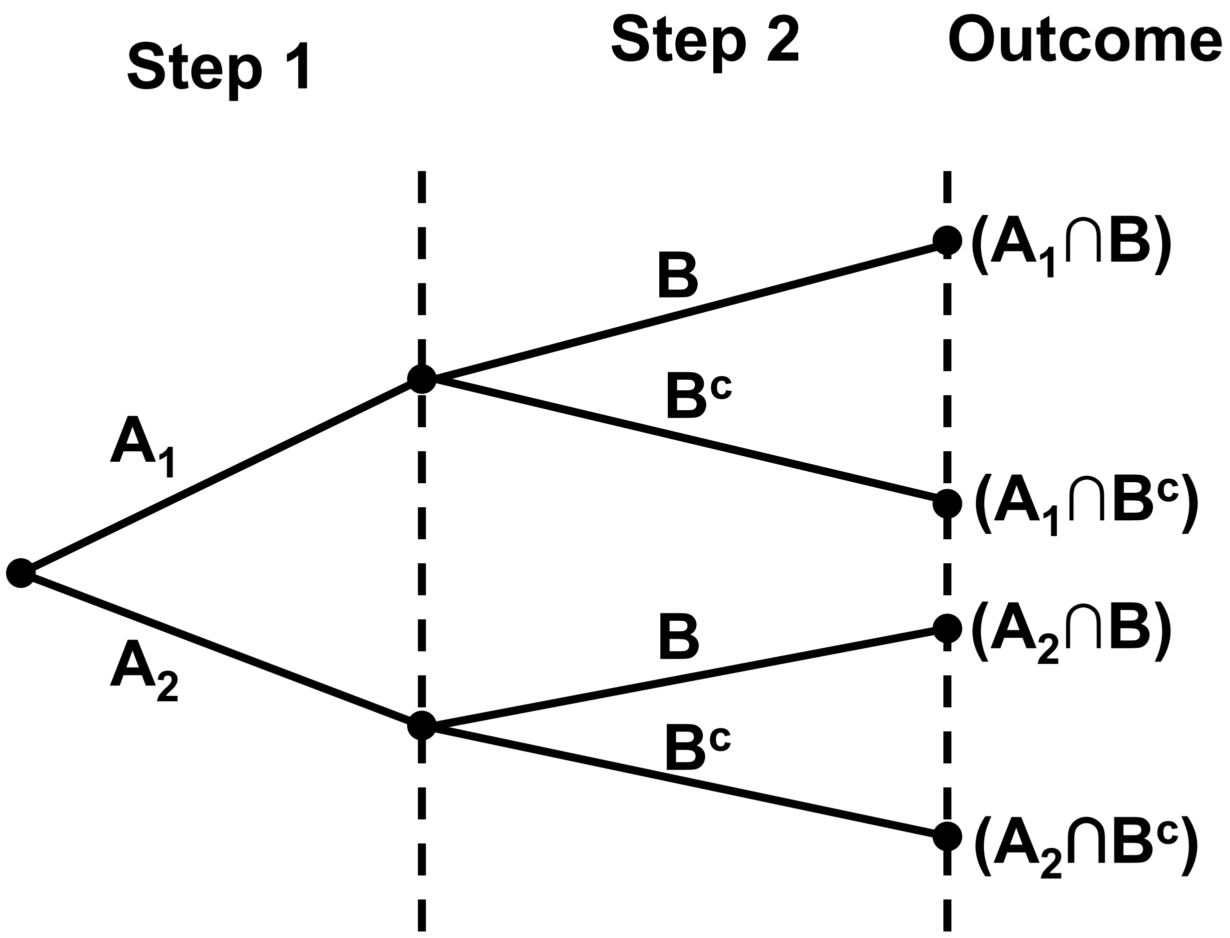
Example

A Nationalized bank sanctioned two types of loans, viz., Short-term loan and Long-term loan respectively in the proportion of 0.45 and 0.55. It was found that the recovery failure for short-term loan was 1.9% and for long-term loan 2.1%. Find the probability of the bank's re-payment failure rate.

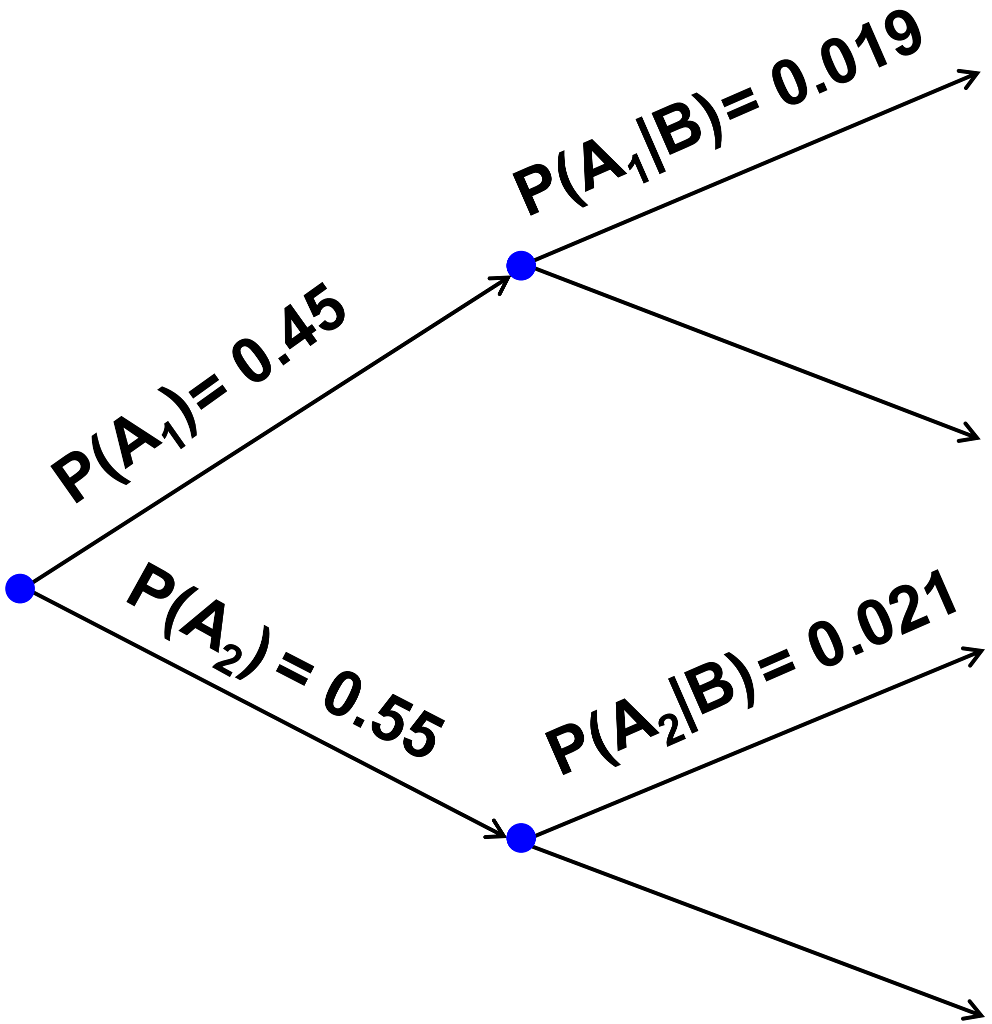


- Let A_1 be the event of sanctioning short-term loan
- Let A_2 be the event of sanctioning long-term loan
- Let B be the event of re-payment failure

Tree Diagram



Probability tree



$P(A_1)\ P(A_1|B) = P(A_1 \cap B) = 0.00855$

$P(A_2)\ P(A_2|B) = P(A_2 \cap B) = 0.00855$

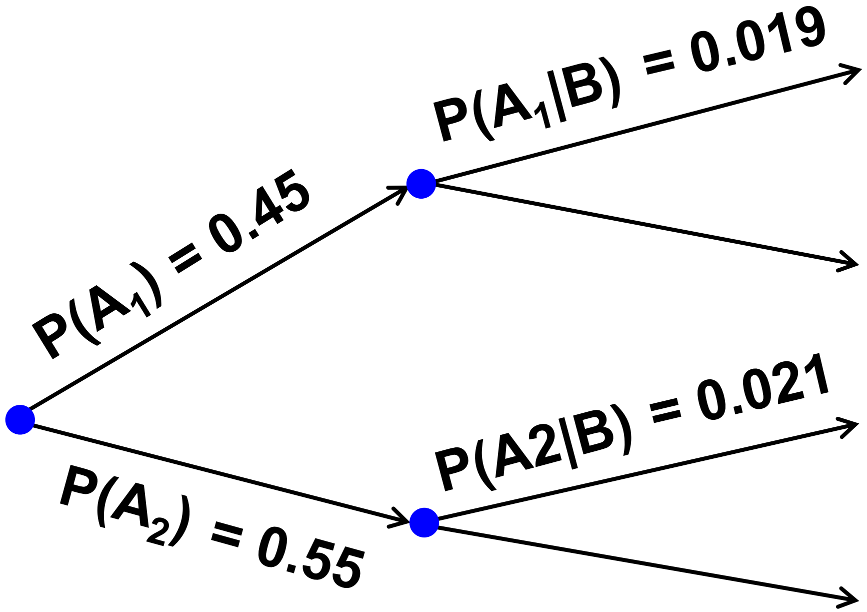
$P(B) = P(A_1)\ P(A_1|B) + P(A_2)\ P(A_2|B) = P(A_1 \cap B) + P(A_2 \cap B) = 0.0201$

Example

Suppose, if given the probability of the bank's re-payment failure rate, find the probability that a randomly selected customer's re-payment failure is from long-term loan.

By the definition of conditional probability, we have.

$$\therefore P(A_2 \mid B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{0.01155}{0.0201} = 0.5746$$



$P(A_1)\ P(A_1|B) = P(A_1 \cap B) = 0.008550$

$P(A_2)\ P(A_2|B) = P(A_2 \cap B) = 0.01155$

$P(B) = P(A_1)\ P(A_1|B) + P(A_2)\ P(A_2|B) = P(A_1 \cap B) + P(A_2 \cap B) = 0.0201$

$$\therefore P(A_2 \mid B) = \frac{P(A_2 \cap B)}{P(B)} = \frac{0.01155}{0.0201} = 0.5746$$

In Bayes' theorem

Bayes' theorem is most commonly used to estimate the state of a hidden, causal variable ***A*** based on the measured state of an observable variable ***B***

Bayes' Equation

Likelihood

Prior

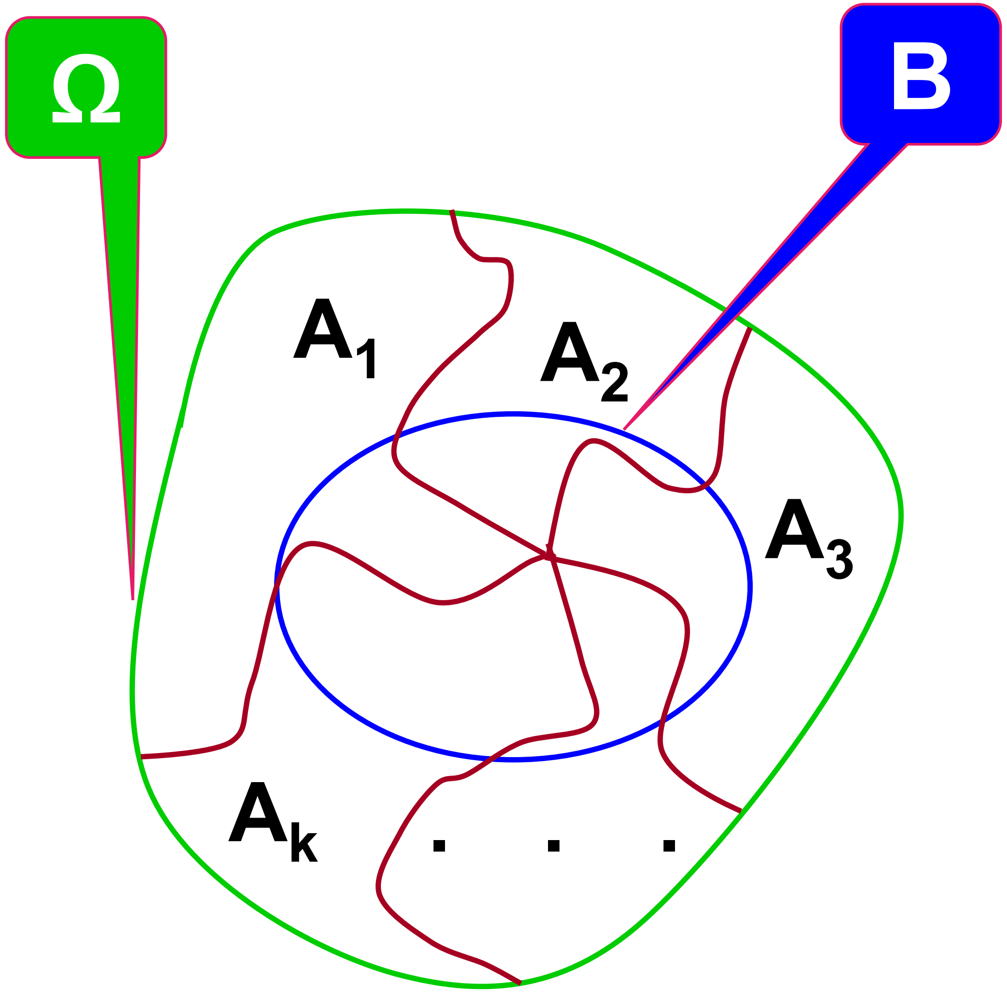
$$\therefore P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B)}$$

Posterior

Evidence

Example

- Whereas the posterior *P(A|B)* is often difficult to estimate directly, reasonable models of the likelihood *P(B|A)* can often be formed. This is typically because A is causal on B.
- Thus Bayes' theorem provides a means for estimating the posterior probability of the causal variable A based on observations B.



- Let Ω be a sample space which is partitioned into k disjoint (mutually exclusive) events $A_1, A_2, A_3, \ldots, A_k$ with respective probabilities $P(A_1), P(A_2), P(A_3), \ldots, P(A_k)$
- Let B be an event common to all k disjoint events.

Example

- **To calculate the evidence *P(B)* in Bayes' equation, we typically have to *marginalize* over all possible states of the causal variable *A***

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

$$\begin{aligned} P(B) &= P(A_1 \cap B) + P(A_2 \cap B) + \ldots + P(A_k \cap B) \\ &= P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \ldots + P(B \mid A_k)P(A_k) \end{aligned}$$

Bayes' Rule - Updating Probabilities

- Let $A_1, \ldots, A_k$ be a set of events which has partitioned the sample space into k mutually exclusive events such that:
  - each set has known $P(A_i) > 0$ (each event can occur)
  - for any 2 sets A_i and A_j, $P(A_i \cap A_j) = P(\phi) = 0$ for $i \neq j$ (events are disjoint)
  - $P(A_1) + \ldots + P(A_k) = 1$ (each outcome belongs to one of events)

Bayes' Rule - Updating Probabilities

- If B is an event such that
  - $0 < P(B) < 1$ (B can occur, but will not necessarily occur)
  - We know the probability will occur given each event A_i:

    $P(B|A_i)$

- Then we can compute probability of A_i given B occurred:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B \mid A_1)P(A_1) + \Lambda + P(B \mid A_k)P(A_k)} = \frac{P(A_i \cap B)}{P(B)}$$

Quality levels differ between suppliers

A product is supplied by two different suppliers. The product is classified as good or bad. 65% of the product was supplied by Supplier 1 and rest by Supplier 2. The details are given in the following table.

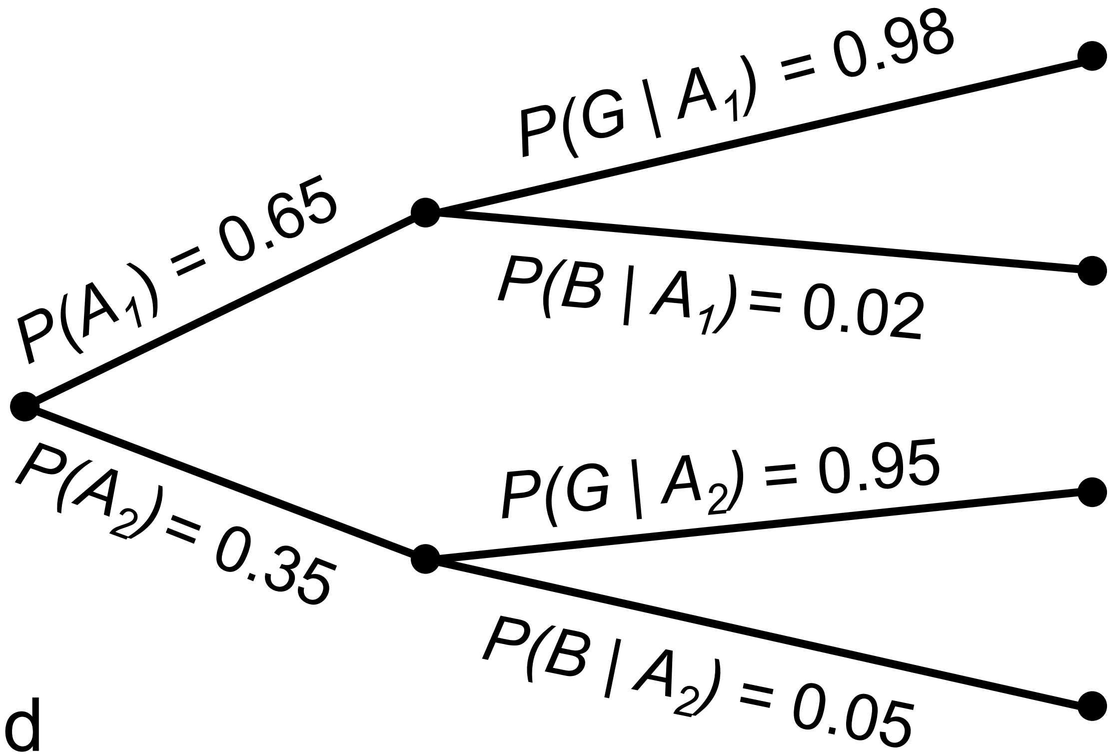
Supplier	Percentage Good Parts	Percentage Bad Parts
1	98	2
2	95	5

Each of the outcomes is the intersection of 2 events. For example, the probability of selecting a part from supplier 1 that is good is given by:

$$P(A_1, G) = P(A_1 \cap G) = P(A_1)P(G \mid A_1)$$

Probability Tree for Two-Supplier Example

Step 1 Suppliers	Step 2 Condition	Probability of Outcome



$$P(A_1 \cap G) = P(A_1)P(G \mid A_1) = 0.6370$$

$$P(A_1 \cap B) = P(A_1)P(B \mid A_1) = 0.0130$$

$$P(A_2 \cap G) = P(A_2)P(G \mid A_2) = 0.3330$$

$$P(A_2 \cap B) = P(A_2)P(G \mid A_2) = 0.01750$$

Find

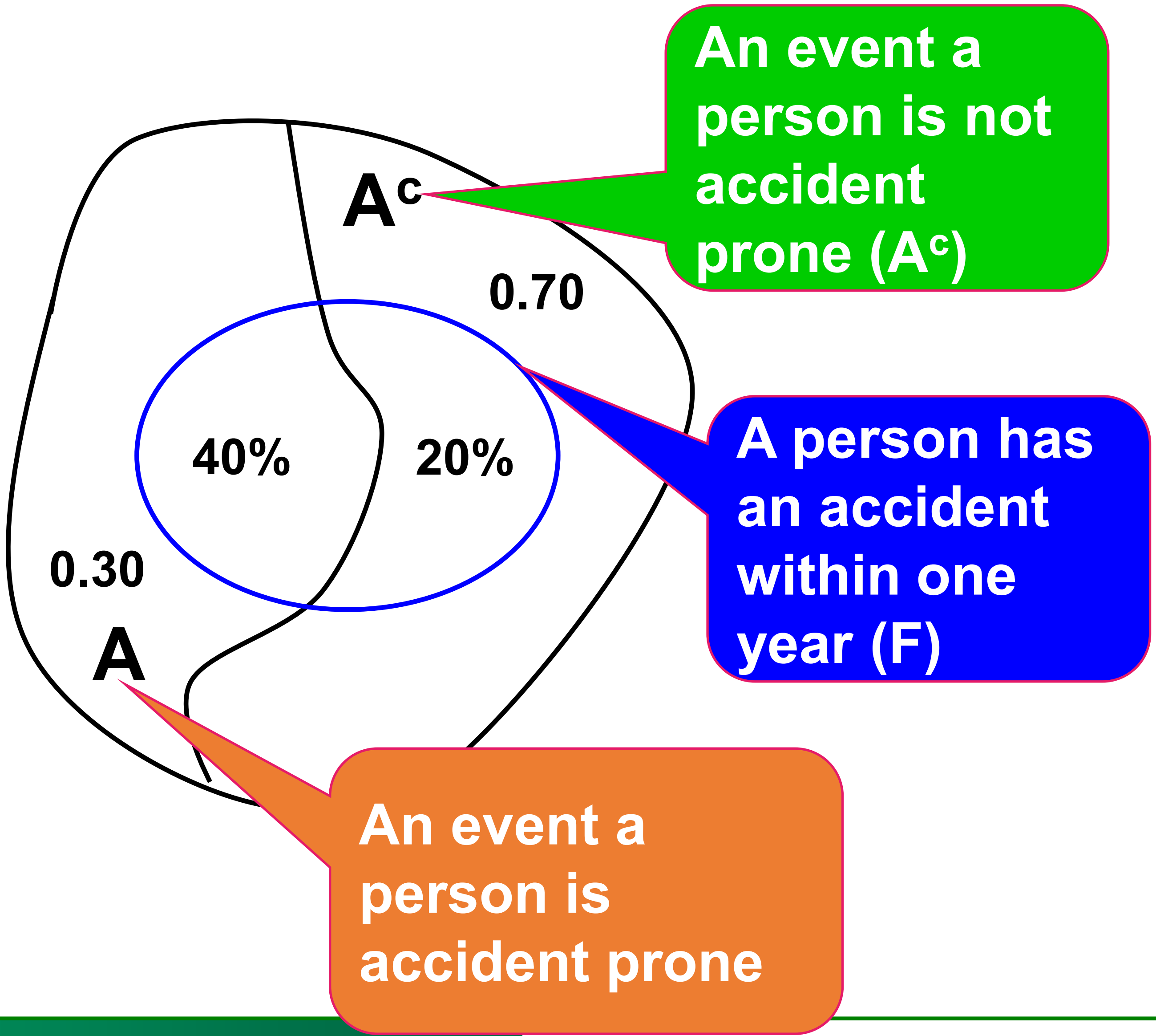
$$P(A_1 \mid G) = \frac{P(A_1)P(G \mid A_1)}{P(G)} = ???$$

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{P(B)} = ???$$

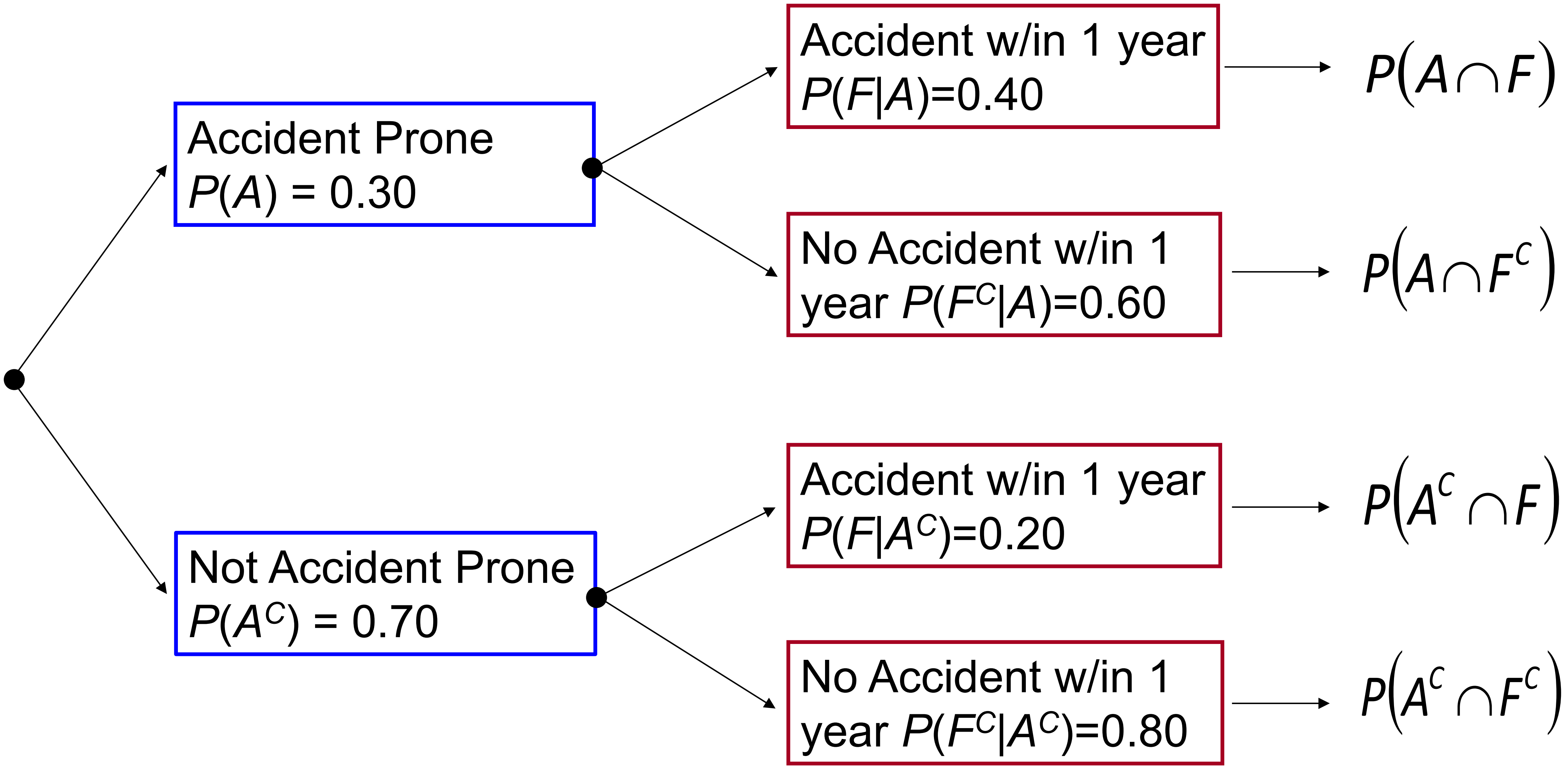
Example

An insurance company divides its clients into two categories: those who are accident prone and who are not. Statistics show there is a 40% chance an accident prone person will have an accident within 1 year whereas a 20% chance non-accident prone people will have an accident within the first year.

If 30% of the population is accident prone, what is the probability that a new policyholder has an accident within 1 year?



- Let A be the event a person is accident prone
- Let A^c be the event a person is not accident prone
- Let F be the event a person has an accident within 1 year



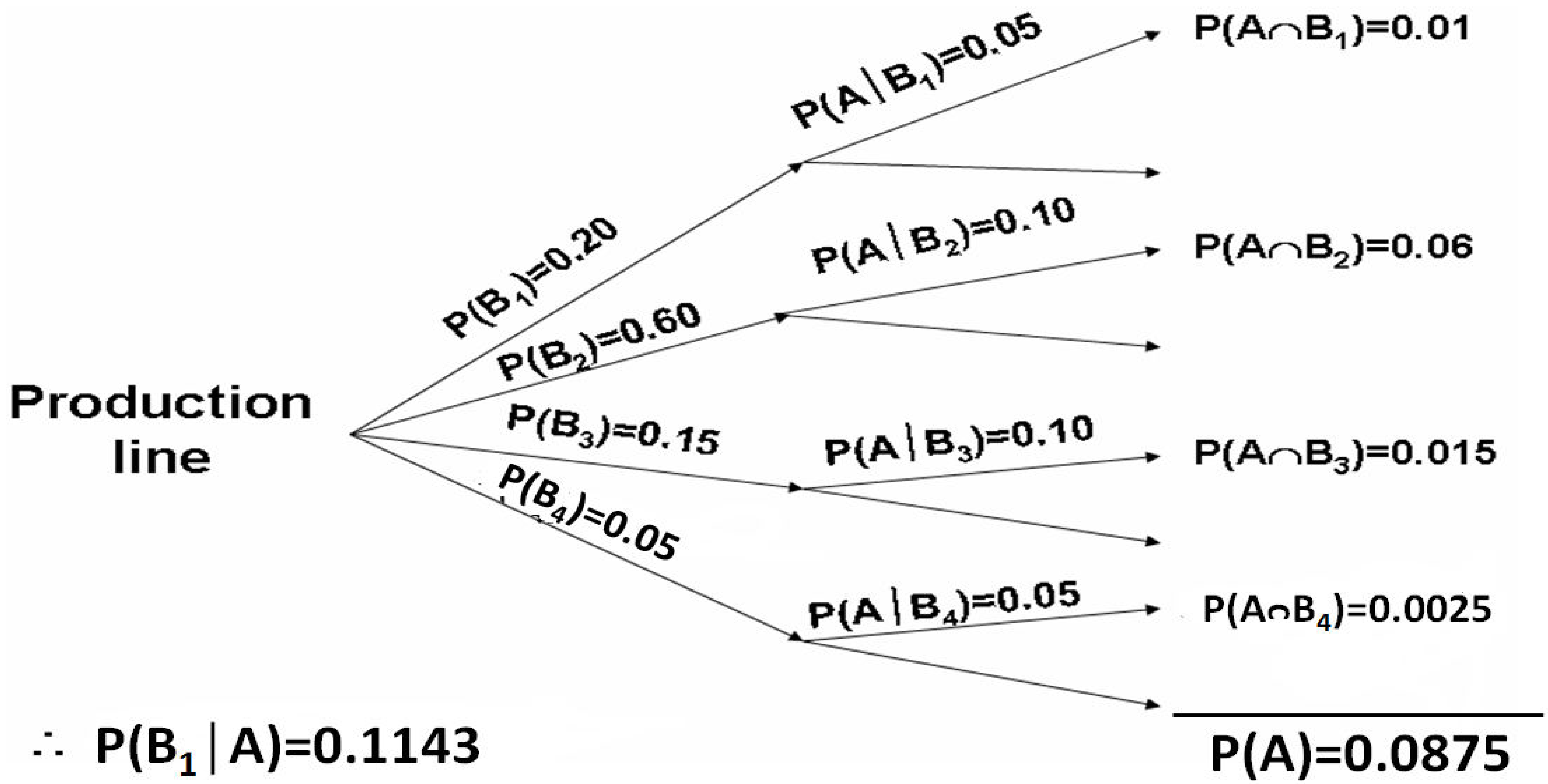
$$P(F) = P(A \cap F) + P(A^c \cap F) = 0.26\ ???$$

Example

- Four technicians regularly make repairs when breakdowns occur on an automated production line.
- Janet, who services 20% of the breakdowns, makes an incomplete repair 1 time in 20.
- Tom, who services 60% of the breakdowns makes an incomplete repair 1 time in 10
- Georgia, who services 15% of the breakdowns, makes an incomplete repair 1 time in 10
- Peter, who services 5% of the breakdowns, makes an incomplete repair 1 time in 20.
- For the next problem with the production line diagnosed as being due to an initial repair that was incomplete, what is the probability that this initial repair was made by Janet?.

Solution

- Let
- A: be the event that the initial repair was incomplete
- B_1: an event that the initial repair was made by Janet
- B_2: an event that the initial repair was made by Tom
- B_3: an event that the initial repair was made by Georgia
- B_4: an event that the initial repair was made by Peter.



$\therefore\ P(B_1 \mid A) = 0.1143$

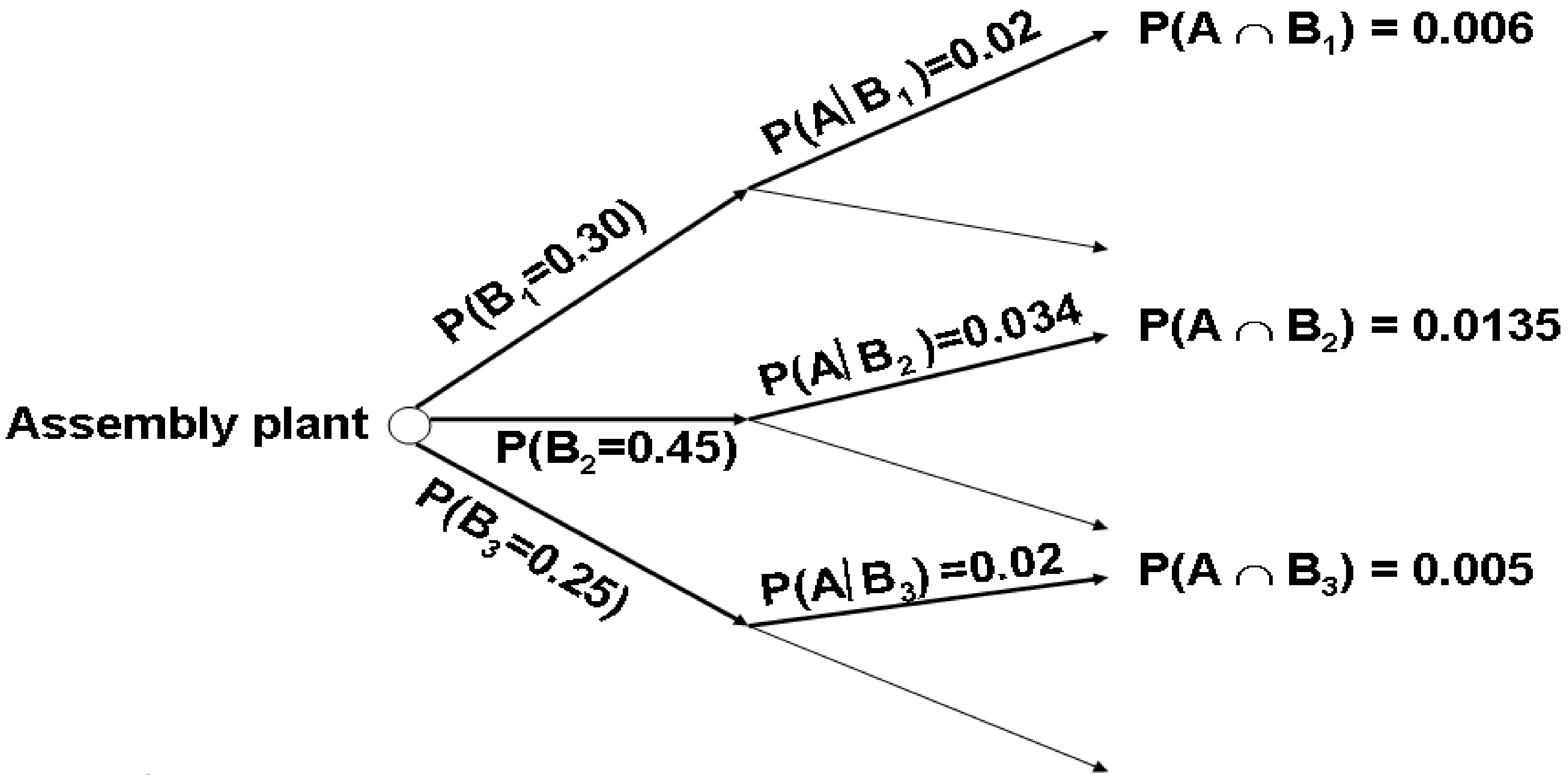
- It is of interest to note that although Janet makes an incomplete repairs only 1 out of times, viz., 5% of the breakdowns she services, more than 11% of the incomplete repairs are her responsibility.

Example

- In a certain assembly plant, three machines, B_1, B_2, and B_3 make 30%, 45%, and 25% respectively of the products. It is known from the past experience that 2%, 3% and 2% of the products made by each machine respectively, are defective.
- (i) Suppose a finished product is randomly selected, what is the probability that it is defective?
- (ii) If a product chosen randomly is found defective, what is the probability that it was made by machine B_3?

Solution

- Let
- A: event that the product is defective
- B_1: an event that product made by machine B_1
- B_2: an event that product made by machine B_2
- B_3: an event that product made by machine B_3



$P(B_3 \mid A) = 0.005/0.0245 = 0.204$

Probability Distributions

Example

- Three RAM chips were examined and detected them as non-defective (N) or defective (D). The possible outcome of are

$$\Omega = \{NNN, NND, NDN, DNN, NDD, DDN, DND, DDD\}$$

- **Define**

  X = No. of defective RAM chips

Based on the Definition
X = No. of defective RAM chips

Outcome	X	Frequency (f)	Probability
NNN	0	1	1/8
NND	1	3	3/8
NDN	1		
DNN	1		
NDD	2	3	3/8
DND	2		
DDN	2		
DDD	3	1	1/8

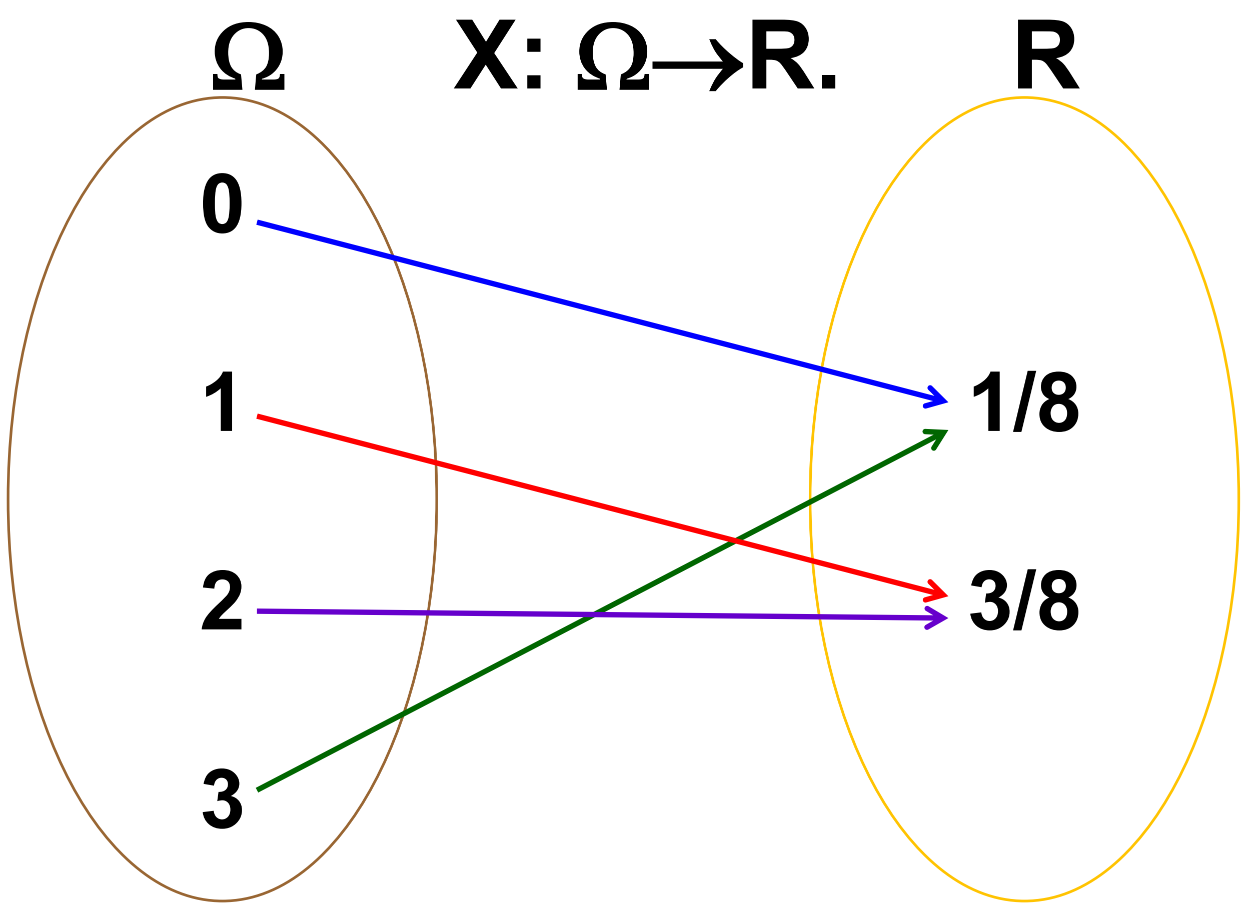
Comparison between frequency and probability distributions

Age distribution

Age (yrs)	No. of Persons
0-1	141
1-5	187
6-14	206
15-24	353
25-34	365
35-44	386
45-54	269
≥55	93
Total	2000

Probability distribution

No. of RAM chips defective	Probability
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$
Total	1


Ω
X: Ω→R.
R
0
1
2
3
1/8
3/8

- With the introduction of X we can write the probabilities as

- $p(0) = P(X=0) = 1/8$
- $p(1) = P(X=1) = 3/8$
- $p(2) = P(X=2) = 3/8$
- $p(3) = P(X=3) = 1/8$

which is such that $p(0) + p(1) + p(2) + p(3) = 1$ and each

$p(x) \geq 0, x = 0, 1, 2, 3$

- $p(0) = P(X=0) = 1/8$
- $p(1) = P(X=1) = 3/8$
- $p(2) = P(X=2) = 3/8$
- $p(3) = P(X=3) = 1/8$

- The values assumed by the random variable X above are discrete and hence, it is of the form $p(x)=P(X=x)$, which is called the probability mass function (pmf). Observe that
- (i) $p(0) + p(1) + p(2) + p(3) = 1$
- (ii) $p(x) \geq 0$, $x = 1, 2, 3, 4$

Random variable

- A random variable is a real valued function which is a mapping from the sample space Ω to the set of real numbers,
- i.e., $X: \Omega \rightarrow R$.
- In other words, it associates a real number with each element in the sample space
- There are two types of random variables

1. Discrete random variable
2. Continuous random variable

Discrete random variable

- A random variable is called discrete when it assumes discrete observations or whole numbers
- The probability that a random variable X assumes different discrete values x is denoted by $p(x)=P(X=x)$, called probability mass function (pmf).

Discrete Probability Distribution

- That is, p(x)=P(X=x) is called a (discrete) probability distribution, because, a total probability of one of a random experiment is divided into different events. Hence, the following are the properties of P(X=x)

$$1.\ 0 \leq p(x) \leq 1, \text{ for all } x$$

$$2.\ \sum_{\text{all } x_i} p(x) = 1$$

Discrete Cumulative Probability Distribution Function

- Let $p(x) = P(X=x)$ is called a (discrete) probability distribution.
- Let $F(x) = P(X \leq x)$. $F(x)$ is called the Distribution Function (DF) of the discrete random variable X. $F(x)$ has the following properties

$$1.\ 0 \leq F(x) \leq 1, \text{ for all } x$$

$$2.\ \sum_{\text{upto } x} P(X \leq x)$$

Developing Discrete Probability Distributions

Probability distributions can be **estimated** from relative frequencies. Consider the discrete (countable) number of televisions per household (X) from India survey data …

No. of televisions	No. of households	X	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
Total	101,501		1.000

1,218 ÷ 101,501 = 0.012

e.g. P(X=4) = P(4) = 0.076 =7.6%

What is the probability there is at least one television but no more than three in any given household?

No. of televisions	No. of households	X	P(x)
0	1,218	0	0.012
1	32,379	1	0.319
2	37,961	2	0.374
3	19,387	3	0.191
4	7,714	4	0.076
5	2,842	5	0.028
Total	101,501		1.000

"at least one television but no more than three"

$P(1 \leq X \leq 3) = P(1) + P(2) + P(3) = 0.319 + 0.374 + 0.191 = 0.884$

Problem 1

- Let X denote the number of tires on a randomly selected automobile that are underinflated.
- (a) Which of the following three probability mass functions is legitimate for X and why are the other two or not allowed?

X = x	0	1	2	3	4
$p_1(x) = P(X=x)$	0.30	0.20	0.10	0.05	0.05
$p_2(x) = P(X=x)$	0.40	0.10	0.10	0.10	0.30
$p_3(x) = P(X=x)$	0.40	0.10	0.20	0.10	0.30

Example 1

$X = x$	0	1	2	3	4
$p_2(x) = P(X=x)$	0.40	0.10	0.10	0.10	0.30

- (b) For legitimate pmf of part (a),

  compute (i) $P(X \le 2)$, (ii) $P(2 \le X \le 4)$ and (iii) $P(X \ne 0)$
- (c) If $p(x) = c(5-x)$, $x = 0, 1, 2, 3, 4$ what is the value of c?

Example 2

- A mail-order computer business has six telephone lines. Let X denote the number of lines in use at a specified time. Suppose that the pmf is as follows:

X = x	0	1	2	3	4	5	6
p(x) = P(X=x)	0.10	0.15	0.20	0.25	0.20	0.06	0.04

- Calculate the probability that
  - (a) At most 3 lines are in use
  - (b) Fewer than 3 line in use
  - (c) At least 3 lines are in use
  - (d) Between 2 and 5 lines are in use
  - (e) At least four lines are not in use

Example 3

- **Find the probability mass function from a given distribution of x**

$$F(x)=\begin{cases}0, & x<0\\ 0.06, & 0\le x<1\\ 0.19, & 1\le x<2\\ 0.39, & 2\le x<3\\ 0.67, & 3\le x<4\\ 0.92, & 4\le x<5\\ 0.97, & 5\le x<6\\ 1.00, & 6\le x\end{cases}$$

Example 4

- Let X be a discrete random variable having the probability mass function

X = x	0	1	2	3	4	5	6	7
p(x) = P(X=x)	0.01	0.03	0.13	0.25	k	0.17	0.10	0.02
# Registered	150	450	1950	3750	k	2550	1500	300

- Find the value of k
- Find the distribution function of X
- Calculate E(X) and V(X)

Define, X = No. of defective RAM chips

Assume that probability of getting a defective chip is 0.15

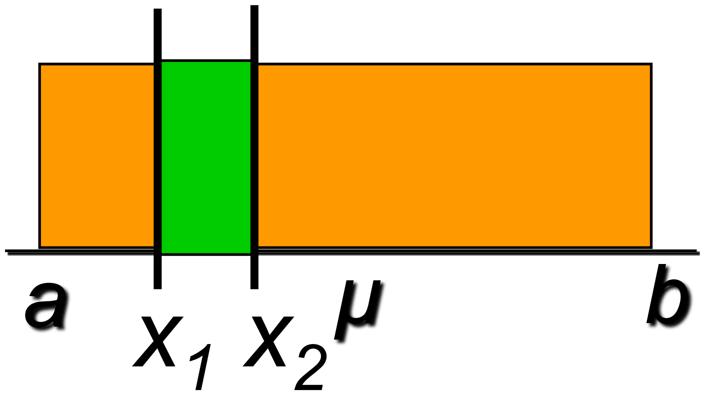
Outcome	Outcome	X	P(X=x)
NNN	$D^cD^cD^c$	0	P(X=0)=0.85*0.85*0.85 = 0.6141
NND	D^cD^cD	1	P(X=1)=0.85*0.85*0.15 = 0.1084
NDN	D^cDD^c	1	?
DNN	DD^cD^c	1	?
NDD	D^cDD	2	P(X=2)=0.85*0.15*0.15=0.0191
DND	DD^cD	2	?
DDN	DDD^c	2	?
DDD	DDD	3	P(X=3)=0.15*0.15*0.15=0.0034

- A continuous random variable can assume any value in an interval on the real line or in a collection of intervals.
- It is not possible to talk about the probability of the random variable assuming a particular value.
- Instead, we talk about the probability of the random variable assuming a value within a given interval.

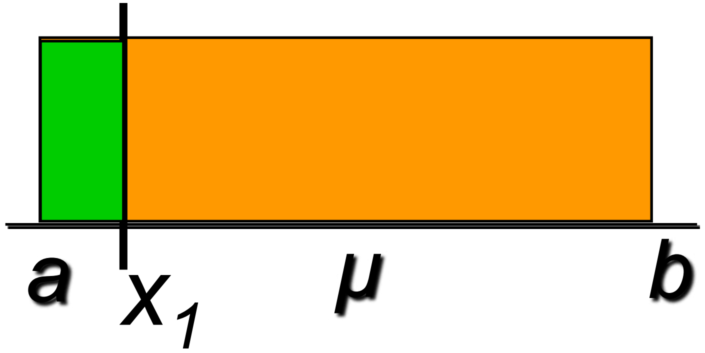
Continuous Probability Distributions

- The probability of the continuous random variable assuming a specific value is 0.

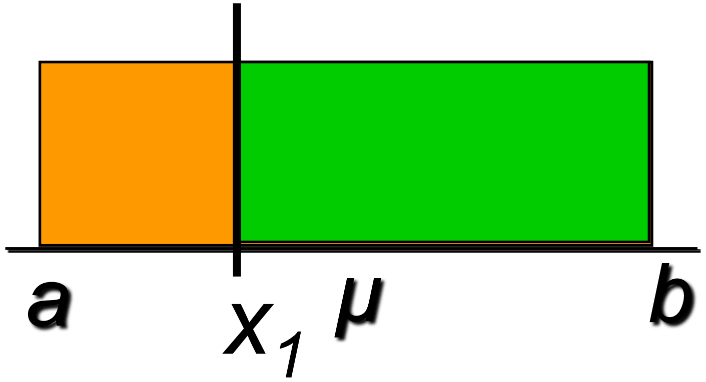
- The probability of the random variable assuming a value within some given interval from x_1 to x_2 is defined to be the <u>area under the graph</u> of the <u>probability density function</u> between x_1 and x_2.



$$P(X_1 \leq X \leq x_2)$$



$$P(X \leq x_1)$$



$$P(X \geq x_1) = 1 - P(X \leq x_1)$$

$$P(X \geq x_1)$$

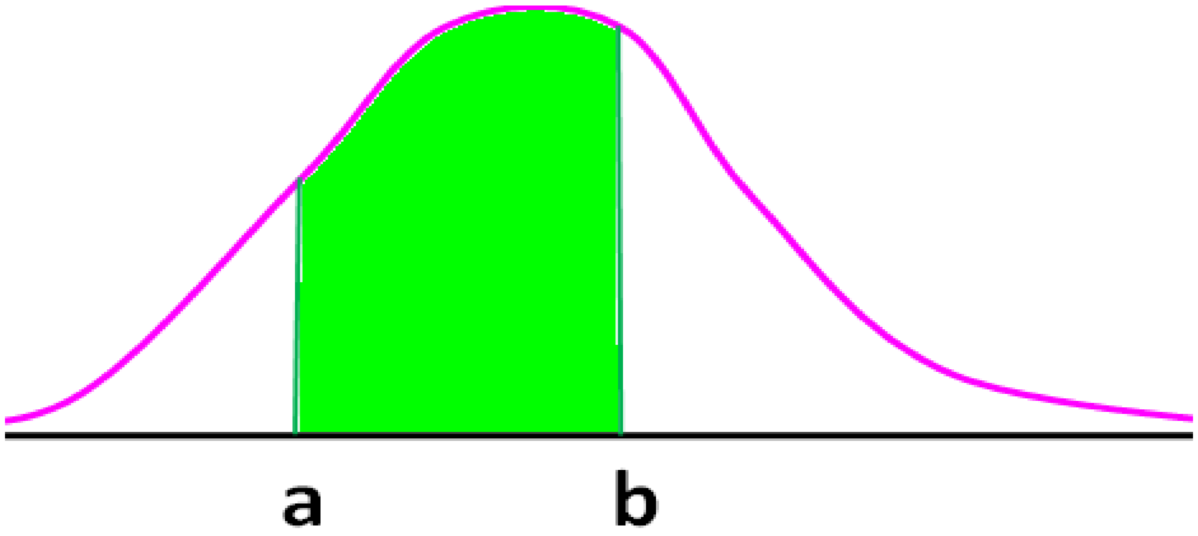
Continuous random variable

- A random variable is called continuous when it assumes values in a given interval.
- The probability that a random variable X assumes different values x in a given interval, say [a, b], is denoted by $f(x) = P(a \leq X \leq b)$, called probability density function (pdf).

Continuous random variable

- Indeed, the probabilities are the area under the curve in a given interval. Thus, a function with values f(x) defined over the set of all real numbers (a, b) is given by

$$\mathrm{f(x)} = P(a \le X \le b) = \int_a^b f\left(x\right)\mathrm{dx}$$



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

Continuous random variable

- It is important to note that f(c), the value of the pdf of X at a constant c does not give P(X=c) as in the discrete case and in continuous case probabilities are always associated with intervals and **P(X=c) = 0**, i.e.,

$$\mathrm{f(c)} = P(X = c) = P(c \le X \le c) = \int_c^c f(x)\,\mathrm{dx} = 0$$

Continuous random variable

- **A f(x) is called a probability density function of a continuous random variable if it satisfy the following conditions:**

(i) $f(x) \geq 0$

(ii) $\int_{a}^{b} f(x)\,\mathrm{dx} = 1$

(iii) $P(a \leq X \leq b) = \int_{a}^{b} f(x)\,\mathrm{dx} = \mathrm{F(b)} - \mathrm{F(a)}$

Continuous Probability Distribution Function

- Let f(x) is called a (continuous) probability distribution.
- Let $F(x) = P(X \leq x)$. F(x) is called the Distribution Function (DF) of the continuous random variable X. F(x) has the following properties

$$1.\ 0 \leq F(x) \leq 1, \text{ for all } x$$

$$2.\ F(x) = \int_a^b f(x)dx = F(b) - F(a)$$

- Let f(x) is called a (continuous) probability distribution.
- Let F(x) = P(X ≤ x). F(x) is called the Distribution Function (DF) of the continuous random variable X. F(x) has the following properties

$$1.\ 0 \le F(x) \le 1, \text{ for all } x$$

$$2.\ F(x) = \int_a^b f(x)dx = F(b) - F(a)$$

Probability Distribution → **Joint Probability Mass Function**

If X and Y are two discrete random variables, the probability distribution of their simultaneous occurrences can be represented by a function with values p(x, y) for each pair of values (x, y) within the range of X and Y. This probability distribution is called the joint probability mass function given by p(x, y) = P(X=x, Y=y).

Probability Distribution → **Joint Probability Mass Function**

The function p(x, y) gives the probability that the outcomes x and y occur simultaneously.

The p(x, y) is called joint probability mass function if it satisfies the following conditions:

(i) p(x, y) ≥ 0, for each pair of values (x, y) within its domain.

(ii) $\sum_{x}\sum_{y} p(x, y) = 1$

where the summation extends over all possible pair of values (x, y) within its domain.

Example 1: Two balls are selected at random from a bag containing three green, two blue and four red balls.

If X and Y are respectively the numbers of green and blue balls included among the two balls drawn from the bag, find the probabilities associated with all possible pairs of value of X and Y.

If X and Y are respectively the numbers of green and blue balls included among the two balls drawn from the bag, find the probabilities associated with all possible pairs of value of X and Y.

If X and Y are respectively the numbers of green and blue balls included among the two balls drawn from the bag, find the probabilities associated with all possible pairs of value of X and Y.

Solution: Here the possible pairs are (0, 0), (0, 1), (1, 0), (1, 1), (0, 2), (2, 0).

To obtain the probability associated with (1, 0), for example, we see that we are dealing with the event of getting one of the three green balls, no blue ball and hence, one of the red ball is the number of ways in which we get this event = ${}^{3}c_{1} \times {}^{2}c_{0} \times {}^{4}c_{1}$ =12. Also, the total number of ways in which two ball are drawn out of nine = ${}^{9}c_{2}$ = 36

As these probabilities are equally likely, the probability of the event associated with (1, 0) is 12/36 =1/3.

Similarly, the probability associated with (1, 1) is $({}^{3}c_1 \times {}^{2}c_0 \times {}^{4}c_1)/36=1/6$.

Likewise all the probabilities can be obtained and these values are shown in the following table

Probability Distribution → **Joint Probability Mass Function**

y	X		
	0	1	2
0	1/6	1/3	1/12
1	2/9	1/6	0
2	1/36	0	0

The above probabilities can also be obtained by

$$P(x,y)=\frac{{}^{3}c_{x}\,{}^{2}c_{y}\,{}^{4}c_{2\text{-}x\text{-}y}}{{}^{9}c_{2}},\ \text{for } x=y=\ 0,1,2 \text{ and } 0\le x+y\le 2$$

Joint Probability Distribution Function

If X and Y are random variables, the function given by

$$F(x, y) = P(X \leq x,\ Y \leq y) = \sum_{s \leq t} \sum_{t \leq y} p(x, y)$$

Example 1: If

$$f(x,y)=\begin{cases}\frac{2}{5}x(2x+3y), & 0\le x\le 1, 0\le y\le 1\\ 0, & \text{Elsewhere}\end{cases}$$

(i) Find P[X≤2/3, Y≤1/4]?

Discrete Marginal Distributions

If X and Y are random variables, then the marginal distribution of X=x is given by

$$g(\mathrm{x}) = \sum_{y} p(x, y)$$

If X and Y are random variables, then the marginal distribution of Y=y is given by

$$h(\mathrm{y}) = \sum_{x} p(x, y)$$

Example 1: If the joint probability mass function of two discrete random variables is given below, obtain the marginal distribution of X=x is

y	x		
	0	1	2
0	3/28	9/28	3/28
1	3/14	3/14	0
2	1/28	0	0

$P(X=0)=5/14$

$P(X=1)=15/28$

$P(X=2)=2/28$

Similarly, the marginal distribution of Y=y is

y	x		
	0	1	2
0	3/28	9/28	3/28
1	3/14	3/14	0
2	1/28	0	0

$P(Y=0)=15/28$

$P(Y=1)=3/7$

$P(Y=2)=1/28$

Summarizing the results:

The marginal distribution of X=x is

x	0	1	2
g(x)	5/14	15/28	3/28

The marginal distribution of Y=y is

y	0	1	2
h(y)	15/28	3/7	1/28

Conditional Distributions

If X and Y are random variables, then the conditional distribution of X=x given Y=y is

$$P(X = x/Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)} = \frac{p(x,y)}{h(y)}, \mathrm{h(y)} > 0$$

Similarly, the conditional distribution of Y=y given X=x is

$$P(Y = y/X = x) = \frac{P(X = x, Y = y)}{P(X = x)} = \frac{p(x,y)}{g(x)}, \mathrm{g(x)} > 0$$

Example 1: If the joint probability mass function of two discrete random variables is given below:

y	x		
	-1	0	1
0	1/15	2/15	1/15
1	3/15	2/15	1/15
2	2/15	1/15	2/15

$P(X=-1 \mid Y=2)=2/5$

$P(X=0 \mid Y=2)=1/5$

$P(X=2 \mid Y=2)=2/5$

Find the conditional distribution of X given Y=2

Independence of random variables

If X and Y are two random variables, with joint probability distributions $p(x, y)$ and the marginal distributions $g(x)$ and $h(y)$ respectively, then they are called as statistically independent if and only if $p(x, y) = g(x)h(y)$ for all (x, y) within their range.

Note: X and Y may be discrete or continuous.

Special Probability Distributions

Prof.Gangaboraiah, PhD

Learning objectives of unit 2

- At the end of the session on unit 2, the student should be able to
  - Distinguish between Discrete and Continuous probability distribution
  - Understand different types of probability distribution like Binomial, Poisson, Normal, Uniform, Exponential distributions

Bernoulli distribution

- Single random experiment
- Dichotomous outcome
- X=No. of success
- Probability of success is ‘p’ and failure ‘q = 1 - p’
- Probability that no. of success in a single trial is
- $P(X=x) = p^x q^{(1-x)}$, $x = 0,1$
- When $x = 0$, $P(X=0) = p^0 q^{(1-0)} = 1 \times q = q$
- When $x = 1$, $P(X=1) = p^1 q^{(1-1)} = p \times 1 = p$

Bernoulli distribution

Probability that no. of success in a single trial is

$$p(x) = P(X = x) = p^x q^{1-x}, x = 0, 1$$

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

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

Binomial distribution

- ‘n’ number of finite trials
- Dichotomous outcome
- X = No. of success
- Probability of success is ‘p’
- Probability of failure is ‘q=1-p’

Example

- A milk vendor sells three grades of milk viz., grade 1, grade 2, and grade 3. The milk is stored in the refrigerator. Due to technical/ mechanical reasons of refrigerator or old stock of milk, the milk may get spoiled. Let N denote not spoiled and S denote spoiled

$$\Omega = \{NNN, NNS, NSN, SNN, NSS, SSN, SNS, SSS\}$$

Outcome	X	f	Probability	
NNN	0	1	⅛	q^3
NNS	1	3	⅜	${}^3c_1q^2p$
NSN	1			
SNN	1			
NSS	2	3	⅜	${}^3c_2qp^2$
SNS	2			
SSN	2			
SSS	3	1	⅛	p^3
	Total	8	1	$(q+p)^3$

$$(q+p)^3 = q^3 + {}^3c_1 q^2 p + {}^3c_2 q p^2 + p^3$$

ie., $(q+p)^3 = q^3 + 3q^2p + 3qp^2 + p^3$

In general,

$$(q+p)^n = q^n + {}^nc_1 q^{n-2} p + {}^nc_2 q^{n-2} p^2 + \ldots + {}^nc_x q^{n-x} p^x + \ldots + p^n$$

Binomial distribution

- Binomial: Suppose that *n* independent experiments, or trials, are performed, where *n* is a fixed number, and that each experiment results in a “success” with probability *p* and a “failure” with probability q = *1-p*. The total number of successes, *X*, is a binomial random variable with parameters *n* and *p*.
- We write: *X ~ Bin (n, p) {reads: “X is distributed binomially with parameters n and p}*
- The probability that *X* = *x* (i.e., that there are exactly *x* successes) is given by:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

n = # of trials

1-p = probability of failure

$$^{n}c_{x}\,p^{x}(1-p)^{n-x}$$

x = # of success

p = probability of success

Probability that x success in n trials is

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \; x = 0,\ 1,\ 2,\ 3, \ldots, n$$

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

$$p(x) = \begin{cases} \binom{n}{x} p^x q^{n-x}, & x = 0,\ 1,\ 2,\ 3, \ldots, n \\ 0 & \text{elsewhere} \end{cases}$$

Example

- If a coin is tossed 6 times, what is the probability of getting 2 or fewer heads?

$$P(X=0)=\binom{6}{0}(0.5)^0(0.5)^6=\frac{6!}{6!0!}(0.5)^6=0.015625$$

$$P(X=1)=\binom{6}{1}(0.5)^1(0.5)^5=\frac{6!}{5!1!}(0.5)^6=0.09375$$

$$P(X=2)=\binom{6}{2}(0.5)^1(0.5)^4=\frac{6!}{4!2!}(0.5)^6=0.078125$$

- $P(x \le 2) = \sum p(x) = 0.015625 + 0.09375 + 0.078125 = 0.1875$

Mean and Standard deviation of Binomial distribution

- All probability distributions are characterized by an expected value and a variance
- If *X* follows a binomial distribution with parameters n and p then we write $X \sim Bin\ (n, p)$
- Then:
- $\mu_x = E(X) = np$
- $\sigma_x^2 = Var\ (X) = np(1-p)$
- $\sigma_x = SD\ (X) = \sqrt{np(1-p)}$

Note: the variance will always lie between

0*n -0.25 *n

p(1-p) reaches maximum at p=0.5

p(1-p)=.025

Problem of Binomial distribution

- After demonetisation customers stand in the queue at ATM's to get new currencies. It was found that the probability of getting a new currency of ` 2000/- by a customer is 0.76. In a randomly selected 5 customers what is the probability that:
- i) None got new ` 2000/- currency
- ii) Between 2 to 4 got new ` 2000/- currency
- iii) More than 4 got new ` 2000/- currency

Poisson distribution

- **Binomial distribution**
  - No. of events-finite (n is finite)
  - Probability of success very large (p is large)
- **When**
  - No. of events are very large ($n \to \infty$)
  - Probability of success very small ($p \to 0$)
- **This results in a probability distribution called**

  POISSON DISTRIBUTION

Probability mass function of Poisson distribution

Probability that x occurrences in an infinite sequence of trials is

$$P(X = x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

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

$$p(x) = \begin{cases} e^{-\lambda}\dfrac{\lambda^x}{x!}, & x = 0, 1, 2, \ldots \\ 0 & \text{elsewhere} \end{cases}$$

Expected value and a variance of Poisson distribution

If X follows Poisson distribution with parameters np = λ, i.e., *X ~ P(λ)* its probability mass function is given by

$$p(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x\,!}, & x = 0, 1, 2, \ldots \\ 0, & \text{elsewhere} \end{cases}$$

Then:

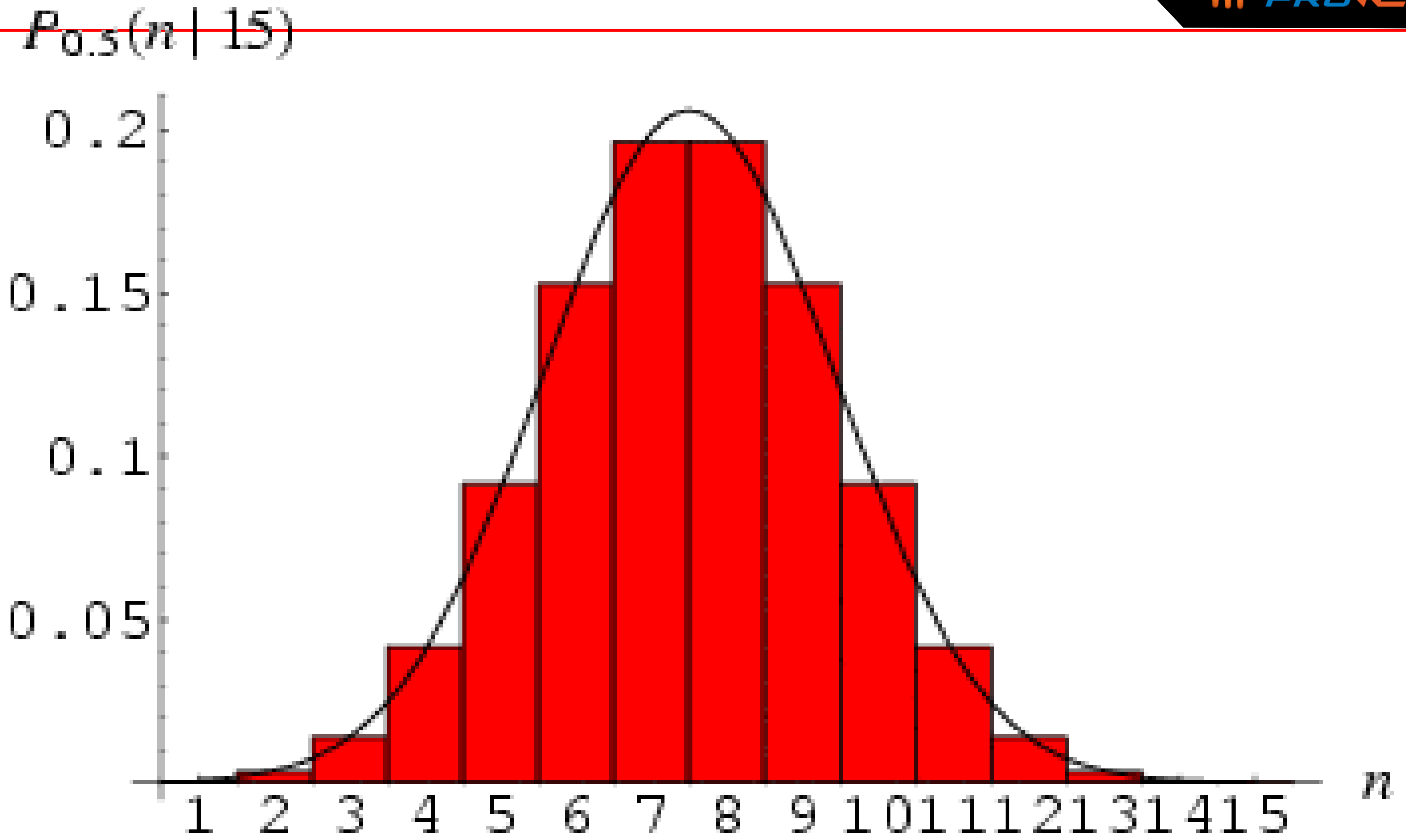
$\mu_x = E(X) = np = \lambda$

$\sigma_x^2 = Var(X) = np(1-p) = \lambda$

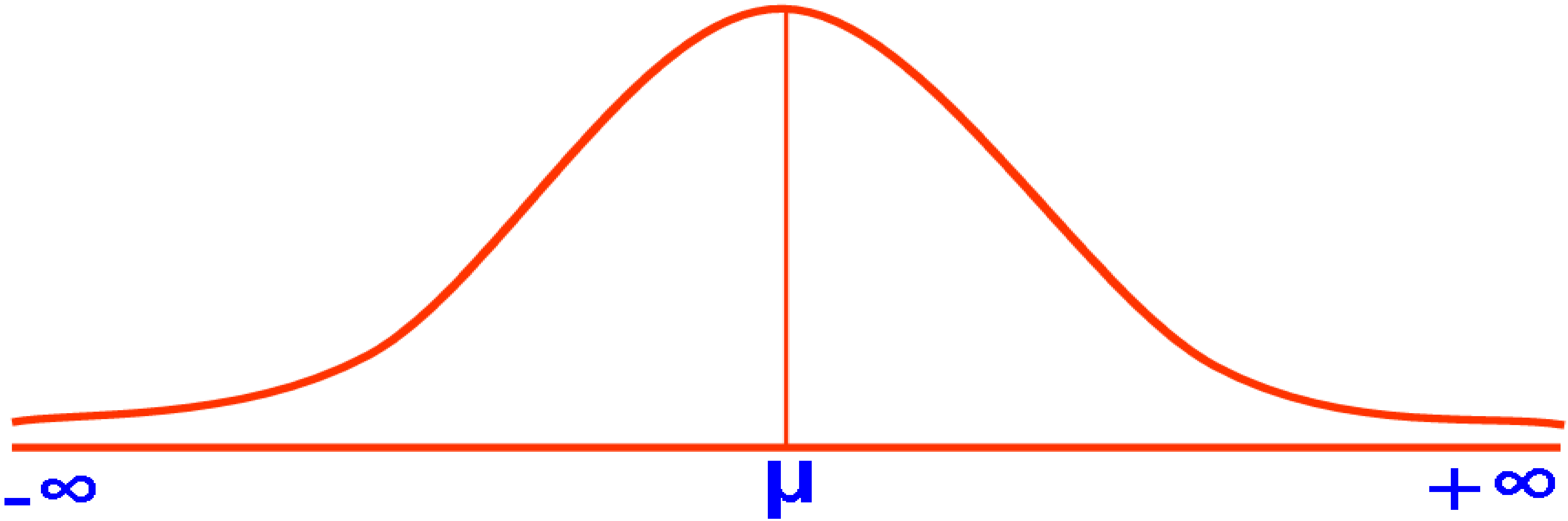
$\sigma_x = SD(X) = \sqrt{\lambda}$

Normal distribution

When mid points of the top of the bars of a histogram are joined by straight lines, we get a frequency polygon. Smoothed shape of the frequency polygon is called a frequency curve. If the mean, median and mode values are same, the curve is symmetrical. This frequency distribution is known as normal distribution.


$P_{0.5}(n \mid 15)$
0.2
0.15
0.1
0.05
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
n

Normal curve



Characterization of Normal distribution

- Bell shaped curve, symmetrical on either side of mean

- Mean = Median = Mode

- The point of inflexion is at $\mu \pm 1\sigma$

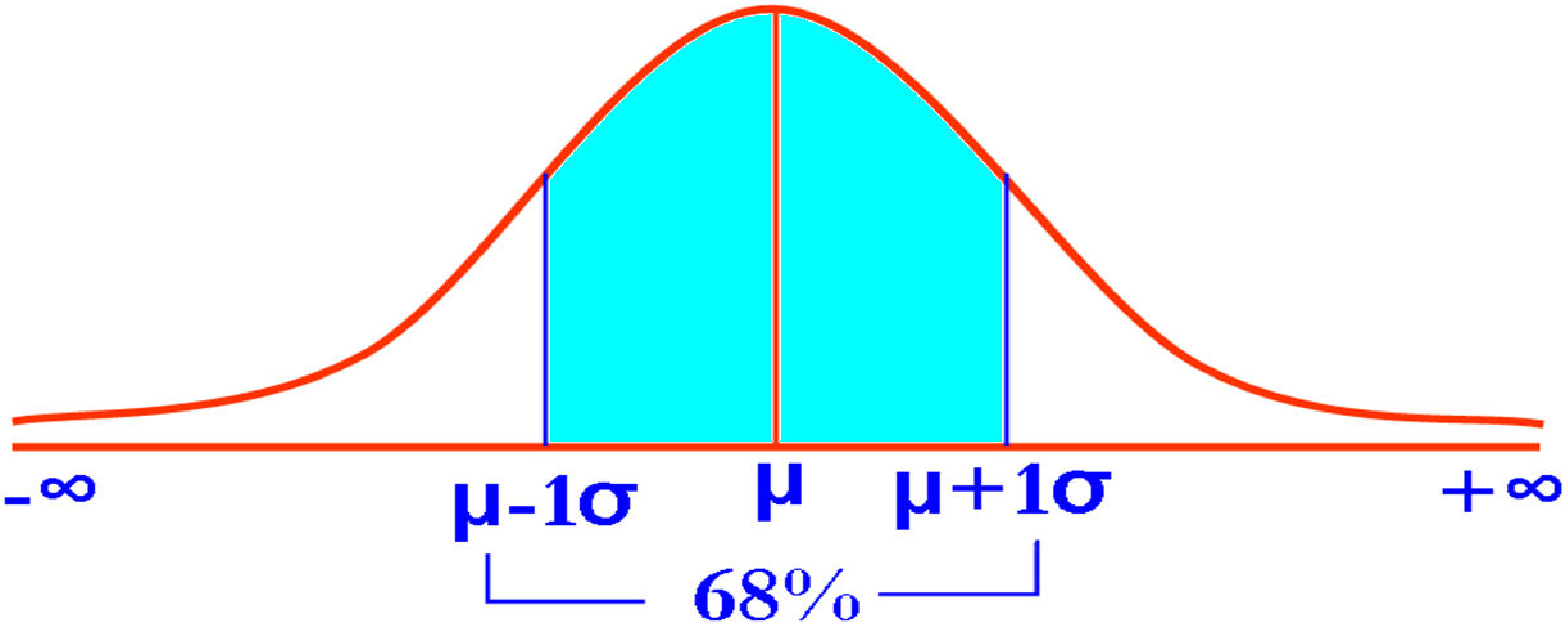
- The measure of skewness

$$S_k = \frac{\text{Mean-Mode}}{\text{SD}} = 0$$

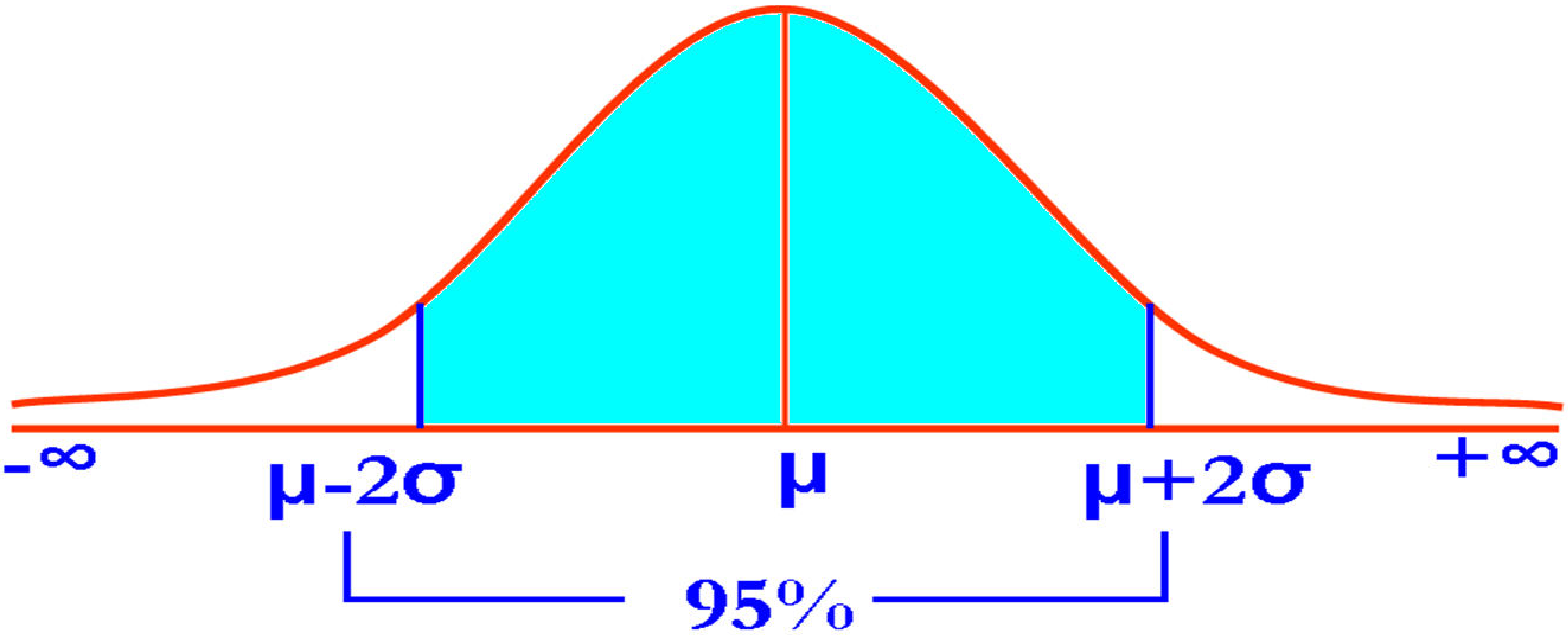
Characterization of Normal distribution

- 68.0% of values in 1σ on either side of the mean.
- 95.0% of values in 2σ on either side of the mean.
- 99.9% of values in 3σ on either side of the mean.

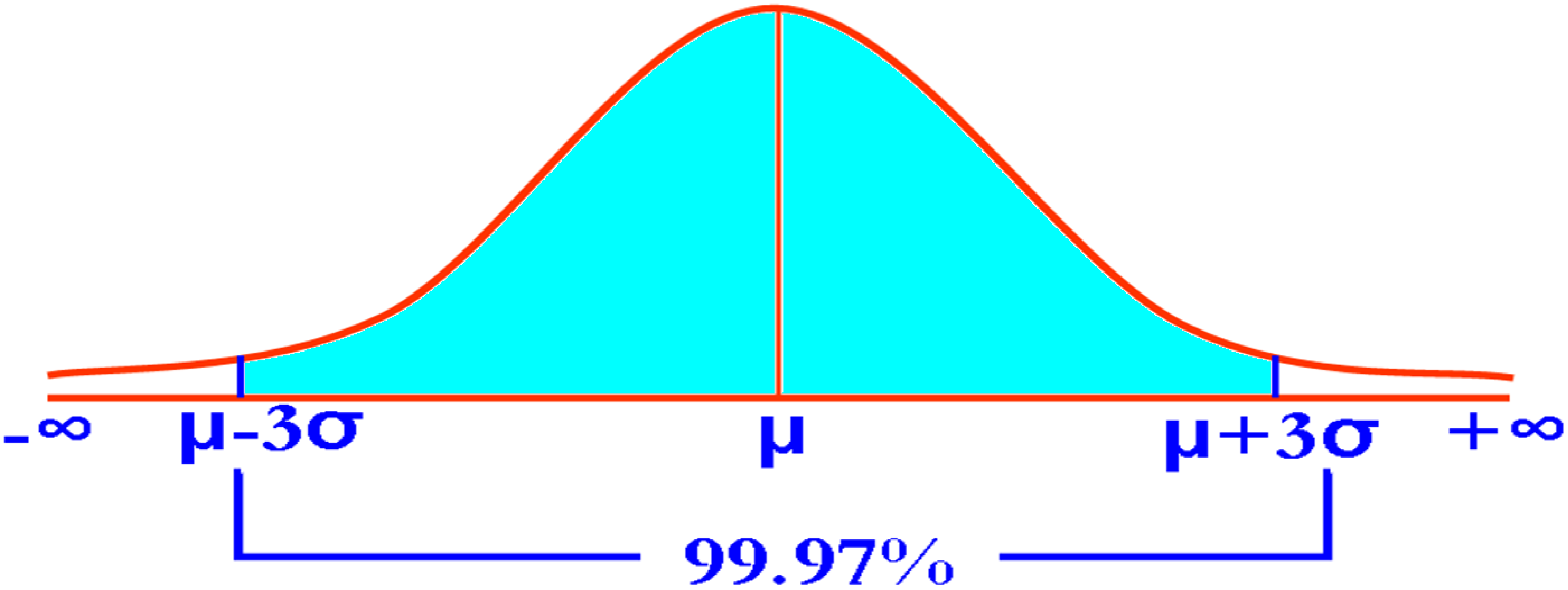
Area under the curve: Mean ±1SD



Area under the curve: Mean ± 2SD



Area under the curve: Mean ± 3SD



Skewness

Positive skew: The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed

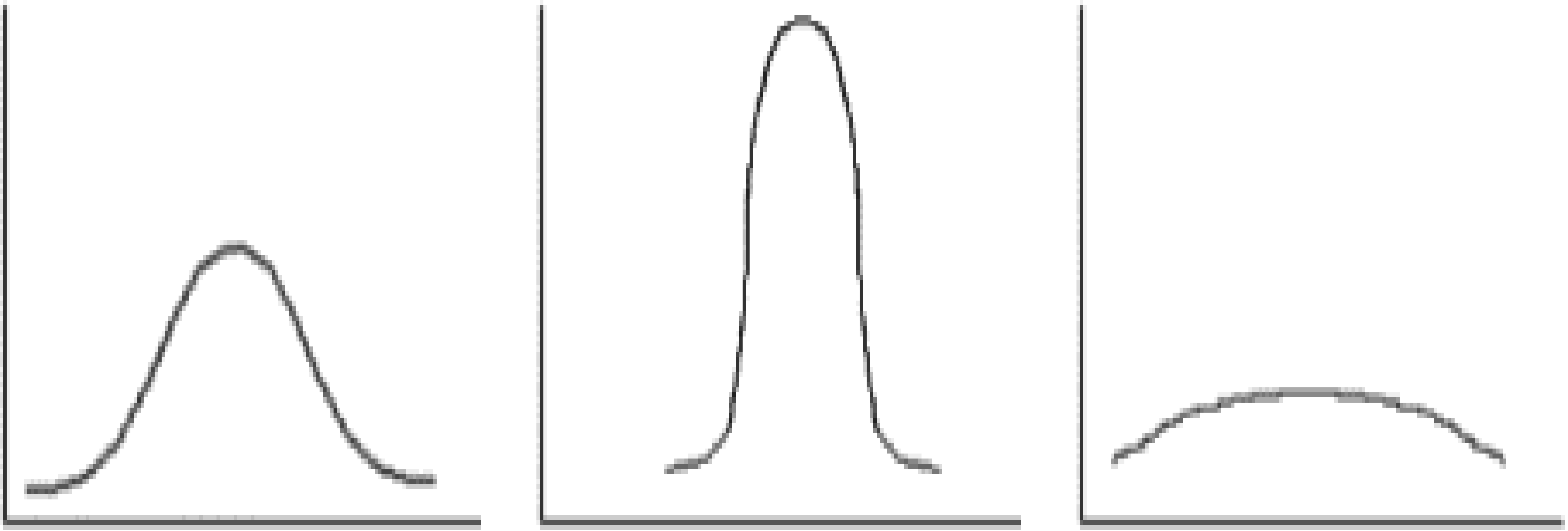
Negative skew: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed

Measure of Skewness: **Karl Pearsons measure of skewness is given by**

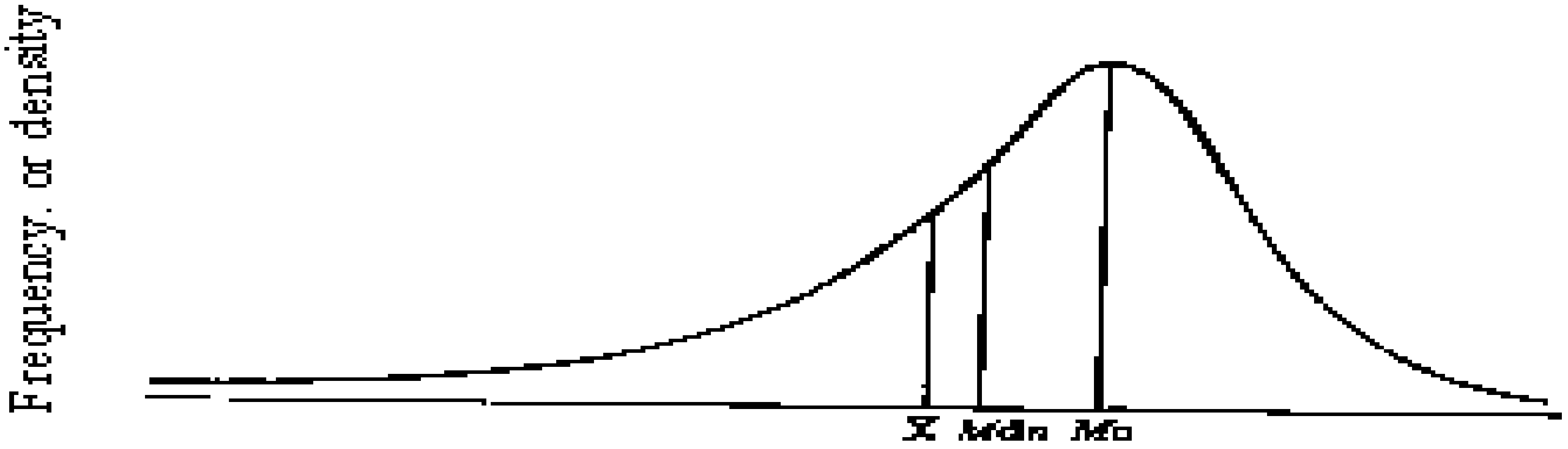
$$z = \frac{\text{Mean-Mode}}{\text{SD}} = 0$$

Kurtosis

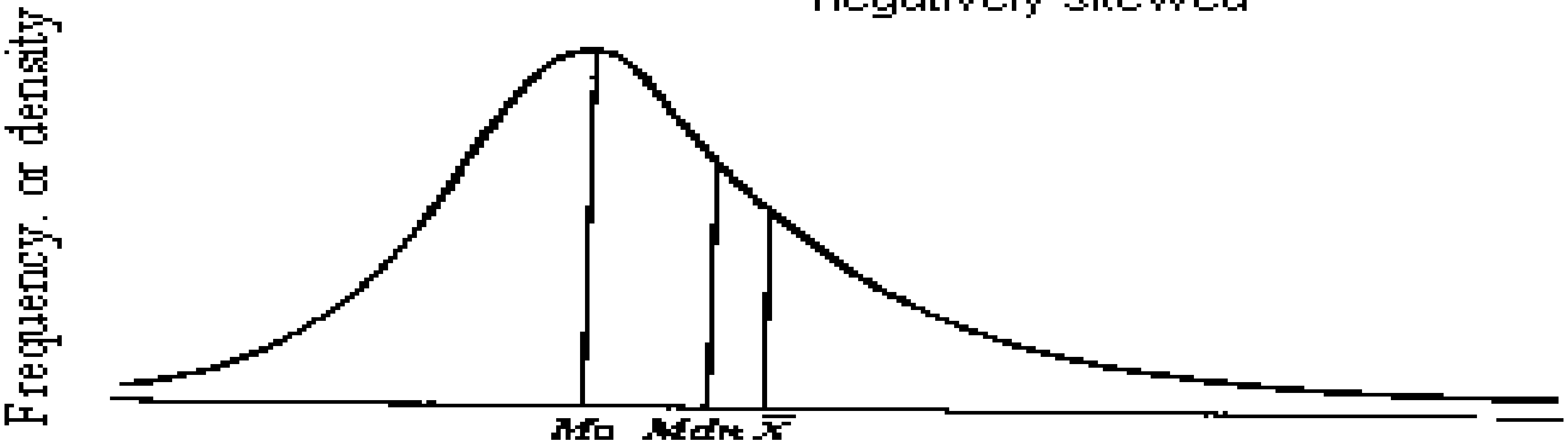
Kurtosis measures the peakedness of a curve. The curve may be leptokurtic, mesokurtic or platykurtic. The ideal curve is mesokurtic. The measure of kurtosis is denoted by ß.



Mesokurtic Curve Leptokurtic Curve Platykurtic Curve



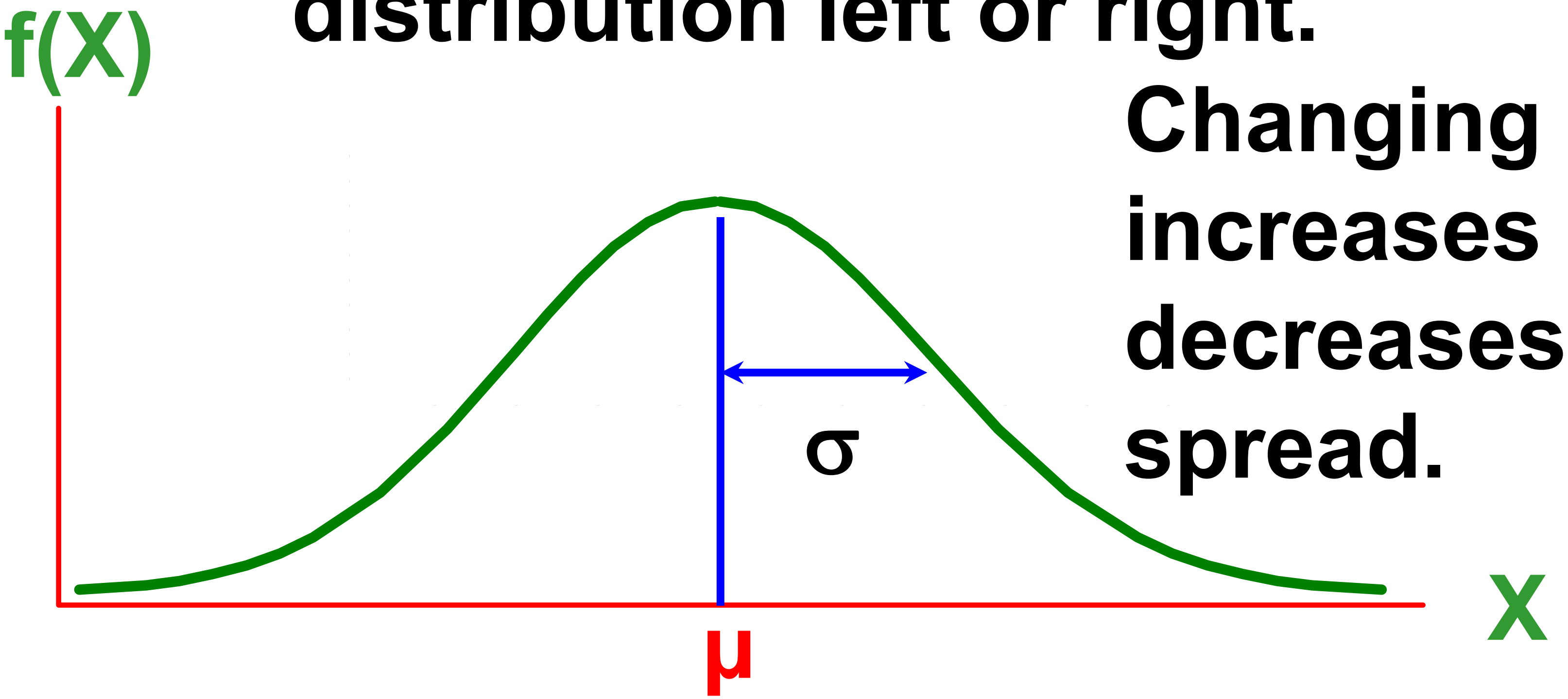
negatively skewed



positively skewed

The Normal Distribution

Changing μ shifts the distribution left or right.



Changing σ increases or decreases the spread.

The Normal Distribution: as mathematical function (pdf)

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

Note constants:

π=3.14159

e=2.71828

This is a bell shaped curve with different centers and spreads depending on μ and σ

The Normal PDF

It's a probability function, so no matter what the values of μ and σ, must integrate to 1.

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

Normal distribution is defined by its mean and standard dev.

$E(X) = \mu$

$Var\ (X) = \sigma^2$

Standard Deviation $(X) = \sigma$

The beauty of the normal curve:

No matter what μ and σ are, the area between

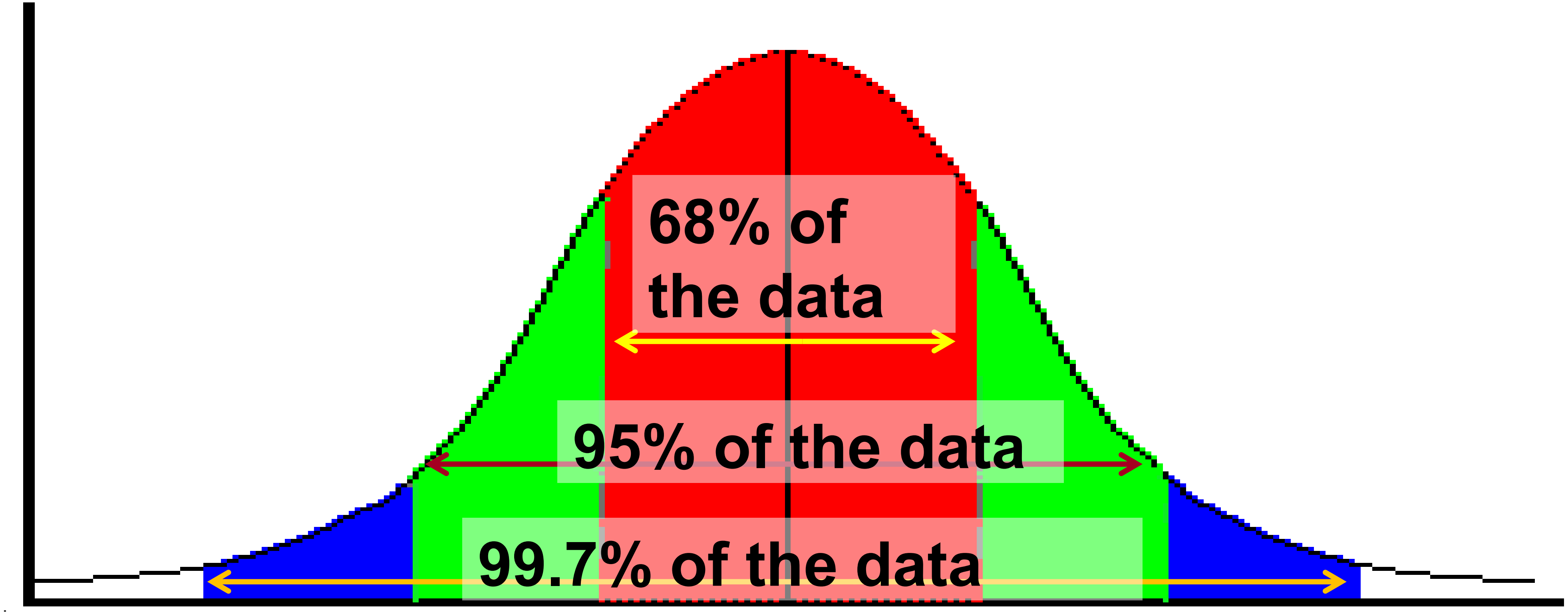
$\mu-\sigma$ and $\mu+\sigma$ is about 68%

$\mu-2\sigma$ and $\mu+2\sigma$ is about 95%

$\mu-3\sigma$ and $\mu+3\sigma$ is about 99.7%

Almost all values fall within 3 standard deviations.

68-95-99.7 Rule



68-95-99.8 Rule: in Math terms…

$$\int_{\mu-\sigma}^{\mu+\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .68$$

$$\int_{\mu-2\sigma}^{\mu+2\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .95$$

$$\int_{\mu-3\sigma}^{\mu+3\sigma} \frac{1}{\sigma\sqrt{2\pi}} \bullet e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = .998$$

How good is rule for real data?

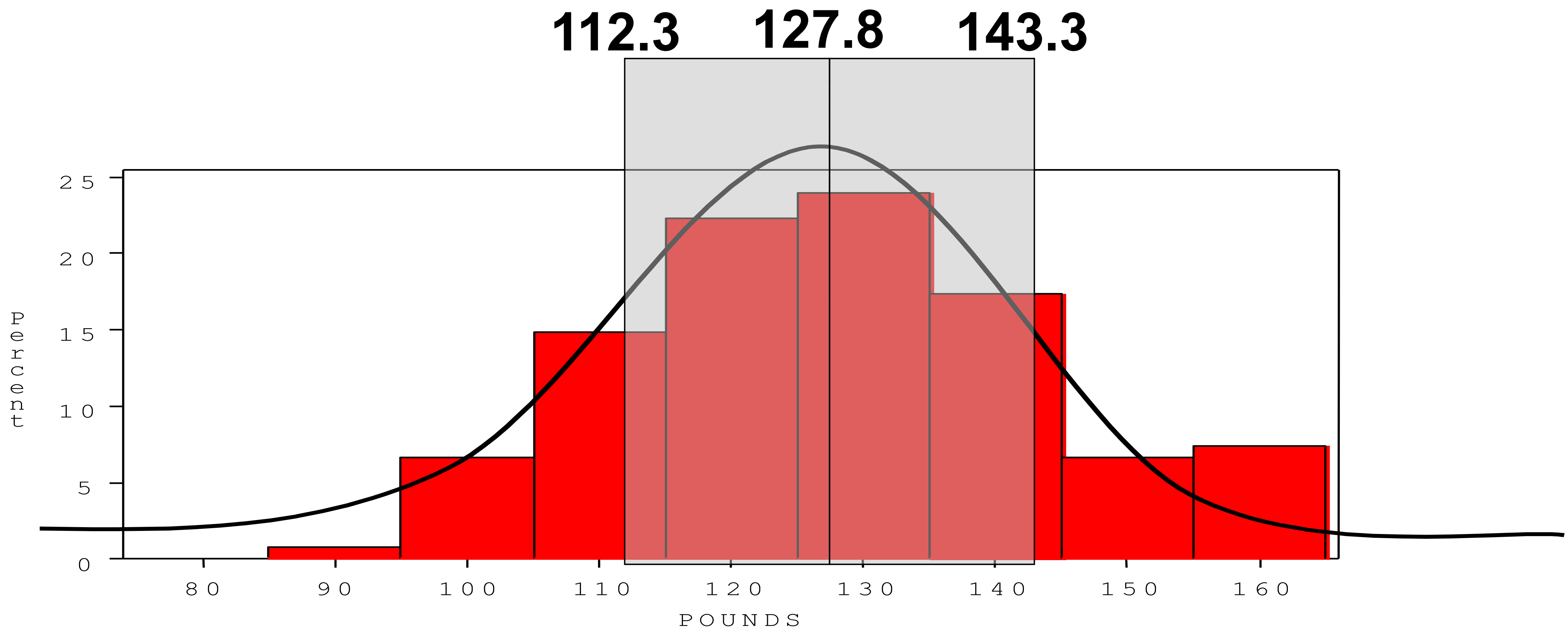
Check some example data on women runners :

The mean of the weight μ= 127.8 lbs

The standard deviation (SD) σ= 15.5 lbs

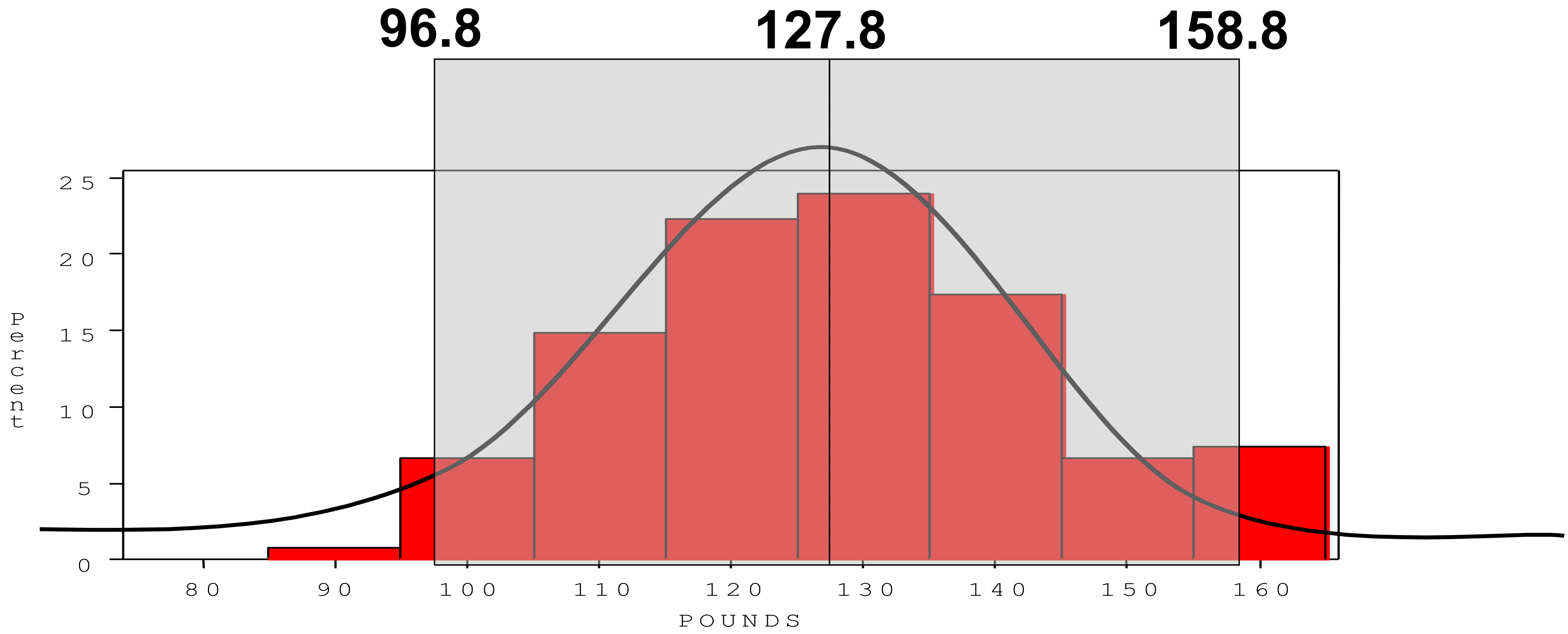
68% of 120 = .68x120 = ~ 82 runners

In fact, 79 runners fall within 1-SD (15.5 lbs) of the mean.



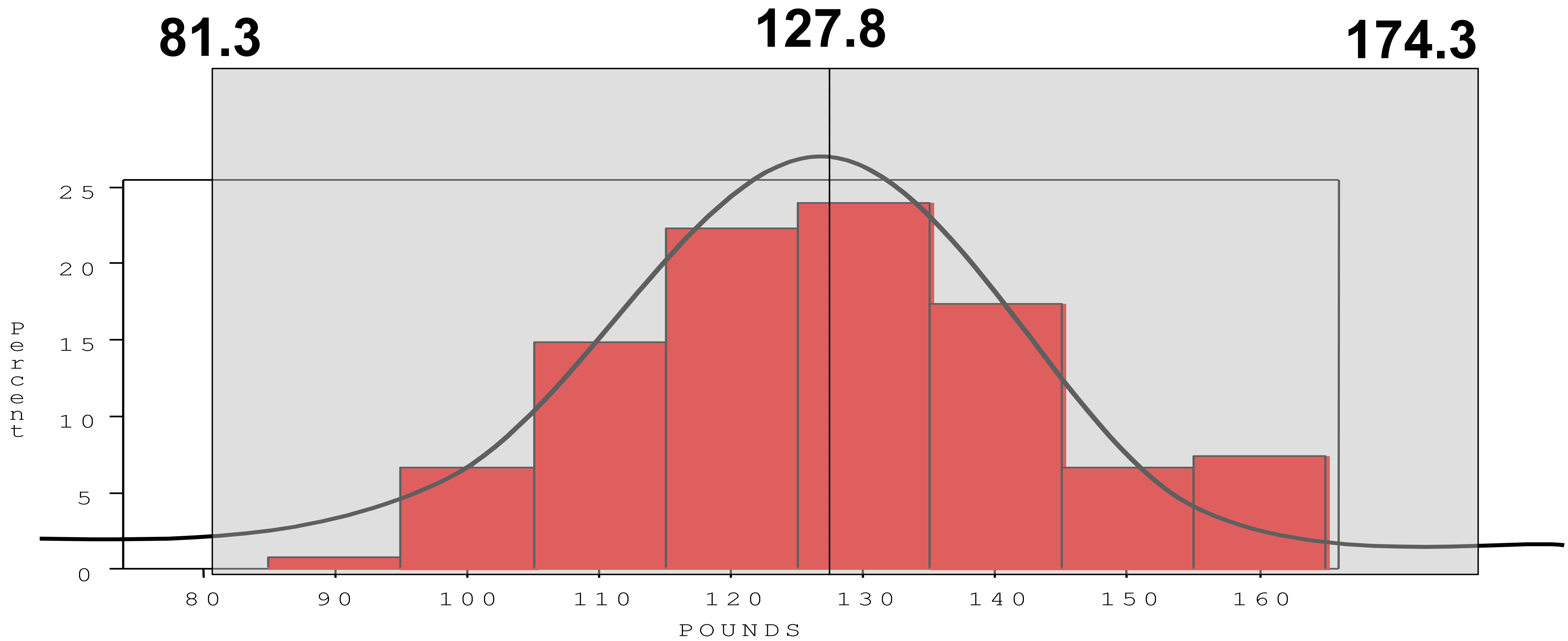
95% of 120 = .95 x 120 = ~ 114 runners

In fact, 115 runners fall within 2-SD's of the mean.



99.7% of 120 = .997 x 120 = 119.6 runners

In fact, all 120 runners fall within 3-SD's of the mean.

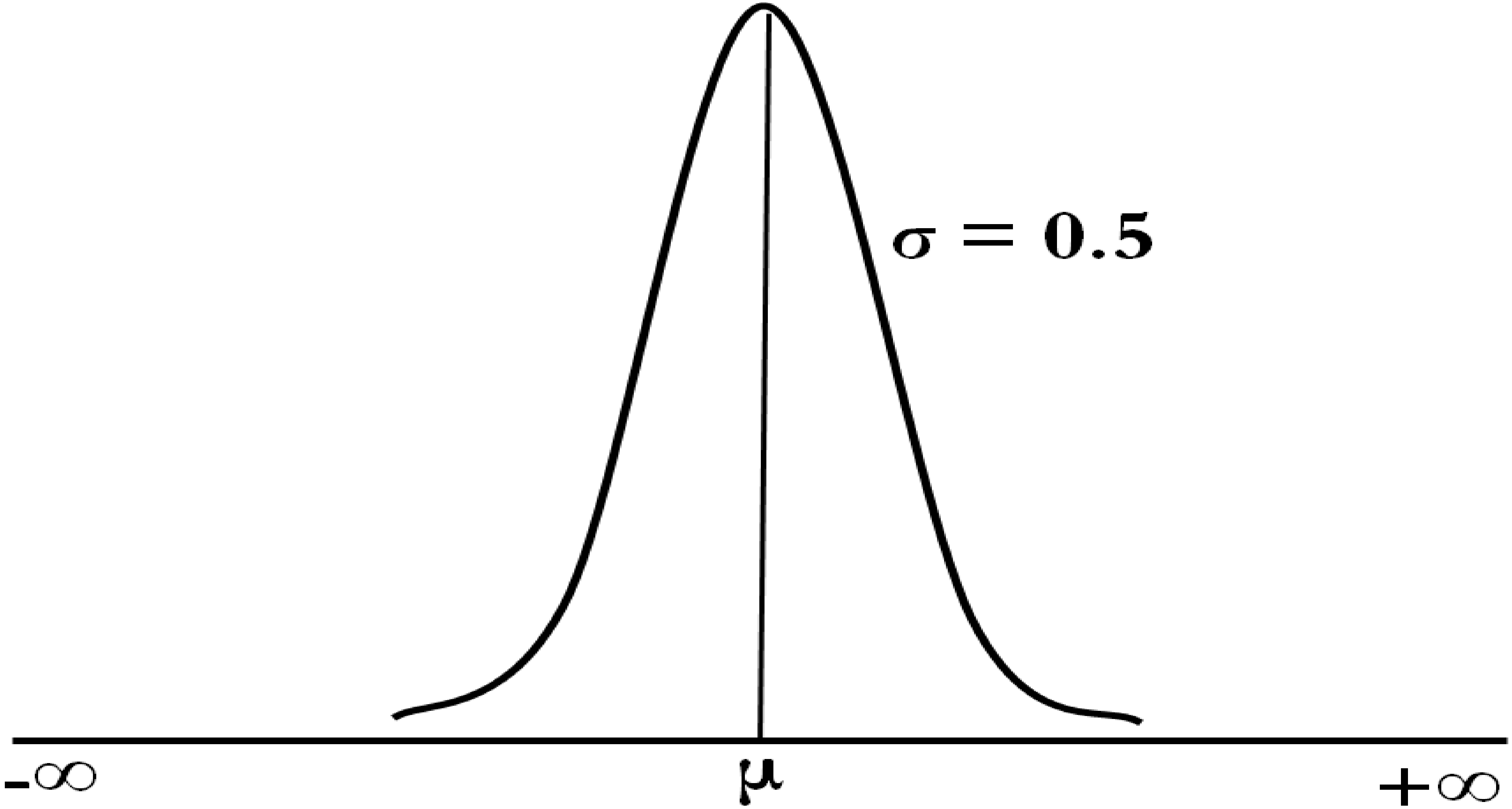


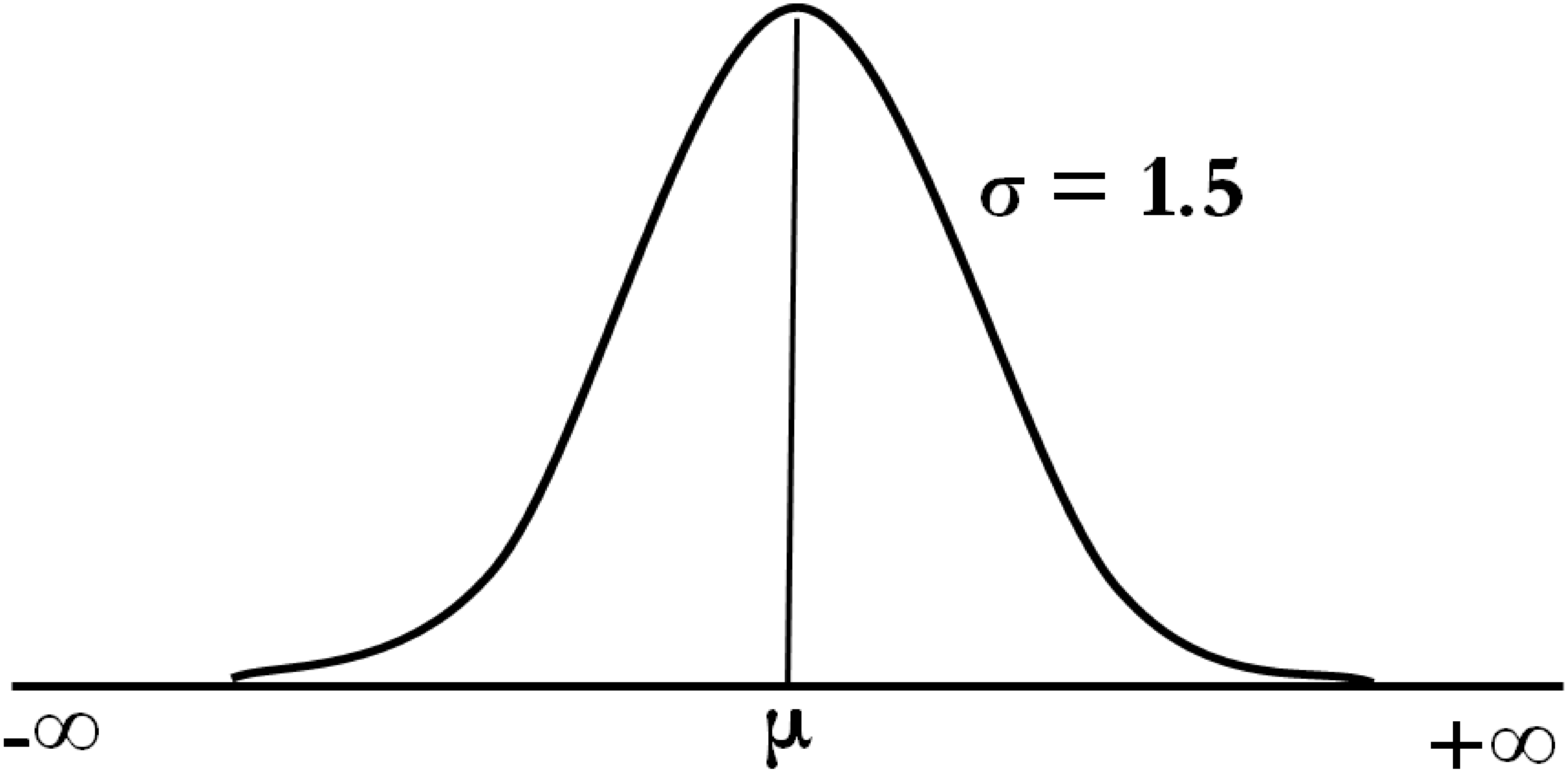
Example

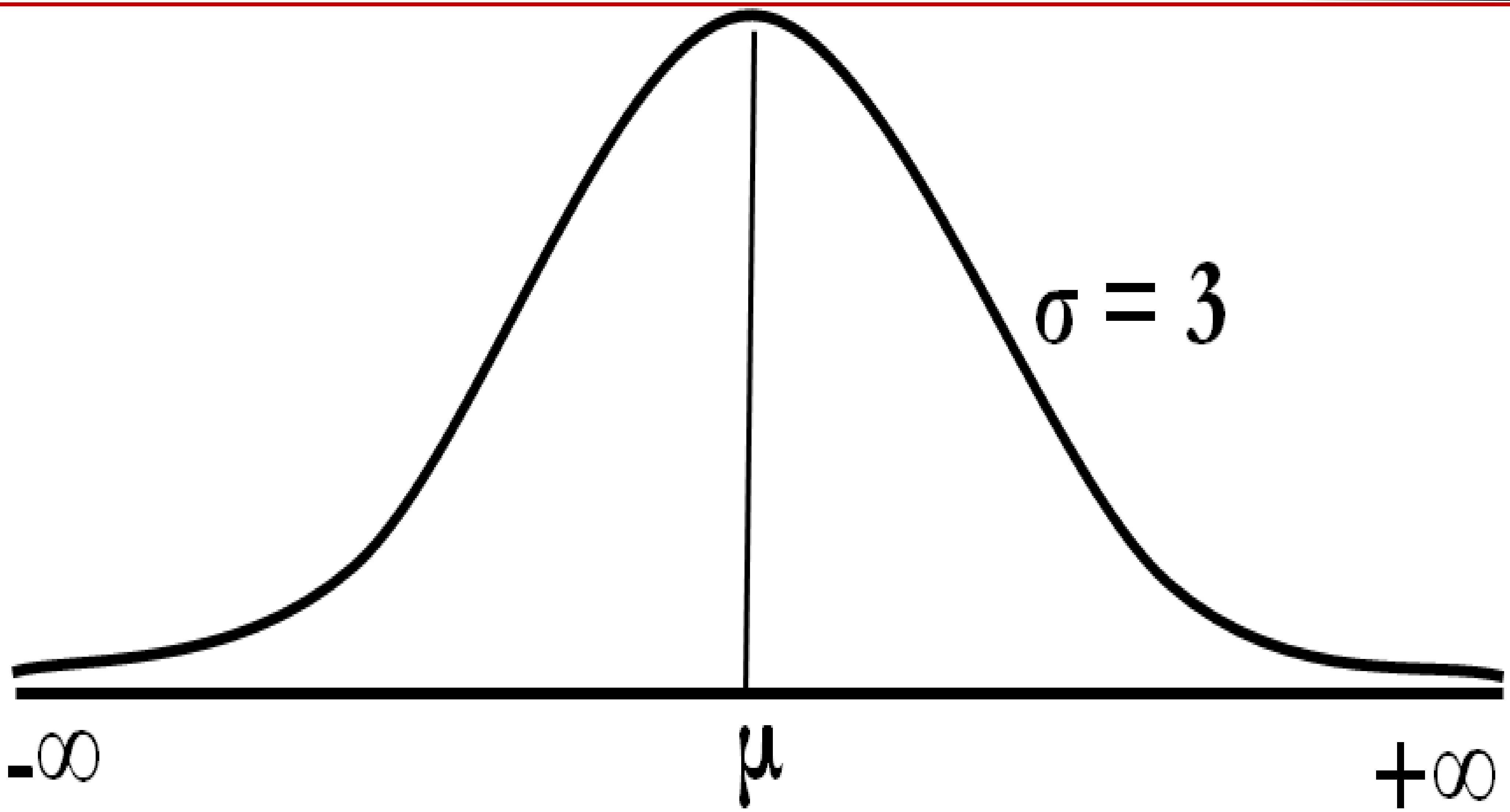
Suppose SAT scores roughly follows a normal distribution in the Indian population of college-bound students (with range restricted to 200-800), and the average math SAT is 500 with a standard deviation of 50, then:

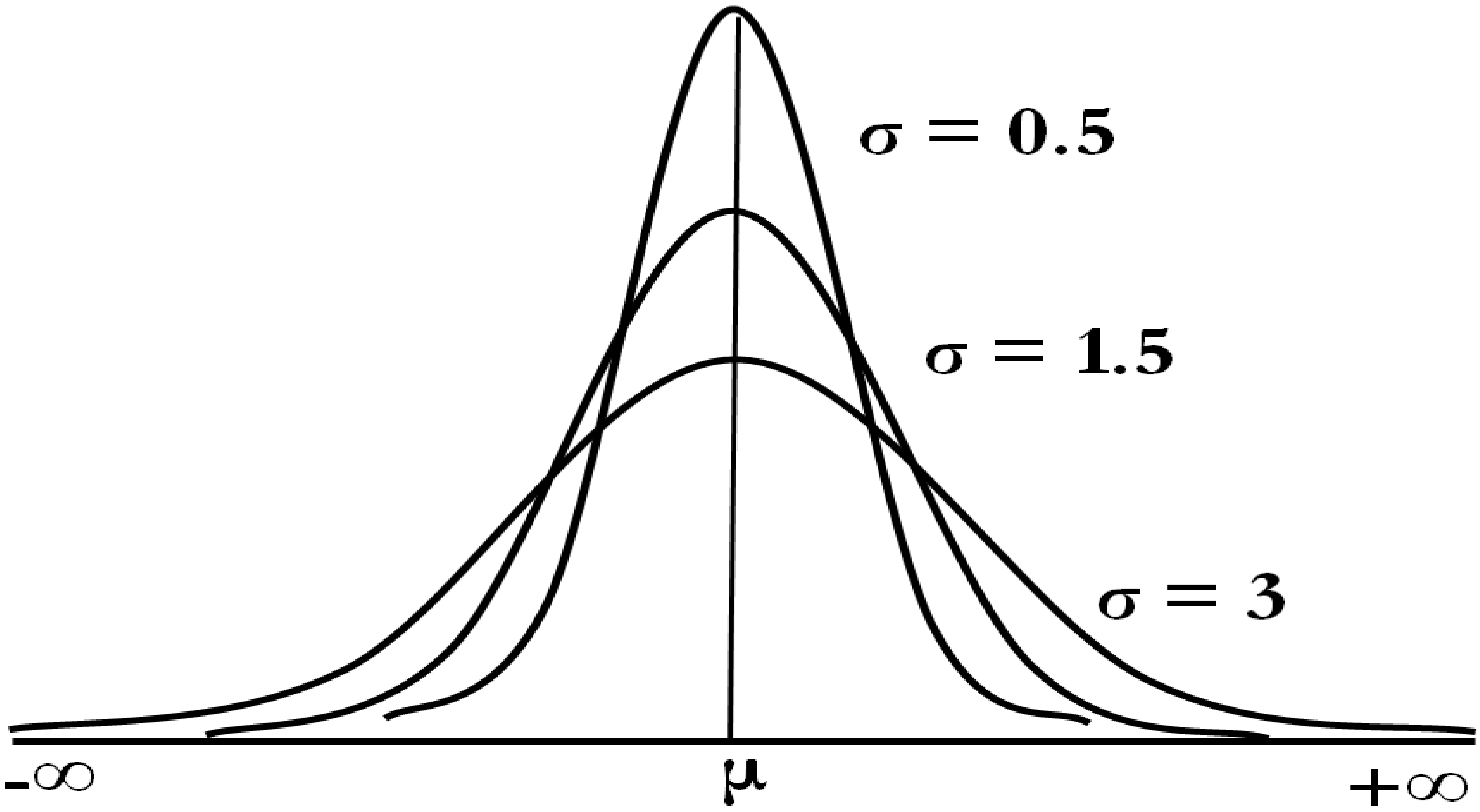
- 68% of students will have scores between 450 and 550
- 95% will be between 400 and 600
- 99.7% will be between 350 and 650

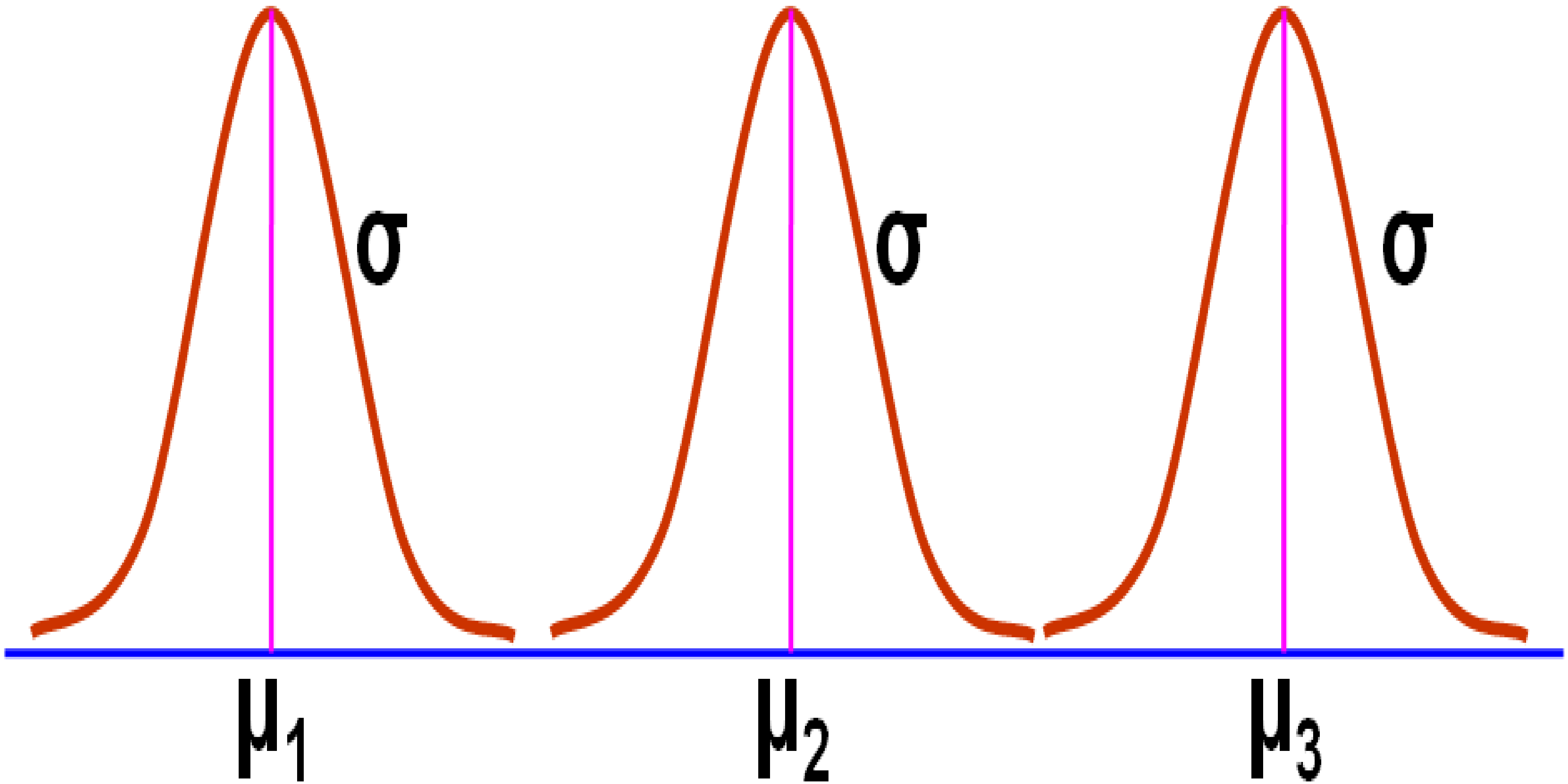
Standard Normal Distribution


σ = 0.5
-∞
μ
+∞


$\sigma = 1.5$
$-\infty$
μ
$+\infty$


σ = 3
-∞
μ
+∞


σ = 0.5
σ = 1.5
σ = 3
-∞
μ
+∞


σ
σ
σ
μ1
μ2
μ3

Standard Normal distribution

Any distribution can be converted to a standardised distribution. However the symmetry of the original distribution remains unchanged. If the original distribution was skewed to start with, it will still be skewed after the z-score transformation.

In the special case where the original distribution can be considered normal, standardising will result in what is known as the standard normal distribution.

The standard Normal distribution is a special member of the normal family that has a mean of 0 and a standard deviation of 1. The standard normal random variable is denoted by Z.

Normal distributions can be transformed to standard normal distributions by using the formula

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

where X is a score from the original normal distribution, μ is the mean of the original normal distribution, and σ is the standard deviation of original normal distribution.

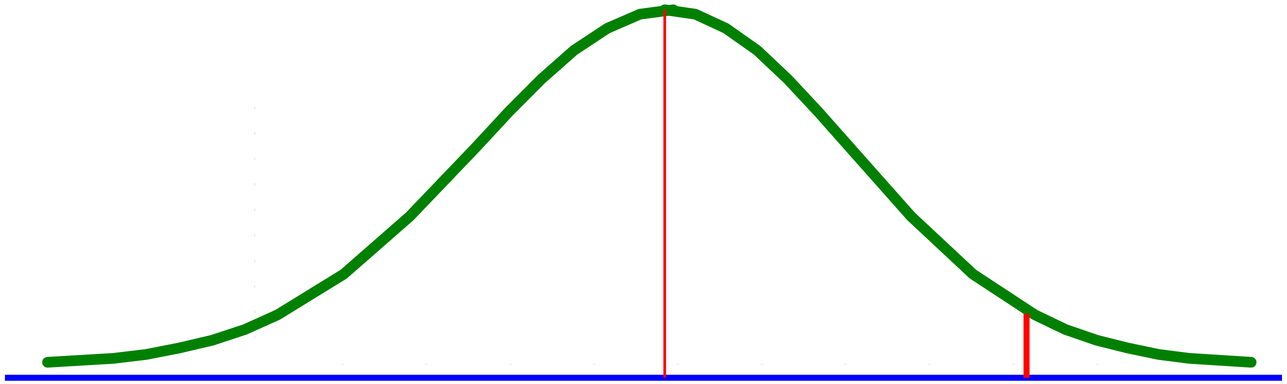
Sl No	X	Z=(X-Mean)	Z=(X-Mean)/S
1	3000	1005	1.19
2	2486	491	0.58
3	827	-1168	-1.39
4	1678	-317	-0.38
5	2070	75	0.09
6	2638	643	0.76
7	2490	495	0.59
8	1865	-130	-0.15
9	1000	-995	-1.18
10	2090	95	0.11
11	596	-1399	-1.66
12	3200	1205	1.43
Mean	1995.00	0	0
SD	842.35		1.00

The Standard Normal (Z): "Universal Currency"

The formula for the standardized normal probability density function is

$$p(Z)=\frac{1}{(1)\sqrt{2\pi}}\cdot e^{-\frac{1}{2}\left(\frac{z-0}{1}\right)^2}=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}(Z)^2}$$

Comparing X and Z units



100 200 X ($\mu = 100, \sigma = 50$)

0 2.0 Z ($\mu = 0, \sigma = 1$)

Example

If birth weights in a population are normally distributed with a mean of 3113 gms with a standard deviation of 195 gms,

a. What is the chance of obtaining a birth weight of 3481 gms *or heavier* when sampling birth records at random?

b. What is the chance of obtaining a birth weight of 3220 gms *or lighter*?

Answer

a. What is the chance of obtaining a birth weight of 3481 oz *or heavier* when sampling birth records at random?

$$\boxed{Z = \frac{3481 - 3113}{195} = 1.89}$$

From the chart Z of 1.89 corresponds to a right tail (greater than) area of: $P(Z \geq 1.89) = 1-(0.9706) = 0.0294$ or 2.94 %

Answer

b. What is the chance of obtaining a birth weight of 120 *or lighter*?

$$\boxed{Z = \frac{3220 - 3113}{195} = 0.54}$$

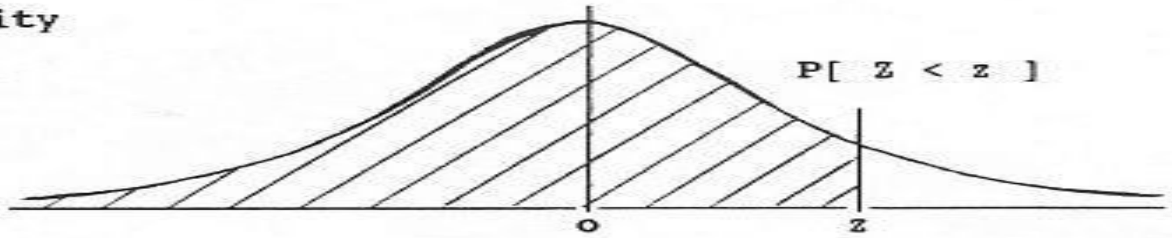
From the chart Z of 0.54 corresponds to a left tail area of: P(Z ≤ 0.54) = 0.7054 = 70.54%

STANDARD STATISTICAL TABLES

1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z
i.e.

$$P[\ Z < z\ ] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\ dZ$$



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

STANDARD STATISTICAL TABLES

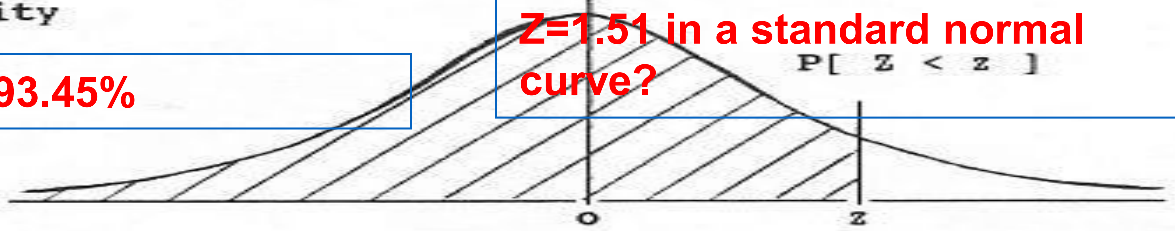
1. Areas under the Normal Distribution

The table gives the cumulative probability up to the standardised normal value z i.e.

$$P[\ Z < z\ ] = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp(-\tfrac{1}{2}Z^2)\, dZ$$

Area is 93.45%

What is the area to the left of Z=1.51 in a standard normal curve?



Z=1.51

Z=1.51

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5159	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7854
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8804	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9874	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9924	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9980	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

z	3.00	3.10	3.20	3.30	3.40	3.50	3.60	3.70	3.80	3.90
P	0.9986	0.9990	0.9993	0.9995	0.9997	0.9998	0.9998	0.9999	0.9999	1.0000

Def:

If $Z \sim N(0,1)$, we shall say that Z has a standard normal distribution and the distribution function of Z is

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-w^2/2} dw.$$

Remark

- It is impossible to evaluate this integration by finding an antiderivative. However, numerical approximattions for this integral are given in Table Va and Vb.
- Because of the symmetry of the standard normal p.d.f., when z > 0,

$$\Phi(-z) = P(Z \le -z) = P(Z \ge z)$$

$$= 1 - P(Z \le z) = 1 - \Phi(z).$$

Example -3 If Z is $N(0,1)$, then find

(i) $P(1.24 \leq Z \leq 2.37)$

(ii) $P(-2.37 \leq Z \leq -1.24)$

$$P(Z \leq 1.24) = \Phi(1.24) = 0.8925,$$

$$P(1.24 \leq Z \leq 2.37) = \Phi(2.37) - \Phi(1.24)$$

$$= 0.9911 - 0.8925 = 0.0986.$$

$$P(-2.37 \leq Z \leq -1.24) = P(1.24 \leq Z \leq 2.37) = 0.0986.$$

Example -3 If Z is $N(0,1)$, then find

(i) P(Z > 1.24) (ii) P(Z ≤ -2.14) (iii) P(-2.14 ≤ Z ≤ 0.77)

$$P(Z > 1.24) = 0.1075,$$

$$P(Z \leq -2.14) = P(Z \geq 2.14) = 0.0162$$

$$P(-2.14 \leq Z \leq 0.77) = P(Z \leq 0.77) - P(Z \leq -2.14)$$

$$= 0.7794 - 0.0162 = 0.7632$$

Example - 4 If Z ~ N(0,1), to find constants a and b such that

$$P(Z \leq a) = 0.9147 \quad \text{and} \quad P(Z \geq b) = 0.0526$$

From Table, we see that a =1.37 and b =1.62.

Definition

- Let Z~N (0, 1) and z_α denote 100 (1-α) percentile of Z.
- That is

$$P\left[Z \geq z_\alpha\right] = \alpha$$

- Then 100 α percentile is denoted by $z_{1-\alpha}$
- That is

$$P\left[Z \geq z_{1-\alpha}\right] = 1 - \alpha$$

Remark

$$P(Z \le -z_{\alpha}) = P(Z \ge z_{\alpha}) = \alpha$$

and

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

Thus, $z_{1-\alpha} = -z_{\alpha}$.

Example

Find from table

(i) $z_{0.0125}$

(ii) $z_{0.05}$

(iii) $z_{0.025}$

Solution

(i) $z_{0.0125} = 2.24$

(ii) $z_{0.05} = 1.645$

(iii) $z_{0.025} = 1.96$

Example

If $X \sim N(3,16)$, then find (i) P(4 < X < 8)

(ii) P(-2 < X < 1)

Solution

$$\text{(i) } P(4 \le X \le 8) = P\left(\frac{4-3}{4} \le \frac{X-3}{4} \le \frac{8-3}{4}\right)$$

$$= P(0.25 \le Z \le 1.25) = \Phi(1.25) - \Phi(0.25)$$

$$= 0.8944 - 0.5987 = 0.2957.$$

Example

$$\text{(ii) } P(-2 \le X \le 1) = P\left(\frac{-2-3}{4} \le \frac{X-3}{4} \le \frac{1-3}{4}\right)$$

$$= P(-1.25 \le Z \le -0.50) = \Phi(-0.50) - \Phi(-1.25)$$

$$= ?$$

Example

If $X \sim N(25,36)$, find a constant c such that

$$P(|X-25| \leq c) = 0.9544.$$

Solution :

$$P(|X-25| \leq c) = 0.9544$$

$$\Rightarrow P\left(\frac{-c}{6} \leq \frac{X-25}{6} \leq \frac{c}{6}\right) = 0.9544$$

$$\Rightarrow \Phi\left(\frac{c}{6}\right) - \left[1 - \Phi\left(\frac{c}{6}\right)\right] = 0.9544$$

$$\Rightarrow \Phi\left(\frac{c}{6}\right) = 0.9772$$

$$\Rightarrow \frac{c}{6} = 2 \Rightarrow c = 12$$

Probit function: the inverse

ϕ(area)= Z: gives the Z-value that goes with the probability you want

For example, recall SAT math scores example. What's the score that corresponds to the 90th percentile?

In Table, find the Z-value that corresponds to area of 0.90 → Z= 1.28

If Z=1.28, convert back to raw SAT score →$1.28 = \dfrac{X - 500}{50}$

$X - 500 = 1.28\ (50)$

$X = 1.28(50) + 500 = 564$ (1.28 standard deviations above the mean!)

Are my data "normal"?

- Not all continuous random variables are normally distributed!!
- It is important to evaluate how well the data are approximated by a normal distribution

Are my data normally distributed?

1. Look at the histogram! Does it appear bell shaped?
2. Compute descriptive summary measures—are mean, median, and mode similar?
3. Do 2/3 of observations lie within 1 std dev of the mean? Do 95% of observations lie within 2 std dev of the mean?
4. Look at a normal probability plot—is it approximately linear?
5. Run tests of normality (such as Kolmogorov-Smirnov). But, be cautious, highly influenced by sample size!

THANK YOU

Manipal ProLearn
#7, Service Road, Pragathi Nagar, Electronic City,
Bengaluru 560100
contact@manipalprolearn.com | manipalprolearn.com