Chapter 3: Probability The Basis o Statistical Inference

3.1 Introduction
3.2 Probability
3.3 Elementary Properties of Probability
3.4 Calculating the Probability of an Event

General Definitions and Concepts:

Probability:

Probability is a measure (or number) used to measure the chance of the occurrence of some event. This number is between 0 and 1.

An Experiment:

An experiment is some procedure (or process) that we do.

Sample Space:

The sample space of an experiment is the set of all possible outcomes of an experiment. Also, it is called the universal set, and is denoted by Ω.

An Event:

Any subset of the sample space Ω is called an event.

- $\phi \subseteq \Omega$ is an event (impossible event)
- $\Omega \subseteq \Omega$ is an event (sure event)

Example:

Experiment: Selecting a ball from a box containing 6 balls numbered from 1 to 6 and observing the number on the selected ball.

This experiment has 6 possible outcomes.

The sample space is: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Consider the following events:

E_1 = getting an even number $= \{2, 4, 6\} \subseteq \Omega$

$E_2 =$ getting a number less than 4 $= \{1, 2, 3\} \subseteq \Omega$

$E_3 =$ getting 1 or 3 $= \{1, 3\} \subseteq \Omega$

$E_4 =$ getting an odd number$= \{1, 3, 5\} \subseteq \Omega$

$E_5 =$ getting a negative number$= \{\ \} = \phi \subseteq \Omega$

$E_6 =$ getting a number less than 10$= \{1, 2, 3, 4, 5, 6\} = \Omega \subseteq \Omega$

Notation: $n(\Omega)$=no. of outcomes (elements) in Ω

$n(E)$=no. of outcomes (elements) in the event E

Equally Likely Outcomes:

The outcomes of an experiment are equally likely if the outcomes have the same chance of occurrence.

Probability of An Event:

If the experiment has $n(\Omega)$ equally likely outcomes, then the probability of the event E is denoted by $P(E)$ and is defined by:

$$P(E) = \frac{n(E)}{n(\Omega)} = \frac{no.\ of\ outcomes\ \ in\ \ E}{no.\ of\ outcomes\ \ in\ \ \Omega}$$

Example:

In the ball experiment in the previous example, suppose the ball is selected at random. Determine the probabilities of the following events:

E_1 =getting an even number

$E_2 =$ getting a number less than 4

$E_3 =$ getting 1 or 3

Solution:

$\Omega = \{1, 2, 3, 4, 5, 6\}$; $n(\Omega) = 6$

$E_1 = \{2, 4, 6\}$; $n(E_1) = 3$

$E_2 = \{1, 2, 3\}$; $n(E_2) = 3$

$E_3 = \{1, 3\}$; $n(E_3) = 2$

The outcomes are equally likely.

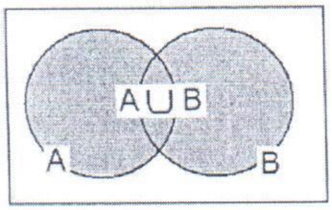
$$\therefore P(E_1) = \frac{3}{6}, \qquad P(E_2) = \frac{3}{6}, \qquad P(E_3) = \frac{2}{6},$$

Some Operations on Events:

Let A and B be two events defined on the sample space Ω.

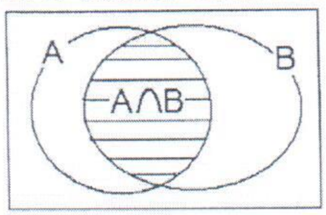
Union of Two events: $(A \cup B)$ or $(A + B)$

The event $A \cup B$ consists of all outcomes in A **or** in B **or** in both A and B. The event $A \cup B$ occurs if A occurs, **or** B occurs, **or** both A and B occur.



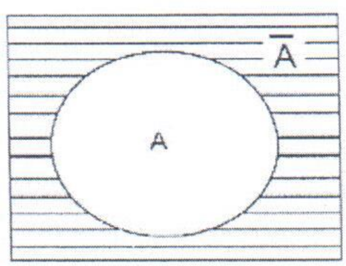
Intersection of Two Events: $(A \cap B)$

The event $A \cap B$ Consists of all outcomes in both A **and** B. The event $A \cap B$ Occurs if both A **and** B occur.



Complement of an Event: $(\overline{A})$ or (A^c) or (A')

The complement of the even A is denoted by $\overline{A}$. The even $\overline{A}$ consists of all outcomes of Ω but are not in A. The even $\overline{A}$ occurs if A does not.



Example:

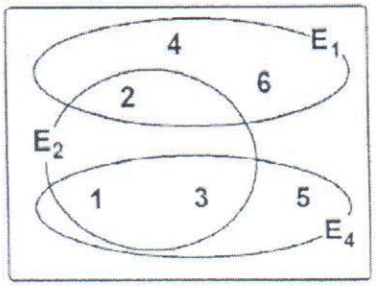
Experiment: Selecting a ball from a box containing 6 balls numbered 1, 2, 3, 4, 5, and 6 randomly.

Define the following events:

$E_1 = \{2, 4, 6\}$ = getting an even number.

$E_2 = \{1, 2, 3\}$ = getting a number $<$ 4.

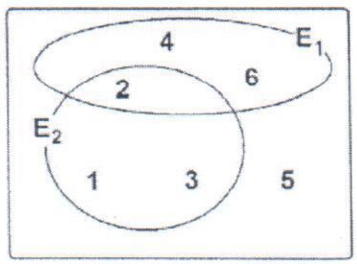
$E_4 = \{1, 3, 5\}$ = getting an odd number.



(1) $E_1 \cup E_2 = \{1, 2, 3, 4, 6\}$

= getting an even number **or** a number less than 4.

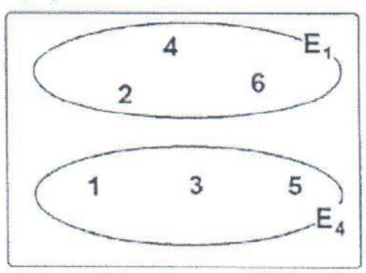
$$P(E_1 \cup E_2) = \frac{n(E_1 \cup E_2)}{n(\Omega)} = \frac{5}{6}$$



(2) $E_1 \cup E_4 = \{1, 2, 3, 4, 5, 6\} = \Omega$

= getting an even number **or** an odd number.

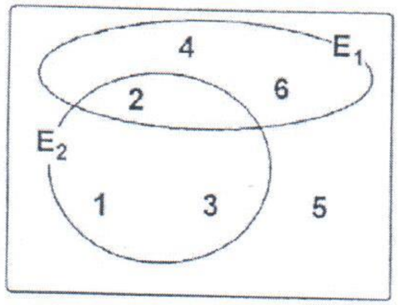
$$P(E_1 \cup E_4) = \frac{n(E_1 \cup E_4)}{n(\Omega)} = \frac{6}{6} = 1$$



Note: $E_1 \cup E_4 = \Omega$. E_1 and E_4 are called exhaustive events. The union of these events gives the whole sample space.

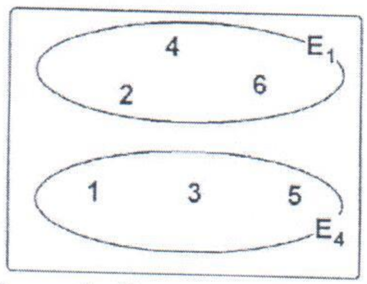
(3) $E_1 \cap E_2 = \{2\}$ = getting an even number **and** a number less than 4.

$$P(E_1 \cap E_2) = \frac{n(E_1 \cap E_2)}{n(\Omega)} = \frac{1}{6}$$



(4) $E_1 \cap E_4 = \phi$ = getting an even number **and** an odd number.

$$P(E_1 \cap E_4) = \frac{n(E_1 \cap E_4)}{n(\Omega)} = \frac{n(\phi)}{6} = \frac{0}{6} = 0$$



Note: $E_1 \cap E_4 = \phi$. E_1 and E_4 are called disjoint (or mutually exclusive) events. These kinds of events can not occurred simultaneously (together in the same time).

(5) The complement of E_1

$\overline{E_1}$ = <u>not</u> getting an even number = $\overline{\{2,4,6\}}$ = $\{1, 3, 5\}$

= getting an odd number.

$= E_4$

<u>Mutually exclusive (disjoint) Events:</u>

The events A and B are disjoint (or mutually exclusive) if:

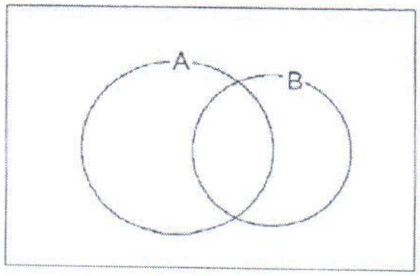
$$A \cap B = \phi .$$

For this case, it is impossible that both events occur simultaneously (i.e., together in the same time). In this case:

(i) $P(A \cap B) = 0$

(ii) $P(A \cup B) = P(A) + P(B)$

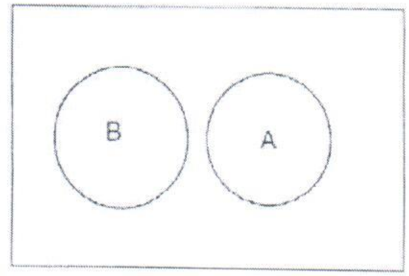
If $A \cap B \neq \phi$, then A and B are not mutually exclusive (not disjoint).



$A \cap B \neq \phi$

A and *B* are not mutually exclusive

(It is possible that both events occur in the same time)



$A \cap B = \phi$

A and *B* are mutually exclusive (disjoint)

(It is impossible that both events occur in the same time)

Exhaustive Events:

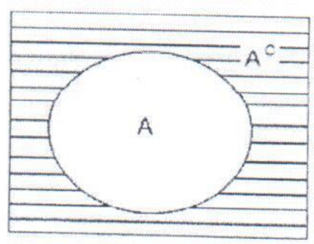
The events $A_1, A_2, \ldots, A_n$ are exhaustive events if:

$$A_1 \cup A_2 \cup \ldots \cup A_n = \Omega .$$

For this case, $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(\Omega) = 1$

Note:

1. $A \cup \overline{A} = \Omega$ (A and $\overline{A}$ are exhaustive events)
2. $A \cap \overline{A} = \phi$ (A and $\overline{A}$ are mutually exclusive (disjoint) events)
3. $n(\overline{A}) = n(\Omega) - n(A)$
4. $P(\overline{A}) = 1 - P(A)$



General Probability Rules:

1. $0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$
3. $P(\phi) = 0$
4. $P(\overline{A}) = 1 - P(A)$

The Addition Rule:

For any two events A and B:

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

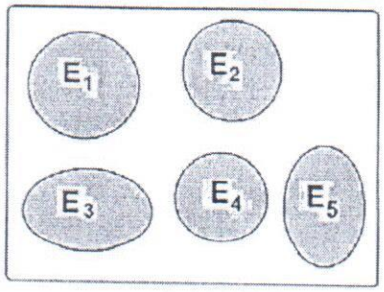
Special Cases:

1. For mutually exclusive (disjoint) events A and B

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

2. For mutually exclusive (disjoint) events $E_1, E_2, \ldots, E_n$:

$$P(E_1 \cup E_2 \cup \ldots \cup E_n) = P(E_1) + P(E_2) + \cdots + P(E_n)$$



Note:

If the events $A_1, A_2, \ldots, A_n$ are exhaustive and mutually exclusive (disjoint) events, then:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = P(\Omega) = 1$$

Marginal Probability:

Given some variable that can be broken down into (m) categories designated by $A_1, A_2, \cdots, A_m$ and another jointly occurring variable that is broken down into (n) categories designated by $B_1, B_2, \cdots, B_n$.

	B_1	B_2	…	B_n	Total
A_1	$n(A_1 \cap B_1)$	$n(A_1 \cap B_2)$	…	$n(A_1 \cap B_n)$	$n(A_1)$
A_2	$n(A_2 \cap B_1)$	$n(A_2 \cap B_2)$	…	$n(A_2 \cap B_n)$	$n(A_2)$
⋮	⋮	⋮	⋮	⋮	⋮
A_m	$n(A_m \cap B_1)$	$n(A_m \cap B_2)$	…	$n(A_m \cap B_n)$	$n(A_m)$
Total	$n(B_1)$	$n(B_2)$	…	$n(B_n)$	$n(\Omega)$

(This table contains the number of elements in each event)

	B_1	B_2	...	B_n	Marginal Probability
A_1	$P(A_1 \cap B_1)$	$P(A_1 \cap B_2)$	...	$P(A_1 \cap B_n)$	$P(A_1)$
A_2	$P(A_2 \cap B_1)$	$P(A_2 \cap B_2)$	...	$P(A_2 \cap B_n)$	$P(A_2)$
. . .	. . .	. . .	. . .	. . .	. . .
A_m	$P(A_m \cap B_1)$	$P(A_m \cap B_2)$	...	$P(A_m \cap B_n)$	$P(A_m)$
Marginal Probability	$P(B_1)$	$P(B_2)$	...	$P(B_n)$	1.00

(This table contains the probability of each event)

The marginal probability of A_i, $P(A_i)$, is equal to the sum of the joint probabilities of A_i with all categories of B. That is:

$$P(A_i) = P(A_i \cap B_1) + P(A_i \cap B_2) + \ldots + P(A_i \cap B_n)$$

$$= \sum_{j=1}^{n} P(A_i \cap B_j)$$

For example,

$$P(A_2) = P(A_2 \cap B_1) + P(A_2 \cap B_2) + \ldots + P(A_2 \cap B_n)$$

$$= \sum_{j=1}^{n} P(A_2 \cap B_j)$$

We define the marginal probability of B_j, $P(B_j)$, in a similar way.

Example:

Table of number of elements in each event:

	B_1	B_2	B_3	Total
A_1	50	30	70	150
A_2	20	70	10	100
A_3	30	100	120	250
Total	100	200	200	500

Table of probabilities of each event:

	B_1	B_2	B_3	Marginal Probability
A_1	0.1	0.06	0.14	0.3
A_2	0.04	0.14	0.02	0.2
A_3	0.06	0.2	0.24	0.5
Marginal Probability	0.2	0.4	0.4	1

For example:

$$P(A_2) = P(A_2 \cap B_1) + P(A_2 \cap B_2) + P(A_2 \cap B_n)$$
$$= 0.04 + 0.14 + 0.02$$
$$= 0.2$$

Applications:

Example:

630 patients are classified as follows:

Blood Type	O (E_1)	A (E_2)	B (E_3)	AB (E_4)	Total
No. of patients	284	258	63	25	630

- Experiment: Selecting a patient at random and observe his/her blood type.
- This experiment has 630 equally likely outcomes $n(\Omega) = 630$

Define the events:

E_1 = The blood type of the selected patient is "O"

E_2 = The blood type of the selected patient is "A"

E_3 = The blood type of the selected patient is "B"

E_4 = The blood type of the selected patient is "AB"

Number of elements in each event:

$n(E_1) = 284$, $n(E_2) = 258$,

$n(E_3) = 63$, $n(E_4) = 25$.

Probabilities of the events:

$P(E_1) = \frac{284}{630} = 0.4508$, $P(E_2) = \frac{258}{630} = 0.4095$,

$P(E_3) = \frac{63}{630} = 0.1$, $P(E_4) = \frac{25}{630} = 0.0397$,

Some operations on the events:

1. $E_2 \cap E_4 =$ the blood type of the selected patients is $''A''$ **and** $''AB''$.

$E_2 \cap E_4 = \phi$ (disjoint events / mutually exclusive events)

$P(E_2 \cap E_4) = P(\phi) = 0$

2. $E_2 \cup E_4 =$ the blood type of the selected patients is $''A''$ **or** $''AB''$

$$P(E_2 \cup E_4) = \begin{cases} \dfrac{n(E_2 \cup E_4)}{n(\Omega)} = \dfrac{258+25}{630} = \dfrac{283}{630} = 0.4492 \\ or \\ P(E_2) + P(E_4) = \dfrac{258}{630} + \dfrac{25}{630} = \dfrac{283}{630} = 0.4492 \end{cases}$$

(since $E_2 \cap E_4 = \phi$)

3. $\overline{E}_1$ = the blood type of the selected patients is not "O".

$n(\overline{E}_1) = n(\Omega) - n(E_1) = 630 - 284 = 346$

$P(\overline{E}_1) = \dfrac{n(\overline{E}_1)}{n(\Omega)} = \dfrac{346}{630} = 0.5492$

another solution:

$P(E_1^C) = 1 - P(E_1) = 1 - 0.4508 = 0.5492$

Notes:

1. E_1, E_2, E_3, E_4 are mutually disjoint, $E_i \cap E_j = \phi \quad (i \neq j)$.
2. E_1, E_2, E_3, E_4 are exhaustive events, $E_1 \cup E_2 \cup E_3 \cup E_4 = \Omega$.

Example:

339 physicians are classified based on their ages and smoking habits as follows.

Smoking Habit

Age		Daily (B_1)	Occasionally (B_2)	Not at all (B_3)	Total
	20 - 29 (A_1)	31	9	7	47
	30 - 39 (A_2)	110	30	49	189
	40 - 49 (A_3)	29	21	29	79
	50+ (A_4)	6	0	18	24
	Total	176	60	103	339

Experiment: Selecting a physician at random

The number of elements of the sample space is $n(\Omega) = 339$.

The outcomes of the experiment are equally likely.

Some events:

- A_3 = the selected physician is aged 40 - 49

$$P(A_3)=\frac{n(A_3)}{n(\Omega)}=\frac{79}{339}=0.2330$$

- $B_2 =$ the selected physician smokes occasionally

$$P(B_2)=\frac{n(B_2)}{n(\Omega)}=\frac{60}{339}=0.1770$$

- $A_3 \cap B_2 =$ the selected physician is aged 40-49 **and** smokes occasionally.

$$P(A_3 \cap B_2)=\frac{n(A_3 \cap B_2)}{n(\Omega)}=\frac{21}{339}=0.06195$$

- $A_3 \cup B_2 =$ the selected physician is aged 40-49 **or** smokes occasionally (**or** both)

$$P(A_3 \cup B_2)=P(A_3)+P(B_2)-P(A_3 \cap B_2)$$
$$=\frac{79}{339}+\frac{60}{339}-\frac{21}{339}$$
$$=0.233+0.177-0.06195=0.3481$$

- $\overline{A}_4 =$ the selected physician is **not** 50 years or older.
  $= A_1 \cup A_2 \cup A_3$

$$P(\overline{A}_4)=1-P(A_4)$$
$$=1-\frac{n(A_4)}{n(\Omega)}=1-\frac{24}{339}=0.9292$$

- $A_2 \cup A_3 =$ the selected physician is aged 30-39 **or** is aged 40-49
  $=$ the selected physician is aged 30-49

$$P(A_2 \cup A_3)=\frac{n(A_2 \cup A_3)}{n(\Omega)}=\frac{189+79}{339}=\frac{268}{339}=0.7906$$

or

$$P(A_2 \cup A_3)=P(A_2)+P(A_3)=\frac{189}{339}+\frac{79}{339}=0.7906$$

(Since $A_2 \cap A_3 = \phi$)

Example:

Suppose that there is a population of pregnant women with:

- 10% of the pregnant women delivered prematurely.
- 25% of the pregnant women used some sort of medication.

- 5% of the pregnant women delivered prematurely and used some sort of medication.

Experiment: Selecting a woman randomly from this population.
Define the events:

- D = The selected woman delivered prematurely.
- M = The selected women used medication.
- $D \cap M$ = The selected woman delivered prematurely and used some sort of medication.

Percentages:

$\%(D) = 10\%$ $\quad\quad$ $\%(M) = 25\%$ $\quad\quad$ $\%(D \cap M) = 5\%$

The complement events:

$\overline{D}$ = The selected woman did not deliver prematurely.
$\overline{M}$ = The selected women did not use medication.

A Two-way table: (Percentages given by a two-way table):

	M	$\overline{M}$	Total
D	**5**	?	**10**
$\overline{D}$	?	?	?
Total	**25**	?	**100**

	M	$\overline{M}$	Total
D	**5**	5	**10**
$\overline{D}$	20	70	90
Total	**25**	75	**100**

The probabilities of the given events are:

$$P(D) = \frac{\%(D)}{100\%} = \frac{10\%}{100\%} = 0.1$$

$$P(M) = \frac{\%(M)}{100\%} = \frac{25\%}{100\%} = 0.25$$

$$P(D \cap M) = \frac{\%(D \cap M)}{100\%} = \frac{5\%}{100\%} = 0.05$$

Calculating probabilities of some events:

$D \cup M$ = the selected woman delivered prematurely or used medication.

$$P(D \cup M) = P(D) + (M) - P(D \cap M)$$
$$= 0.1 + 0.25 - 0.05 = 0.3 \quad\quad \text{(by the rule)}$$

$\overline{M}$ = The selected woman did not use medication

$P(\overline{M}) = 1 - P(M) = 1 - 0.25 = 0.75$ (by the rule)

$P(\overline{M}) = \frac{75}{100} = 0.75$ (from the table)

$\overline{D}$ = The selected woman did not deliver prematurely

$P(\overline{D}) = 1 - P(D) = 1 - 0.10 = 0.90$ (by the rule)

$P(\overline{D}) = \frac{90}{100} = 0.90$ (from the table)

$\overline{D} \cap \overline{M}$ = the selected woman did not deliver prematurely and did not use medication.

$P(\overline{D} \cap \overline{M}) = \frac{70}{100} = 0.70$ (from the table)

$\overline{D} \cap M$ = the selected woman did not deliver prematurely and used medication.

$P(\overline{D} \cap M) = \frac{20}{100} = 0.20$ (from the table)

$D \cap \overline{M}$ = the selected woman delivered prematurely and did not use medication.

$P(D \cap \overline{M}) = \frac{5}{100} = 0.05$ (from the table)

$D \cup \overline{M}$ = the selected woman delivered prematurely or did not use medication.

$$P(D \cup \overline{M}) = P(D) + (\overline{M}) - P(D \cap \overline{M})$$
$$= 0.1 + 0.75 - 0.05 = 0.8$$ (by the rule)

$\overline{D} \cup M$ = the selected woman did not deliver prematurely or used medication.

$$P(\overline{D} \cup M) = P(\overline{D}) + (M) - P(\overline{D} \cap M)$$
$$= 0.9 + 0.25 - 0.20 = 0.95$$ (by the rule)

$\overline{D} \cup \overline{M}$ = the selected woman did not deliver prematurely or did not use medication.

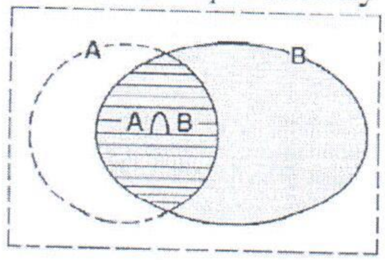
$$P(\overline{D} \cup \overline{M}) = P(\overline{D}) + (\overline{M}) - P(\overline{D} \cap \overline{M})$$
$$= 0.9 + 0.75 - 0.70 = 0.95$$ (by the rule)

Conditional Probability:

- The conditional probability of the event A when we know that the event B has already occurred is defined by:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad ; P(B) \neq 0$$

- P(A | B) = The conditional probability of *A* given *B*.



Notes:

(1) $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{n(A \cap B)/n(\Omega)}{n(B)/n(\Omega)} = \dfrac{n(A \cap B)}{n(B)}$

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

(3) For calculating $P(A \mid B)$, we may use any one of the following:

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

(ii) $P(A \mid B) = \dfrac{n(A \cap B)}{n(B)}$

(iii) Using the restricted table directly.

Multiplication Rules of Probability:

For any two events A and B, we have:

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

Example:

Smoking Habit

Age	Daily (B_1)	Occasionally (B_2)	Not at all (B_3)	Total
20-29 (A_1)	31	9	7	47
30-39 (A_2)	110	30	49	189
40-49 (A_3)	29	21	29	79
50+ (A_4)	6	0	18	24
Total	176	60	103	339

Consider the following event:

$(B_1 \mid A_2)$ = the selected physician smokes daily given that his age is between 30 and 39

- $P(B_1)=\frac{n(B_1)}{n(\Omega)}=\frac{176}{339}=0.519$
- $P(B_1 \mid A_2)=\frac{P(B_1 \cap A_2)}{P(A_2)}$

$$=\frac{0.324484}{0.557522}=0.5820$$

$$\begin{cases} P(B_1 \cap A_2)=\frac{n(B_1 \cap A_2)}{n(\Omega)}=\frac{110}{339}=0.324484 \\ P(A_2)=\frac{n(A_2)}{n(\Omega)}=\frac{189}{339}=0.557522 \end{cases}$$

another solution:

$$P(B_1 \mid A_2)=\frac{n(B_1 \cap A_2)}{n(A_2)}=\frac{110}{189}=0.5820$$

Notice that:

$P(B_1)=0.519$

$P(B_1 \mid A_2)=0.5820$

$P(B_1 \mid A_2) > P(B_1)$!! ... $P(B_1) \neq P(B_1 \mid A_2)$

What does this mean?

We will answer this question after talking about the concept of independent events.

Example: (Multiplication Rule of Probability)

A training health program consists of two consecutive parts. To pass this program, the trainee must pass both parts of the program. From the past experience, it is known that 90% of the trainees pass the first part, and 80% of those who pass the first part pass the second part. If you are admitted to this program, what is the probability that you will pass the program? What is the percentage of trainees who pass the program?

Solution:

Define the following events:

A = the event of passing the first part

B = the event of passing the second part

A∩B = the event of passing the first part and the second Part

= the event of passing both parts

= the event of passing the program

Therefore, the probability of passing the program is P(A∩B).

From the given information:

The probability of passing the first part is:

$$P(A) = 0.9 \qquad \left(\frac{90\%}{100\%} = 0.9\right)$$

The probability of passing the second part given that the trainee has already passed the first part is:

$$P(B|A) = 0.8 \qquad \left(\frac{80\%}{100\%} = 0.8\right)$$

Now, we use the multiplication rule to find P(A∩B) as follows:

$$P(A \cap B) = P(A)\,P(B|A) = (0.9)(0.8) = 0.72$$

We can conclude that 72% of the trainees pass the program.

Independent Events

There are 3 cases:

- $P(A\,|\,B) > P(A)$
  (knowing B increases the probability of occurrence of A)
- $P(A\,|\,B) < P(A)$
  (knowing B decreases the probability of occurrence of A)
- $P(A\,|\,B) = P(A)$
  (knowing B has no effect on the probability of occurrence of A). In this case A is independent of B.

Independent Events:

- Two events A and B are independent if one of the following conditions is satisfied:

$$(i)\ \ P(A\,|\,B) = P(A)$$
$$\Leftrightarrow (ii)\ \ P(B\,|\,A) = P(B)$$
$$\Leftrightarrow (iii)\ \ P(B \cap A) = P(A)P(B)$$

Note: The third condition is the multiplication rule of independent events.

Example:

Suppose that A and B are two events such that:

$P(A) = 0.5$, $P(B)=0.6$, $P(A \cap B)=0.2$.

Theses two events are not independent (they are dependent) because:

$P(A)\ P(B) = 0.5 \times 0.6 = 0.3$

$P(A \cap B)=0.2$.

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

Also, $P(A)= 0.5 \neq P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.2}{0.6} = 0.3333$.

Also, $P(B) = 0.6 \neq P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.2}{0.5} = 0.4$.

For this example, we may calculate probabilities of all events. We can use a two-way table of the probabilities as follows:

	B	$\overline{B}$	Total
A	**0.2**	?	**0.5**
$\overline{A}$	?	?	?
Total	**0.6**	?	**1.00**

We complete the table:

	B	$\overline{B}$	Total
A	**0.2**	0.3	**0.5**
$\overline{A}$	0.4	0.1	0.5
Total	**0.6**	0.4	**1.00**

$P(\overline{A}) = 0.5$

$P(\overline{B}) = 0.4$

$P(A \cap \overline{B}) = 0.3$

$P(\overline{A} \cap B) = 0.4$

$P(\overline{A} \cap \overline{B}) = 0.1$

$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.6 - 0.2 = 0.9$

$P(A \cup \overline{B}) = P(A) + P(\overline{B}) - P(A \cap \overline{B}) = 0.5 + 0.4 - 0.3 = 0.6$

$P(\overline{A} \cup B) = exercise$

$P(\overline{A} \cup \overline{B}) = exercise$

Note: The Addition Rule for Independent Events:

If the events A and B are independent, then

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad \text{(Addition rule)}$$
$$= P(A) + P(B) - P(A)\,P(B)$$

Example: (Reading Assignment)

Suppose that a dental clinic has 12 nurses classified as follows:

Nurse	1	2	3	4	5	6	7	8	9	10	11	12
Has children	Yes	No	No	No	No	Yes	No	No	Yes	No	No	No
Works at night	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes

The experiment is to randomly choose one of these nurses. Consider the following events:

C = the chosen nurse has children

N = the chosen nurse works night shift

a) Find The probabilities of the following events:
   1. the chosen nurse has children.
   2. the chosen nurse works night shift.
   3. the chosen nurse has children and works night shift.
   4. the chosen nurse has children and does not work night shift.

b) Find the probability of choosing a nurse who woks at night given that she has children.

c) Are the events C and N independent? Why?

d) Are the events C and N disjoint? Why?

e) Sketch the events C and N with their probabilities using Venn diagram.

Solution:

We can classify the nurses as follows:

	N (Night shift)	$\overline{N}$ (No night shift)	total
C (Has Children)	2	1	3
$\overline{C}$ (No Children)	6	3	9
total	8	4	12

a) The experiment has $n(\Omega) = 12$ equally likely outcomes.

P(The chosen nurse has children) = P(C) = $\frac{n(C)}{n(\Omega)} = \frac{3}{12} = 0.25$

P(The chosen nurse works night shift) = P(N) = $\frac{n(N)}{n(\Omega)} = \frac{8}{12} = 0.6667$

P(The chosen nurse has children and works night shift)

= P(C∩N)= $\frac{n(C \cap N)}{n(\Omega)} = \frac{2}{12} = 0.16667$

P(The chosen nurse has children and does not work night shift)

$$= P(C\cap\overline{N}) = \frac{n(C\cap\overline{N})}{n(\Omega)} = \frac{1}{12} = 0.0833$$

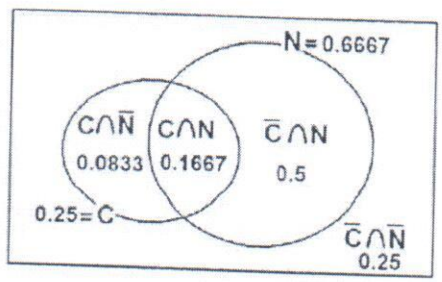
b) The probability of choosing a nurse who woks at night given that she has children:

$$P(N\mid C) = \frac{P(C\cap N)}{P(C)} = \frac{2/12}{0.25} = 0.6667$$

c) The events C and N are independent because $P(N\mid C)=P(N)$.

d) The events C and N not are disjoint because $C\cap N\neq\phi$.
(Note: $n(C\cap N)=2$)

e) Venn diagram



3.5 Bayes' Theorem, Screening Tests, Sensitivity, Specificity, and Predictive Value Positive and Negative: (pp.79-83)

There are two states regarding the disease and two states regarding the result of the screening test:

State of the Disease: **Present (D)** / **Absent ($\overline{D}$)**

Result of the Test: **Positive (T)** / **Negative ($\overline{T}$)**

We define the following events of interest:

D : the individual has the disease (presence of the disease)

$\overline{D}$: the individual does not have the disease (absence of The disease)

T : the individual has a positive screening test result

$\overline{T}$: the individual has a negative screening test result

- There are 4 possible situations:

		True status of the disease	
		+ve (D: Present)	−ve ($\overline{D}$:Absent)
Result of the test	+ve (T)	Correct diagnosing	false positive result
	−ve ($\overline{T}$)	false negative result	Correct diagnosing

Definitions of False Results:

There are two false results:

1. **A false positive** result:
   This result happens when a test indicates a positive status when the true status is negative. Its probability is:
   $P(T \mid \overline{D})$ = P(positive result | absence of the disease)

2. **A false negative** result:
   This result happens when a test indicates a negative status when the true status is positive. Its probability is:
   $P(\overline{T} \mid D)$ = P(negative result | presence of the disease)

Definitions of the Sensitivity and Specificity of the test:

1. **The Sensitivity:**

The sensitivity of a test is the probability of a positive test result given the presence of the disease.

$P(T \mid D)$= P(positive result of the test | presence of the disease)

2. The specificity:

The specificity of a test is the probability of a negative test result given the absence of the disease.

$P(\overline{T} \mid \overline{D})$ = P(negative result of the test | absence of the disease)

To clarify these concepts, suppose we have a sample of (n) subjects who are cross-classified according to Disease Status and Screening Test Result as follows:

	Disease		
Test Result	**Present (D)**	**Absent ($\overline{D}$)**	**Total**
Positive (T)	a	b	a + b = n(T)
Negative ($\overline{T}$)	c	d	c + d = n($\overline{T}$)
Total	a + c = n(D)	b + d = n($\overline{D}$)	n

For example, there are (a) subjects who have the disease and whose screening test result was positive.

From this table we may compute the following conditional probabilities:

1. The probability of false positive result:

$$P(T \mid \overline{D}) = \frac{n(T \cap \overline{D})}{n(\overline{D})} = \frac{b}{b+d}$$

2. The probability of false negative result:

$$P(\overline{T} \mid D) = \frac{n(\overline{T} \cap D)}{n(D)} = \frac{c}{a+c}$$

3. The sensitivity of the screening test:

$$P(T \mid D) = \frac{n(T \cap D)}{n(D)} = \frac{a}{a+c}$$

4. The specificity of the screening test:

$$P(\overline{T} \mid \overline{D}) = \frac{n(\overline{T} \cap \overline{D})}{n(\overline{D})} = \frac{d}{b+d}$$

Definitions of the Predictive Value Positive and Predictive Value Negative of a Screening Test:

1. The predictive value positive of a screening test:
The predictive value positive is the probability that a subject has the disease, given that the subject has a positive screening test result:

$$P(D \mid T) = \text{P(the subject has the disease | positive result)}$$
$$= \text{P(presence of the disease | positive result)}$$

2. The predictive value negative of a screening test:
The predictive value negative is the probability that a subject does not have the disease, given that the subject has a negative screening test result:

$$P(\overline{D} \mid \overline{T}) = \text{P(the subject does not have the disease | negative result)}$$
$$= \text{P(absence of the disease | negative result)}$$

Calculating the Predictive Value Positive and Predictive Value Negative:

(How to calculate $P(D \mid T)$ and $P(\overline{D} \mid \overline{T})$):

We calculate these conditional probabilities using the knowledge of:

1. The sensitivity of the test = $P(T \mid D)$
2. The specificity of the test = $P(\overline{T} \mid \overline{D})$
3. The probability of the relevant disease in the general population, P(D). (It is usually obtained from another independent study)

Calculating the Predictive Value Positive, $P(D \mid T)$:

$$P(D \mid T) = \frac{P(T \cap D)}{P(T)}$$

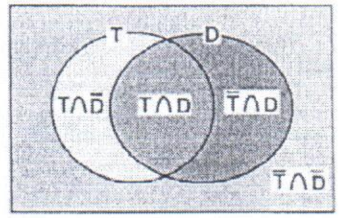
But we know that:

$$P(T) = P(T \cap D) + P(T \cap \overline{D})$$

$$P(T \cap D) = P(T \mid D)\, P(D) \quad \text{(multiplication rule)}$$

$$P(T \cap \overline{D}) = P(T \mid \overline{D})\, P(\overline{D}) \quad \text{(multiplication rule)}$$

$$P(T) = P(T \mid D)\, P(D) + P(T \mid \overline{D})\, P(\overline{D})$$



Therefore, we reach the following version of Bayes' Theorem:

$$P(D \mid T) = \frac{P(T \mid D)\, P(D)}{P(T \mid D)\, P(D) + P(T \mid \overline{D})\, P(\overline{D})} \quad \ldots\ldots\ldots\ldots (1)$$

Note:

$P(T \mid D)$ = sensitivity.

$P(T \mid \overline{D}) = 1 - P(\overline{T} \mid \overline{D})$ = 1 − specificity.

$P(D)$ = The probability of the relevant disease in the general population.

$P(\overline{D}) = 1 - P(D)$.

Calculating the Predictive Value Negative, $P(\overline{D} \mid \overline{T})$:

To obtain the predictive value negative of a screening test, we use the following statement of Bayes' theorem:

$$P(\overline{D} \mid \overline{T}) = \frac{P(\overline{T} \mid \overline{D})\, P(\overline{D})}{P(\overline{T} \mid \overline{D})\, P(\overline{D}) + P(\overline{T} \mid D)\, P(D)} \quad \ldots\ldots\ldots\ldots (2)$$

Note:

$P(\overline{T} \mid \overline{D})$ = specificity.

$P(\overline{T} \mid D) = 1 - P(T \mid D)$ = 1 − sensitivity.

Example:

A medical research team wished to evaluate a proposed screening test for Alzheimer's disease. The test was given to a random sample of 450 patients with Alzheimer's disease and an independent random sample of 500 patients without symptoms of the disease. The two samples were drawn from populations of subjects who were 65 years of age or older. The results are as follows:

	Alzheimer Disease		
Test Result	**Present (D)**	**Absent ($\bar{D}$)**	**Total**
Positive (T)	436	5	441
Negative ($\bar{T}$)	14	495	509
Total	450	500	950

Based on another independent study, it is known that the percentage of patients with Alzheimer's disease (the rate of prevalence of the disease) is 11.3% out of all subjects who were 65 years of age or older.

Solution:

Using these data we estimate the following quantities:

1. The sensitivity of the test:

$$P(T \mid D) = \frac{n(T \cap D)}{n(D)} = \frac{436}{450} = 0.9689$$

2. The specificity of the test:

$$P(\bar{T} \mid \bar{D}) = \frac{n(\bar{T} \cap \bar{D})}{n(\bar{D})} = \frac{495}{500} = 0.99$$

3. The probability of the disease in the general population, P(D): The rate of disease in the relevant general population, P(D), cannot be computed from the sample data given in the table. However, it is given that the percentage of patients with Alzheimer's disease is 11.3% out of all subjects who were 65 years of age or older. Therefore P(D) can be computed to be:

$$P(D) = \frac{11.3\,\%}{100\,\%} = 0.113$$

4. The predictive value positive of the test:
We wish to estimate the probability that a subject who is positive on the test has Alzheimer disease. We use the Bayes' formula of Equation (1):

$$P(D \mid T) = \frac{P(T \mid D)\,P(D)}{P(T \mid D)\,P(D) + P(T \mid \bar{D})\,P(\bar{D})}.$$

From the tabulated data we compute:

$$P(T \mid D) = \frac{436}{450} = 0.9689 \quad \text{(From part no. 1)}$$

$$P(T \mid \overline{D}) = \frac{n(T \cap \overline{D})}{n(\overline{D})} = \frac{5}{500} = 0.01$$

Substituting of these results into Equation (1), we get:

$$P(D \mid T) = \frac{(0.9689)\,P(D)}{(0.9689)\,P(D) + (0.01)\,P(\overline{D})}$$

$$= \frac{(0.9689)\,(0.113)}{(0.9689)\,(0.113) + (0.01)\,(1 - 0.113)} = 0.93$$

As we see, in this case, the predictive value positive of the test is very high.

5. The predictive value negative of the test:

We wish to estimate the probability that a subject who is negative on the test does not have Alzheimer disease. We use the Bayes' formula of Equation (2):

$$P(\overline{D} \mid \overline{T}) = \frac{P(\overline{T} \mid \overline{D})\,P(\overline{D})}{P(\overline{T} \mid \overline{D})\,P(\overline{D}) + P(\overline{T} \mid D)\,P(D)}$$

To compute $P(\overline{D} \mid \overline{T})$, we first compute the following probabilities:

$$P(\overline{T} \mid \overline{D}) = \frac{495}{500} = 0.99 \quad \text{(From part no. 2)}$$

$$P(\overline{D}) = 1 - P(D) = 1 - 0.113 = 0.887$$

$$P(\overline{T} \mid D) = \frac{n(\overline{T} \cap D)}{n(D)} = \frac{14}{450} = 0.0311$$

Substitution in Equation (2) gives:

$$P(\overline{D} \mid \overline{T}) = \frac{P(\overline{T} \mid \overline{D})\,P(\overline{D})}{P(\overline{T} \mid \overline{D})\,P(\overline{D}) + P(\overline{T} \mid D)\,P(D)}$$

$$= \frac{(0.99)(0.887)}{(0.99)(0.887) + (0.0311)(0.113)}$$

$$= 0.996$$

As we see, the predictive value negative is also very high.

CHAPTER 4: Probabilistic Features of Certain Data Distribution (Probability Distributions)

4.1 Introduction:

The concept of random variables is very important in Statistics. Some events can be defined using random variables.

There are two types of random variables:

$$\text{Random variables} \begin{cases} \textit{Discrete Random Variables} \\ \textit{Continuous Random Variables} \end{cases}$$

4.2 Probability Distributions of Discrete Random Variables:

Definition:

The probability distribution of a discrete random variable is a table, graph, formula, or other device used to specify all possible values of the random variable along with their respective probabilities.

Examples of discrete r v.'s

- The no. of patients visiting KKUH in a week.
- The no. of times a person had a cold in last year.

Example:

Consider the following discrete random variable.

X = The number of times a Saudi person had a cold in January 2010.

Suppose we are able to count the no. of Saudis which $X = x$:

x (no. of colds a Saudi person had in January 2010)	Frequency of x (no. of Saudi people who had a cold x times in January 2010)
0	10,000,000
1	3,000,000
2	2,000,000
3	1,000,000
Total	$N = 16{,}000{,}000$

Note that the possible values of the random variable X are:

$$x = 0, 1, 2, 3$$

Experiment: Selecting a person at random

Define the event:

$(X = 0)$ = The event that the selected person had no cold.
$(X = 1)$ = The event that the selected person had 1 cold.
$(X = 2)$ = The event that the selected person had 2 colds.
$(X = 3)$ = The event that the selected person had 3 colds.

In general:

$(X = x)$ =The event that the selected person had x colds.

For this experiment, there are $n(\Omega) = 16{,}000{,}000$ equally likely outcomes.

The number of elements of the event (X = x) is:

n(X=x) = no. of Saudi people who had a cold x times in January 2010.
= frequency of x.

The probability of the event (X = x) is:

$$P(X = x) = \frac{n(X = x)}{n(\Omega)} = \frac{n(X = x)}{16000000} \quad \text{, for x=0, 1, 2, 3}$$

x	freq. of x $n(X = x)$	$P(X = x) = \frac{n(X = x)}{16000000}$ (Relative frequency)
0	10000000	0.6250
1	3000000	0.1875
2	2000000	0.1250
3	1000000	0.0625
Total	16000000	1.0000

Note:

$$P(X = x) = \frac{n(X = x)}{16000000} = \text{Relative Frequency} = \frac{frequency}{16000000}$$

The probability distribution of the discrete random variable X is given by the following table:

x	$P(X=x)=f(x)$
0	0.6250
1	0.1874
2	0.1250
3	0.0625
Total	1.0000

Notes:

- The probability distribution of any discrete random variable X must satisfy the following two properties:
  (1) $0 \le P(X=x) \le 1$
  (2) $\sum_x P(X=x)=1$
- Using the probability distribution of a discrete r.v. we can find the probability of any event expressed in term of the r.v. X.

Example:

Consider the discrete r.v. X in the previous example.

x	$P(X=x)$
0	0.6250
1	0.1875
2	0.1250
3	0.0625
Total	1.0000

(1) $P(X \ge 2)=P(X=2)+P(X=3)=0.1250+0.0625=0.1875$

(2) $P(X>2)=P(X=3)=0.0625$ [note: $P(X>2) \ne P(X \ge 2)$]

(3) $P(1 \le X<3)=P(X=1)+P(X=2)=0.1875+0.1250=0.3125$

(4) $P(X \le 2)=P(X=0)+P(X=1)+P(X=2)$
$=0.6250+0.1875+0.1250=0.9375$

another solution:

$P(X \le 2)=1-P((\overline{X \le 2}))$
$=1-P(X>2)=1-P(X=3)=1-0.625=0.9375$

(5) $P(-1 \le X<2)=P(X=0)+P(X=1)$
$=0.6250+0.1875=0.8125$

(6) $P(-1.5 \leq X < 1.3) = P(X=0) + P(X=1)$
$= 0.6250 + 0.1875 = 0.8125$

(7) $P(X=3.5) = P(\phi) = 0$

(8) $P(X \leq 10) = P(X=0) + P(X=1) + P(X=2) + P(X=3) = P(\Omega) = 1$

(9) The probability that the selected person had at least 2 cold:

$$P(X \geq 2) = P(X=2) + P(X=3) = 0.1875$$

(10) The probability that the selected person had at most 2 colds:

$$P(X \leq 2) = 0.9375$$

(11) The probability that the selected person had more than 2 colds:

$$P(X > 2) = P(X=3) = 0.0625$$

(12) The probability that the selected person had less than 2 colds:

$$P(X < 2) = P(X=0) + P(X=1) = 0.8125$$

Graphical Presentation:

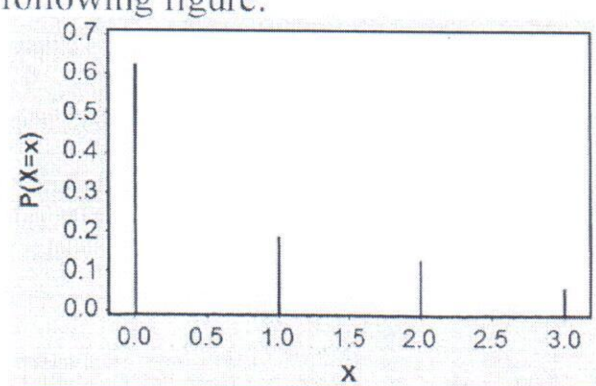
The probability distribution of a discrete r. v. X can be graphically represented.

Example:

The probability distribution of the random variable in the previous example is:

x	$P(X=x)$
0	0.6250
1	0.1875
2	0.1250
3	0.0625

The graphical presentation of this probability distribution is given by the following figure:



Mean and Variance of a Discrete Random Variable

Mean: The mean (or expected value) of a discrete random variable X is denoted by μ or μ_X. It is defined by:

$$\mu = \sum_x x\, P(X = x)$$

Variance: The variance of a discrete random variable X is denoted by σ^2 or σ_X^2. It is defined by:

$$\sigma^2 = \sum_x (x-\mu)^2 P(X = x)$$

Example:

We wish to calculate the mean μ and the variance of the discrete r. v. X whose probability distribution is given by the following table:

x	$P(X=x)$
0	0.05
1	0.25
2	0.45
3	0.25

Solution:

x	$P(X=x)$	$xP(X=x)$	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 P(X=x)$
0	0.05	0	-1.9	3.61	0.1805
1	0.25	0.25	-0.9	0.81	0.2025
2	0.45	0.9	0.1	0.01	0.0045
3	0.25	0.75	1.1	1.21	0.3025
Total		$\mu = \sum x\, P(X=x) = 1.9$			$\sigma^2 = \sum (x-\mu)^2 P(X=x) = 0.69$

$$\mu = \sum_x x\; P(X = x) = (0)(0.05) + (1)(0.25) + (2)(0.45) + (3)(0.25) = 1.9$$

$$\sigma^2 = \sum_x (x-1.9)^2\, P(X = x)$$

$$= (0-1.9)^2(0.05) + (1-1.9)^2(0.25) + (2-1.9)^2(0.45) + (3-1.9)^2(0.25)$$

$$= 0.69$$

Cumulative Distributions:

The cumulative distribution function of a discrete r. v. X is defined by:

$$P(X \le x) = \sum_{a \le x} P(X = a) \qquad \text{(Sum over all values } \le x\text{)}$$

Example:

Calculate the cumulative distribution of the discrete r. v. X whose probability distribution is given by the following table:

x	$P(X=x)$
0	0.05
1	0.25
2	0.45
3	0.25

Use the cumulative distribution to find:

P(X≤2), P(X<2), P(X≤1.5), P(X<1.5), P(X>1), P(X≥1)

Solution:

The cumulative distribution of X is:

x	$P(X \le x)$	
0	0.05	$P(X \le 0) = P(X=0)$
1	0.30	$P(X \le 1) = P(X=0) + P(X=1)$
2	0.75	$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$
3	1.0000	$P(X \le 3) = P(X=0) + \cdots + P(X=3)$

Using the cumulative distribution,

P(X≤2) = 0.75

P(X<2) = P(X≤1) = 0.30

P(X≤1.5) = P(X≤1) = 0.30

P(X<1.5) = P(X≤1) = 0.30

P(X>1) = 1- P($\overline{(X>1)}$) = 1-P(X≤1) = 1- 0.30 = 0.70

P(X≥1) = 1- P($\overline{(X \ge 1)}$) = 1-P(X<1) = 1- P(X≤0)

= 1- 0.05 = 0.95

Example: (Reading Assignment)

Given the following probability distribution of a discrete random variable X representing the number of defective teeth of the patient visiting a certain dental clinic:

x	P(X = x)
1	0.25
2	0.35
3	0.20
4	0.15
5	K

a) Find the value of K.
b) Find the flowing probabilities:
   1. $P(X < 3)$
   2. $P(X \leq 3)$
   3. $P(X < 6)$
   4. $P(X < 1)$
   5. $P(X = 3.5)$
c) Find the probability that the patient has at least 4 defective teeth.
d) Find the probability that the patient has at most 2 defective teeth.
e) Find the expected number of defective teeth (mean of X).
f) Find the variance of X.

Solution:

a) $1 = \sum P(X = x) = 0.25 + 0.35 + 0.20 + 0.15 + K$

$1 = 0.95 + K$

$K = 1 - 0.95$

$K = 0.05$

The probability distribution of X is:

x	P(X = x)
1	0.25
2	0.35
3	0.20
4	0.15
5	0.05
Total	1.00

b) Finding the probabilities:
   1. $P(X < 3) = P(X=1)+P(X=2) = 0.25+0.35 = 0.60$
   2. $P(X \leq 3) = P(X=1)+P(X=2)+P(X=3) = 0.8$
   3. $P(X < 6) = P(X=1)+P(X=2)+P(X=3)+P(X=4)+P(X=5)= P(\Omega)=1$
   4. $P(X < 1) = P(\phi)=0$
   5. $P(X = 3.5) = P(\phi)=0$

c) The probability that the patient has at least 4 defective teeth

$P(X\geq4) = P(X=4)+P(X=5) =0.15+0.05=0.2$

d) The probability that the patient has at most 2 defective teeth

$P(X\leq2) = P(X=1)+P(X=2) = 0.25+0.35=0.6$

e) The expected number of defective teeth (mean of X)

x	P(X = x)	x P(X = x)
1	0.25	0.25
2	0.35	0.70
3	0.20	0.60
4	0.15	0.60
5	0.05	0.25
Total	$\sum P(X=x)=1$	$\mu=\sum x\,P(X=x)=2.4$

The expected number of defective teeth (mean of X) is

$\mu=\sum x\,P(X=x)=(1)(0.25)+(2)(0.35)+(3)(0.2)+(4)(0.15)+(5)(0.05)=2.4$

f) The variance of X:

x	$P(X=x)$	$(x-\mu)$	$(x-\mu)^2$	$(x-\mu)^2 P(X=x)$
1	0.25	-1.4	1.96	0.49
2	0.35	-0.4	0.16	0.056
3	0.20	0.6	0.36	0.072
4	0.15	1.6	2.56	0.384
5	0.05	2.6	6.76	0.338
Total				$\sigma^2=\sum(x-\mu)^2 P(X=x)=1.34$

The variance is $\sigma^2=\sum(x-\mu)^2 P(X=x)=1.34$

Combinations:

Notation (n!):

$n!$ is read "n factorial". It defined by:

$$n! = n(n-1)(n-2)\cdots(2)(1) \qquad for \quad n \geq 1$$

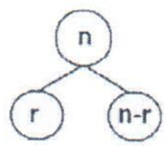
$$0! = 1$$

Example: $5! = (5)(4)(3)(2)(1) = 120$

Combinations:

The number of different ways for selecting r objects from n distinct objects is denoted by ${}_nC_r$ or $\binom{n}{r}$ and is given by:

$$_nC_r = \frac{n!}{r!\ (n-r)!}; \qquad for \quad r = 0, 1, 2, \ldots, n$$



Notes:

1. ${}_nC_r$ is read as " n " choose " r ".
2. ${}_nC_n = 1$, ${}_nC_0 = 1$,
3. ${}_nC_r = {}_nC_{n-r}$ (for example: ${}_{10}C_3 = {}_{10}C_7$)
4. ${}_nC_r$ = number of unordered subsets of a set of (n) objects such that each subset contains (r) objects.

Example:

For n = 4 and r = 2:

$$_4C_2 = = \frac{4!}{2!\ (4-2)!} = \frac{4!}{2! \times 2!} = \frac{4 \times 3 \times 2 \times 1}{(2 \times 1) \times (2 \times 1)} = 6$$

${}_4C_2 = 6$ = The number of different ways for selecting 2 objects from 4 distinct objects.

Example:

Suppose that we have the set {a, b, c, d} of (n=4) objects.
We wish to choose a subset of two objects. The possible subsets of this set with 2 elements in each subset are:

{a , b}, {a , c}, {a , d}, {b , d}, {b , c}, {c , d}

The number of these subsets is ${}_4C_2 = 6$.

4.3 Binomial Distribution:

- **Bernoulli Trial**: is an experiment with only two possible outcomes: S = success and F= failure (Boy or girl, Saudi or non-Saudi, sick or well, dead or alive).
- Binomial distribution is a discrete distribution.
- Binomial distribution is used to model an experiment for which:
  1. The experiment has a sequence of n Bernoulli trials.
  2. The probability of success is $P(S) = p$, and the probability of failure is $P(F) = 1 - p = q$.
  3. The probability of success $P(S) = p$ is constant for each trial.
  4. The trials are independent; that is the outcome of one trial has no effect on the outcome of any other trial.

In this type of experiment, we are interested in the discrete r. v. representing the number of successes in the n trials.

X = The number of successes in the n trials

The possible values of X (number of success in n trails) are:

x = 0, 1, 3, … , n

The r.v. X has a binomial distribution with parameters n and p , and we write:

$$X \sim Binomial(n, p)$$

The probability distribution of X is given by:

$$P(X = x) = \begin{cases} {}_nC_x \; p^x \; q^{n-x} & for\, x = 0, 1, 2, \ldots, n \\ 0 & otherwise \end{cases}$$

Where: $${}_nC_x = \frac{n!}{x! \; (n-x)!}$$

We can write the probability distribution of X as a table as follows.

x	$P(X = x)$
0	${}_nC_0 \; p^0 \; q^{n-0} = q^n$
1	${}_nC_1 \; p^1 \; q^{n-1}$

x	$P(X=x)$
2	${}_nC_2\ p^2\ q^{n-2}$
$\vdots$	$\vdots$
$n-1$	${}_nC_{n-1}\ p^{n-1}\ q^1$
n	${}_nC_n\ p^n\ q^0 = p^n$
Total	1.00

Result: (Mean and Variance for normal distribution)
If $X\sim$ Binomial(n, p) , then

- The mean: $\mu = np$ (expected value)
- The variance: $\sigma^2 = npq$

Example:

Suppose that the probability that a Saudi man has high blood pressure is 0.15. Suppose that we randomly select a sample of 6 Saudi men.

(1) Find the probability distribution of the random variable (X) representing the number of men with high blood pressure in the sample.

(2) Find the expected number of men with high blood pressure in the sample (mean of X).

(3) Find the variance X.

(4) What is the probability that there will be exactly 2 men with high blood pressure?

(5) What is the probability that there will be at most 2 men with high blood pressure?

(6) What is the probability that there will be at lease 4 men with high blood pressure?

Solution:

We are interested in the following random variable:

X = The number of men with high blood pressure in the sample of 6 men.

Notes:

- Bernoulli trial: diagnosing whether a man has a high blood pressure or not. There are two outcomes for each trial:

S = Success: The man has high blood pressure
F = failure: The man does not have high blood pressure.

- Number of trials = 6 (we need to check 6 men)
- Probability of success: $P(S) = p = 0.15$
- Probability of failure: $P(F) = q = 1 - p = 0.85$
- Number of trials: $n = 6$
- The trials are independent because of the fact that the result of each man does not affect the result of any other man since the selection was made ate random.

The random variable X has a binomial distribution with parameters: n=6 and p=0.15, that is:

$$X \sim \text{Binomial (n, p)}$$
$$X \sim \text{Binomial (6, 0.15)}$$

The possible values of X are:

$$x = 0, 1, 3, 4, 5, 6$$

(1) The probability distribution of X is:

$$P(X = x) = \begin{cases} {}_6C_x \ (0.15)^x (0.85)^{6-x} & ; x = 0, 1, 2, 3, 4, 5, 6 \\ 0 & ; \ otherwise \end{cases}$$

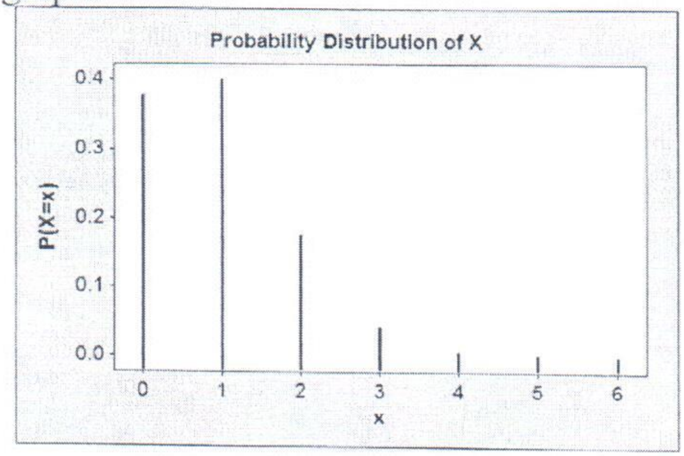
The probabilities of all values of X are:

$$P(X = 0) = {}_6C_0 \ (0.15)^0 (0.85)^6 = (1)(0.15)^0 (0.85)^6 = 0.37715$$
$$P(X = 1) = {}_6C_1 \ (0.15)^1 (0.85)^5 = (6)(0.15)(0.85)^5 = 0.39933$$
$$P(X = 2) = {}_6C_2 \ (0.15)^2 (0.85)^4 = (15)(0.15)^2 (0.85)^4 = 0.17618$$
$$P(X = 3) = {}_6C_3 \ (0.15)^3 (0.85)^3 = (20)(0.15)^3 (0.85)^3 = 0.04145$$
$$P(X = 4) = {}_6C_4 \ (0.15)^4 (0.85)^2 = (15)(0.15)^4 (0.85)^2 = 0.00549$$
$$P(X = 5) = {}_6C_5 \ (0.15)^5 (0.85)^1 = (6)(0.15)^5 (0.85)^1 = 0.00039$$
$$P(X = 6) = {}_6C_6 \ (0.15)^6 (0.85)^0 = (1)(0.15)^6 (1) = 0.00001$$

The probability distribution of X can by presented by the following table:

x	$P(X=x)$
0	0.37715
1	0.39933
2	0.17618
3	0.04145
4	0.00549
5	0.00039
6	0.00001

The probability distribution of X can by presented by the following graph:



(2) The mean of the distribution (the expected number of men out of 6 with high blood pressure) is:

$$\mu = np = (6)(0.15) = 0.9$$

(3) The variance is:

$$\sigma^2 = npq = (6)(0.15)(0.85) = 0.765$$

(4) The probability that there will be exactly 2 men with high blood pressure is:

$$P(X=2) = 0.17618$$

(5) The probability that there will be at most 2 men with high blood pressure is:

$$\begin{aligned} P(X \le 2) &= P(X=0) + P(X=1) + P(X=2) \\ &= 0.37715 + 0.39933 + 0.17618 \\ &= 0.95266 \end{aligned}$$

(6) The probability that there will be at lease 4 men with high blood pressure is:

$$P(X \geq 4) = P(X=4) + P(X=5) + P(X=6)$$
$$= 0.00549 + 0.00039 + 0.00001$$
$$= 0.00589$$

Example: (Reading Assignment)

Suppose that 25% of the people in a certain population have low hemoglobin levels. The experiment is to choose 5 people at random from this population. Let the discrete random variable X be the number of people out of 5 with low hemoglobin levels.

a) Find the probability distribution of X.
b) Find the probability that at least 2 people have low hemoglobin levels.
c) Find the probability that at most 3 people have low hemoglobin levels.
d) Find the expected number of people with low hemoglobin levels out of the 5 people.
e) Find the variance of the number of people with low hemoglobin levels out of the 5 people.

Solution:

X = the number of people out of 5 with low hemoglobin levels
The Bernoulli trail is the process of diagnosing the person
Success = the person has low hemoglobin
Failure = the person does not have low hemoglobin

$n = 5$ (no. of trials)
$p = 0.25$ (probability of success)
$q = 1 - p = 0.75$ (probability of failure)

a) X has a binomial distribution with parameter $n = 5$ and $p = 0.25$

$$X \sim Binomial(n, p)$$
$$X \sim Binomial(5, 0.25)$$

The possible values of X are:

x=0, 1, 2, 3, 4, 5

The probability distribution is:

$$P(X = x) = \begin{cases} {}_nC_x\ p^x\ q^{n-x}\ ; & for\ x = 0, 1, 2, \ldots, n \\ 0\ ; & otherwise \end{cases}$$

$$P(X = x) = \begin{cases} {}_5C_x\ (0.25)^x\ (0.75)^{5-x}\ ; & for\ x = 0, 1, 2, 3, 4, 5 \\ 0\ ; & otherwise \end{cases}$$

x	P(X = x)
0	${}_5C_0 \times 0.25^0 \times 0.75^{5-0} = 0.23730$

x	P(X = x)
1	${}_5C_1 \times 0.25^1 \times 0.75^{5-1} = 0.39551$
2	${}_5C_2 \times 0.25^2 \times 0.75^{5-2} = 0.26367$
3	${}_5C_3 \times 0.25^3 \times 0.75^{5-3} = 0.08789$
4	${}_5C_4 \times 0.25^4 \times 0.75^{5-4} = 0.01465$
5	${}_5C_5 \times 0.25^5 \times 0.75^{5-5} = 0.00098$
Total	$\sum P(X = x) = 1$

b) The probability that at least 2 people have low hemoglobin levels:

$P(X\geq 2) = P(X=2)+P(X=3)+P(X=4)+P(X=5)$
$= 0.26367+ 0.08789+ 0.01465+ 0.00098$
$= 0. 0.36719$

c) The probability that at most 3 people have low hemoglobin levels:

$P(X\leq 3) = P(X=0)+P(X=1)+P(X=2)+P(X=3)$
$= 0.23730+ 0.39551+ 0.26367+ 0.08789$
$= 0.98437$

d) The expected number of people with low hemoglobin levels out of the 5 people (the mean of X):

$$\mu = n\,p = 5\times 0.25 = 1.25$$

e) The variance of the number of people with low hemoglobin levels out of the 5 people (the variance of X) is:

$$\sigma^2 = n\,pq = 5\times 0.25\times 0.75 = 0.9375$$

4.4 The Poisson Distribution:

- It is a discrete distribution.
- The Poisson distribution is used to model a discrete r. v. representing the number of occurrences of some random event in an interval of time or space (or some volume of matter).
- The possible values of X are:

  x= 0, 1, 2, 3, …
- The discrete r. v. X is said to have a Poisson distribution with parameter (average or mean) λ if the probability distribution of X is given by

$$P(X=x)=\begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & ; \quad for \quad x=0,1,2,3,\ldots \\ 0 & ; \quad otherwise \end{cases}$$

where e = 2.71828 (the natural number).
We write :

$$X \sim \text{Poisson }(\lambda)$$

Result: (Mean and Variance of Poisson distribution)
If $X \sim$ Poisson (λ), then:

- The mean (average) of X is : $\mu=\lambda$ (Expected value)
- The variance of X is: $\sigma^2=\lambda$

Example:
Some random quantities that can be modeled by Poisson distribution:

- No. of patients in a waiting room in an hours.
- No. of surgeries performed in a month.
- No. of rats in each house in a particular city.

Note:

- λ is the average (mean) of the distribution.
- If X = The number of patients seen in the emergency unit in a day, and if X ~Poisson (λ), then:
    1. The average (mean) of patients seen every day in the emergency unit = λ.
    2. The average (mean) of patients seen every month in the emergency unit =30λ.
    3. The average (mean) of patients seen every year in the emergency unit = 365λ.
    4. The average (mean) of patients seen every hour in the emergency unit = λ/24.

Also, notice that:
(i) If Y = The number of patients seen every month, then:

$Y \sim$ Poisson (λ^*), where $\lambda^*=30\lambda$

(ii) W = The number of patients seen every year, then:

$W \sim$ Poisson (λ^*), where $\lambda^*=365\lambda$

(iii) V = The number of patients seen every hour, then:

$V \sim$ Poisson (λ^*), where $\lambda^*=\frac{\lambda}{24}$

Example:

Suppose that the number of snake bites cases seen at KKUH in a year has a Poisson distribution with average 6 bite cases.

(1) What is the probability that in a year:

(i) The no. of snake bite cases will be 7?

(ii) The no. of snake bite cases will be less than 2?

(2) What is the probability that there will be 10 snake bite cases in 2 years?

(3) What is the probability that there will be no snake bite cases in a month?

Solution:

(1) X = no. of snake bite cases in a year.

$X \sim$ Poisson (6) $\qquad (\lambda=6)$

$$P(X=x)=\frac{e^{-6}6^x}{x!}\,;\; x=0,1,2,\ldots$$

(i) $$P(X=7)=\frac{e^{-6}6^7}{7!}=0.13768$$

(ii) $$P(X<2)=P(X=0)+P(X=1)$$
$$=\frac{e^{-6}6^0}{0!}+\frac{e^{-6}6^1}{1!}=0.00248+0.01487=0.01735$$

(2) Y = no of snake bite cases in 2 years

$Y \sim$ Poisson(12) $\qquad (\lambda^*=2\lambda=(2)(6)=12)$

$$P(Y=y)=\frac{e^{-12}12^y}{y!}\,;\quad y=0,1,2\ldots$$

$$\therefore P(Y=10)=\frac{e^{-12}12^{10}}{10!}=0.1048$$

(3) W = no. of snake bite cases in a month.

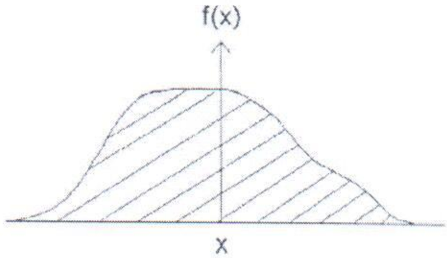
$W \sim$ Poisson (0.5) $\qquad (\lambda^{**}=\frac{\lambda}{12}=\frac{6}{12}=0.5)$

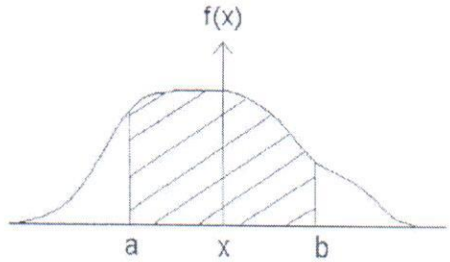
$$P(W = w) = \frac{e^{-0.5}\ 0.5^{w}}{w!}: \qquad w = 0\ , 1\ , 2 \ldots$$

$$P(W = 0) = \frac{e^{-0.5}(0.5)^{0}}{0!} = 0.6065$$

4.5 Continuous Probability Distributions:

For any continuous r. v. X, there exists a function $f(x)$, called the probability density function (pdf) of X , for which:
(1) The total area under the curve of $f(x)$ equals to 1.

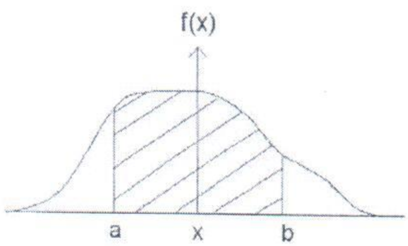


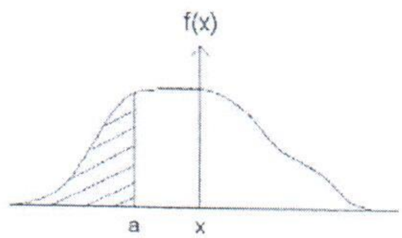


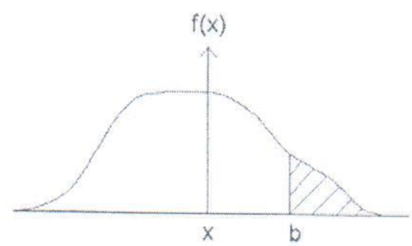
Total area= $\int_{-\infty}^{\infty} f(x)\,dx = 1$ $\qquad$ $P(a \le X \le b) = \int_a^b f(x)dx = area$

(2) The probability hat X is between the points (a) and (b) equals to the area under the curve of f(x) which is bounded by the point a and b.

(3) In general, the probability of an interval event is given by the area under the curve of $f(x)$ and above that interval.







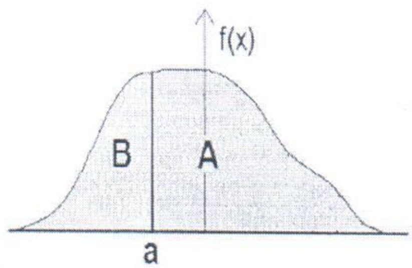
$P(a \le X \le b) = \int_a^b f(x)dx = area$ $\qquad$ $P(X \le a) = \int_{-\infty}^{a} f(x)dx = area$ $\qquad$ $P(X \ge b) = \int_b^{\infty} f(x)\,dx = area$

Note:

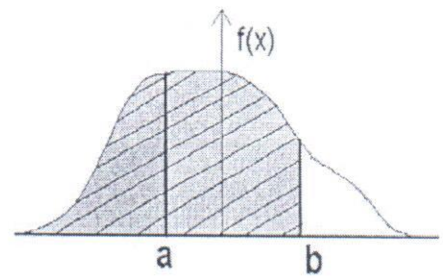
If X is continuous r.v. then:

1. $P(X = a) = 0$ for any a.
2. $P(X \le a) = P(X < a)$

3. $P(X \geq b) = P(X > b)$
4. $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$
5. $P(X \leq x) =$ cumulative probability
6. $P(X \geq a) = 1 - P(X < a) = 1 - P(X \leq a)$
7. $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$



$P(X \geq a) = 1 - P(X \leq a)$
$A = 1 - B$
Total area = 1



$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$

$$\int_a^b f(x)dx = \int_{-\infty}^b f(x)dx - \int_{-\infty}^a f(x)dx$$

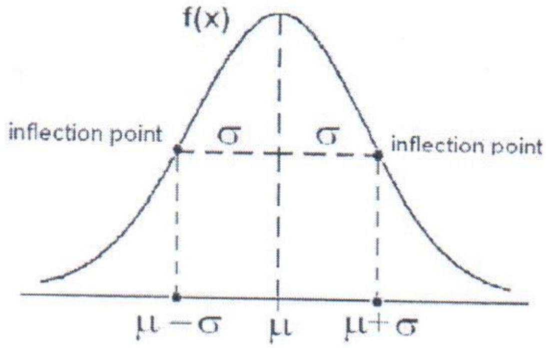
4.6 The Normal Distribution:

- One of the most important continuous distributions.
- Many measurable characteristics are normally or approximately normally distributed.
 (Examples: height, weight, ...)
- The probability density function of the normal distribution is given by:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} \quad ; -\infty < x < \infty$$

where (e=2.71828) and (π=3.14159).
The parameters of the distribution are the mean (μ) and the standard deviation (σ).

- The continuous r.v. X which has a normal distribution has several important characteristics:
  1. $-\infty < X < \infty$,
  2. The density function of X, $f(x)$, has a bell-Shaped curve:

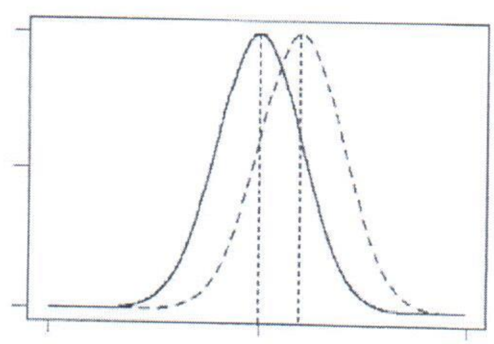


mean = μ
standard deviation = σ
variance = σ^2

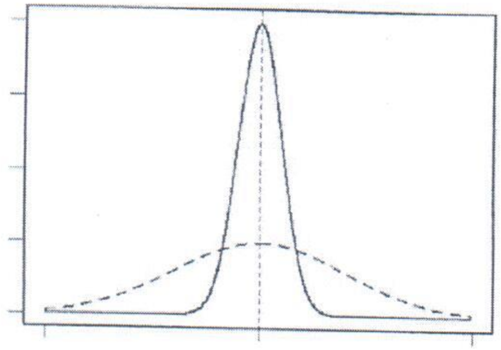
3. The highest point of the curve of f(x) at the mean μ. (Mode = μ)
4. The curve of f(x) is symmetric about the mean μ.
   μ = mean = mode = median
5. The normal distribution depends on two parameters:
   mean = μ (determines the location)
   standard deviation = σ (determines the shape)
6. If the r.v. X is normally distributed with mean μ and standard deviation σ (variance σ^2), we write:
   $X \sim \text{Normal}(\mu, \sigma^2)$ or $X \sim N(\mu, \sigma^2)$
7. The location of the normal distribution depends on μ. The shape of the normal distribution depends on σ.

Note: The location of the normal distribution depends on μ and its shape depends on σ.
Suppose we have two normal distributions:

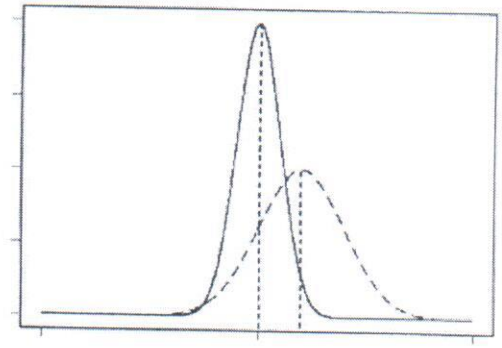
———— $N(\mu_1, \sigma_1)$
---------- $N(\mu_2, \sigma_2)$



$\mu_1 < \mu_2, \sigma_1 = \sigma_2$



$\mu_1 = \mu_2, \sigma_1 < \sigma_2$



$\mu_1 < \mu_2, \sigma_1 < \sigma_2$

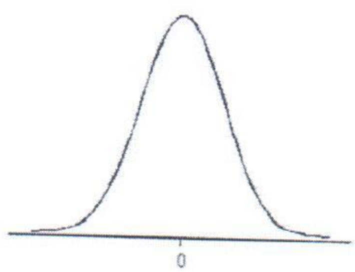
The Standard Normal Distribution:

The normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$ is called the standard normal distribution and is denoted by Normal (0,1) or N(0,1). The standard normal random variable is denoted by (Z), and we write:

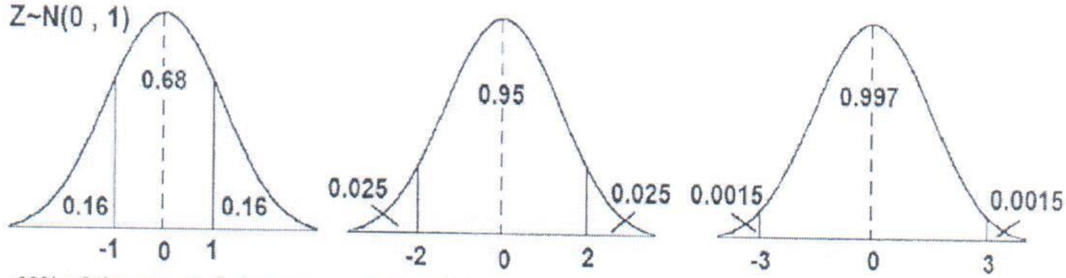
$$Z \sim N(0, 1)$$

The probability density function (pdf) of Z~N(0,1) is given by:

$$f(z) = n(z;0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$



The standard normal distribution, Normal (0,1), is very important because probabilities of any normal distribution can be calculated from the probabilities of the standard normal distribution.



68% of the area is between -1 and 1 (approximately)

95% of the area is between -2 and 2 (approximately)

99.7% of the area is between -3 and 3 (approximately)

Result:

If $X \sim \text{Normal}\left(\mu, \sigma^2\right)$, then $Z = \dfrac{X-\mu}{\sigma} \sim \text{Normal}\ (0,1)$.

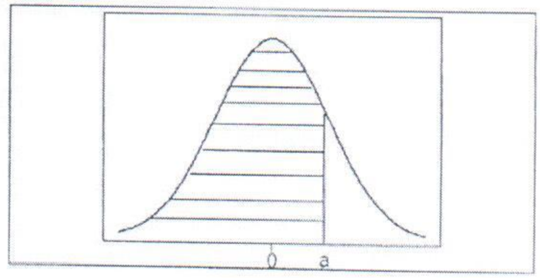
Calculating Probabilities of Normal (0,1):

Suppose Z ~ Normal (0,1).

For the standard normal distribution Z ~ N(0,1), there is a special table used to calculate probabilities of the form:

$$P(Z \le a)$$

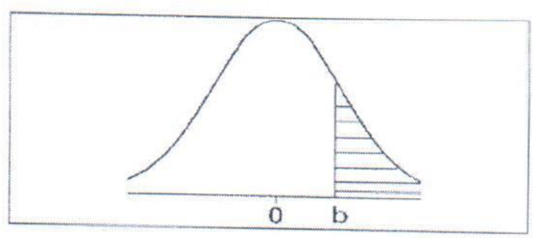
(i) $P(Z \le a) =$ From the table



(ii) $P(Z \ge b) = 1 - P(Z \le b)$

Where:

$P(Z \le b)$ = From the table

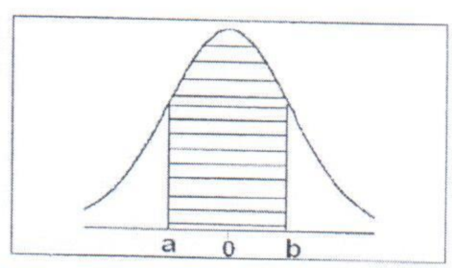


(iii) $P(a \le Z \le b) = P(Z \le b) - P(z \le a)$

Where:

$P(Z \le b)$ = from the table

$P(z \le a)$ = from the table

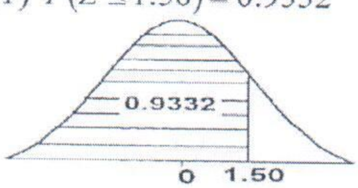


(iv) $P(Z = a) = 0$ for every a .

Example:

Suppose that $Z \sim N(0,1)$

(1) $P(Z \leq 1.50) = 0.9332$



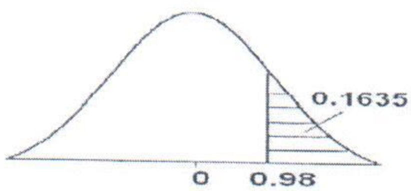
Z	0.00	0.01	...
:	⇓		
1.50 ⇒	0.9332		
:			

(2)

$P(Z \geq 0.98) = 1 - P(Z \leq 0.98)$

$= 1 - 0.8365$

$= 0.1635$



Z	0.00	...	0.08
:	:	:	⇓
:	...	...	⇓
0.90⇒	⇒	⇒	0.8365

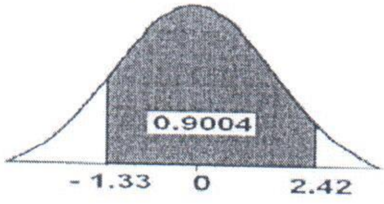
(3)

$P(-1.33 \leq Z \leq 2.42) =$

$P(Z \leq 2.42) - P(Z \leq -1.33)$

$= 0.9922 - 0.0918$

$= 0.9004$

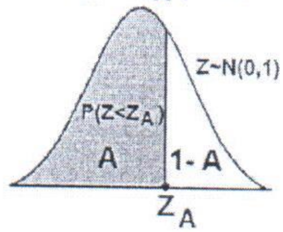


Z	...		−0.03
:	:		⇓
−1.30	⇒		0.0918
:			

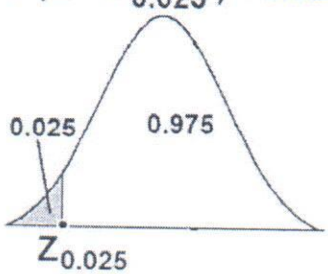
(4) $P(Z \leq 0) = P(Z \geq 0) = 0.5$

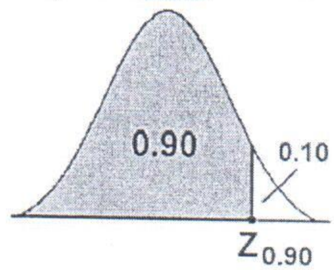
Notation:

$P(Z \leq Z_A) = A$



For example:





Result:

Since the pdf of Z~N(0,1) is symmetric about 0, we have:

$$Z_A = -Z_{1-A}$$

For example: $Z_{0.35} = -Z_{1-0.35} = -Z_{0.65}$

$Z_{0.86} = -Z_{1-0.86} = -Z_{0.14}$

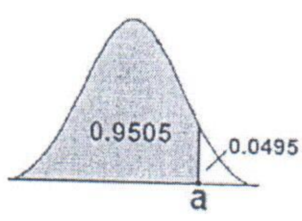
Example:

Suppose that $Z \sim N(0,1)$.

If $P(Z \leq a) = 0.9505$

Then $a = 1.65$

Z	…	0.05	…
⋮		⇑	
1.60	⇐	0.9505	
⋮			



$P(Z < a) = 0.9505$

$P(Z < Z_{0.9505}) = 0.9505$

$a = Z_{0.9505}$

$a = Z_{0.9505} = 1.65$

Example:

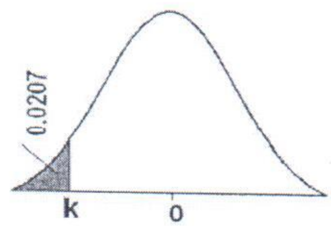
Suppose that $Z \sim N(0,1)$. Find the value of k such that $P(Z \le k) = 0.0207$.

Solution:

$k = -2.04$

Notice that $k = Z_{0.0207} = -2.04$

Z	...	-0.04	
:	:	⇑ ⇑	
-2.0	⇐⇐	0.0207	
:			



Example:

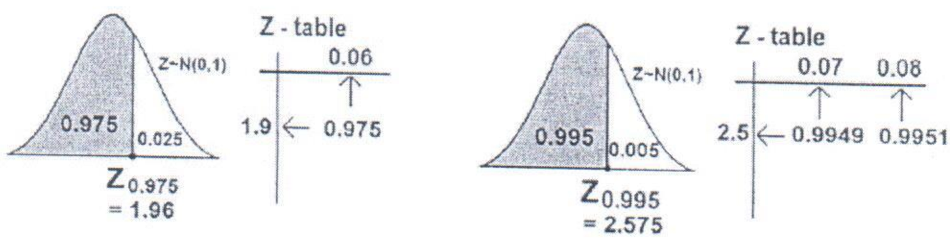
If $Z \sim N(0,1)$, then:

$Z_{0.90} = 1.285$

$Z_{0.95} = 1.645$

$Z_{0.975} = 1.96$

$Z_{0.99} = 2.325$



Using the result: $Z_A = -Z_{1-A}$

$Z_{0.10} = -Z_{0.90} = -1.285$

$Z_{0.05} = -Z_{0.95} = -1.645$

$Z_{0.025} = -Z_{0.975} = -1.96$

$Z_{0.01} = -Z_{0.99} = -2.325$

<u>Calculating Probabilities of Normal (μ, σ^2):</u>

- Recall the result:

$$X \sim \text{Normal}(\mu, \sigma^2) \quad \Leftrightarrow \quad Z = \frac{X-\mu}{\sigma} \sim \text{Normal}(0,1)$$

- $X \le a \Leftrightarrow \frac{X-\mu}{\sigma} \le \frac{a-\mu}{\sigma} \Leftrightarrow Z \le \frac{a-\mu}{\sigma}$

1. $P(X \le a) = P\left(Z \le \frac{a-\mu}{\sigma}\right)$ = From the table.
2. $P(X \ge a) = 1 - P(X \le a) = 1 - P\left(Z \le \frac{a-\mu}{\sigma}\right)$
3. $P(a \le X \le b) = P(X \le b) - P(X \le a)$
   $= P\left(Z \le \frac{b-\mu}{\sigma}\right) - P\left(Z \le \frac{a-\mu}{\sigma}\right)$
4. $P(X = a) = 0$, for every a.

4.7 Normal Distribution Application:

Example

Suppose that the hemoglobin levels of healthy adult males are approximately normally distributed with a mean of 16 and a variance of 0.81.

(a) Find that probability that a randomly chosen healthy adult male has a hemoglobin level less than 14.

(b) What is the percentage of healthy adult males who have hemoglobin level less than 14?

(c) In a population of 10,000 healthy adult males, how many would you expect to have hemoglobin level less than 14?

Solution:

X = hemoglobin level for healthy adults males

Mean: $\mu = 16$

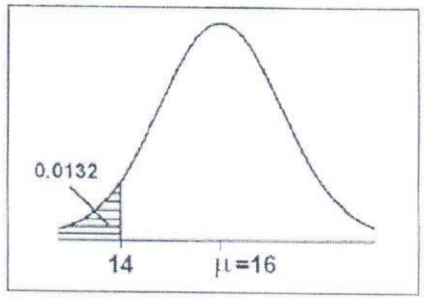
Variance: $\sigma^2 = 0.81$

Standard deviation: $\sigma = 0.9$

$X \sim$ Normal (16, 0.81)

(a) The probability that a randomly chosen healthy adult male has hemoglobin level less than 14 is $P(X \le 14)$.

$$P(X \le 14) = P\left(Z \le \frac{14-\mu}{\sigma}\right)$$
$$= P\left(Z \le \frac{14-16}{0.9}\right)$$
$$= P(Z \le -2.22)$$
$$= 0.0132$$



(b) The percentage of healthy adult males who have hemoglobin level less than 14 is:

$$P(X \le 14) \times 100\% = 0.0132 \times 100\% = 1.32\%$$

(c) In a population of 10000 healthy adult males, we would expect that the number of males with hemoglobin level less than 14 to be:

$$P(X \le 14) \times 10000 = 0.0132 \times 10000 = 132 \ males$$

Example:

Suppose that the birth weight of Saudi babies has a normal distribution with mean μ=3.4 and standard deviation σ=0.35.

(a) Find the probability that a randomly chosen Saudi baby has a birth weight between 3.0 and 4.0 kg.

(b) What is the percentage of Saudi babies who have a birth weight between 3.0 and 4.0 kg?

(c) In a population of 100000 Saudi babies, how many would you expect to have birth weight between 3.0 and 4.0 kg?

Solution:

X = birth weight of Saudi babies

Mean: μ = 3.4

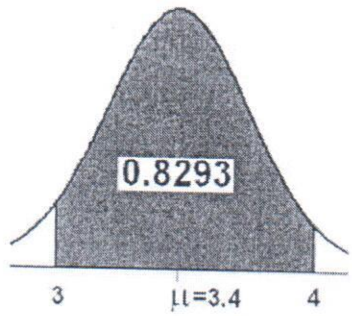
Standard deviation: σ = 0.35

Variance: $\sigma^2 = (0.35)^2 = 0.1225$

X ~ Normal (3.4, 0.1225)

(a) The probability that a randomly chosen Saudi baby has a birth weight between 3.0 and 4.0 kg is $P(3.0 < X < 4.0)$

$$P(3.0 < X < 4.0) = P(X \leq 4.0) - P(X \leq 3.0)$$
$$= P\left(Z \leq \frac{4.0-\mu}{\sigma}\right) - P\left(Z \leq \frac{3.0-\mu}{\sigma}\right)$$
$$= P\left(Z \leq \frac{4.0-3.4}{0.35}\right) - P\left(Z \leq \frac{3.0-3.4}{0.35}\right)$$
$$= P(Z \leq 1.71) - P(Z \leq -1.14)$$
$$= 0.9564 - 0.1271 = 0.8293$$



(b) The percentage of Saudi babies who have a birth weight between 3.0 and 4.0 kg is

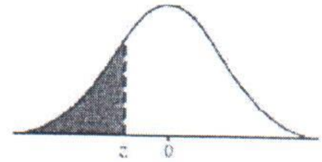
$P(3.0<X<4.0) \times 100\% = 0.8293 \times 100\% = 82.93\%$

(c) In a population of 100,000 Saudi babies, we would expect that the number of babies with birth weight between 3.0 and 4.0 kg to be:

$P(3.0<X<4.0) \times 100000 = 0.8293 \times 100000 = 82930$ babies

Standard Normal Table

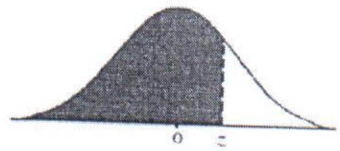
Areas Under the Standard Normal Curve



z	-0.09	-0.08	-0.07	-0.06	-0.05	-0.04	-0.03	-0.02	-0.01	-0.00	z
-3.50	0.00017	0.00017	0.00018	0.00019	0.00019	0.00020	0.00021	0.00022	0.00022	0.00023	-3.50
-3.40	0.00024	0.00025	0.00026	0.00027	0.00028	0.00029	0.00030	0.00031	0.00032	0.00034	-3.40
-3.30	0.00035	0.00036	0.00038	0.00039	0.00040	0.00042	0.00043	0.00045	0.00047	0.00048	-3.30
-3.20	0.00050	0.00052	0.00054	0.00056	0.00058	0.00060	0.00062	0.00064	0.00066	0.00069	-3.20
-3.10	0.00071	0.00074	0.00076	0.00079	0.00082	0.00084	0.00087	0.00090	0.00094	0.00097	-3.10
-3.00	0.00100	0.00104	0.00107	0.00111	0.00114	0.00118	0.00122	0.00126	0.00131	0.00135	-3.00
-2.90	0.00139	0.00144	0.00149	0.00154	0.00159	0.00164	0.00169	0.00175	0.00181	0.00187	-2.90
-2.80	0.00193	0.00199	0.00205	0.00212	0.00219	0.00226	0.00233	0.00240	0.00248	0.00256	-2.80
-2.70	0.00264	0.00272	0.00280	0.00289	0.00298	0.00307	0.00317	0.00326	0.00336	0.00347	-2.70
-2.60	0.00357	0.00368	0.00379	0.00391	0.00402	0.00415	0.00427	0.00440	0.00453	0.00466	-2.60
-2.50	0.00480	0.00494	0.00508	0.00523	0.00539	0.00554	0.00570	0.00587	0.00604	0.00621	-2.50
-2.40	0.00639	0.00657	0.00676	0.00695	0.00714	0.00734	0.00755	0.00776	0.00798	0.00820	-2.40
-2.30	0.00842	0.00866	0.00889	0.00914	0.00939	0.00964	0.00990	0.01017	0.01044	0.01072	-2.30
-2.20	0.01101	0.01130	0.01160	0.01191	0.01222	0.01255	0.01287	0.01321	0.01355	0.01390	-2.20
-2.10	0.01426	0.01463	0.01500	0.01539	0.01578	0.01618	0.01659	0.01700	0.01743	0.01786	-2.10
-2.00	0.01831	0.01876	0.01923	0.01970	0.02018	0.02068	0.02118	0.02169	0.02222	0.02275	-2.00
-1.90	0.02330	0.02385	0.02442	0.02500	0.02559	0.02619	0.02680	0.02743	0.02807	0.02872	-1.90
-1.80	0.02938	0.03005	0.03074	0.03144	0.03216	0.03288	0.03362	0.03438	0.03515	0.03593	-1.80
-1.70	0.03673	0.03754	0.03836	0.03920	0.04006	0.04093	0.04182	0.04272	0.04363	0.04457	-1.70
-1.60	0.04551	0.04648	0.04746	0.04846	0.04947	0.05050	0.05155	0.05262	0.05370	0.05480	-1.60
-1.50	0.05592	0.05705	0.05821	0.05938	0.06057	0.06178	0.06301	0.06426	0.06552	0.06681	-1.50
-1.40	0.06811	0.06944	0.07078	0.07215	0.07353	0.07493	0.07636	0.07780	0.07927	0.08076	-1.40
-1.30	0.08226	0.08379	0.08534	0.08691	0.08851	0.09012	0.09176	0.09342	0.09510	0.09680	-1.30
-1.20	0.09853	0.10027	0.10204	0.10383	0.10565	0.10749	0.10935	0.11123	0.11314	0.11507	-1.20
-1.10	0.11702	0.11900	0.12100	0.12302	0.12507	0.12714	0.12924	0.13136	0.13350	0.13567	-1.10
-1.00	0.13786	0.14007	0.14231	0.14457	0.14686	0.14917	0.15151	0.15386	0.15625	0.15866	-1.00
-0.90	0.16109	0.16354	0.16602	0.16853	0.17106	0.17361	0.17619	0.17879	0.18141	0.18406	-0.90
-0.80	0.18673	0.18943	0.19215	0.19489	0.19766	0.20045	0.20327	0.20611	0.20897	0.21186	-0.80
-0.70	0.21476	0.21770	0.22065	0.22363	0.22663	0.22965	0.23270	0.23576	0.23885	0.24196	-0.70
-0.60	0.24510	0.24825	0.25143	0.25463	0.25785	0.26109	0.26435	0.26763	0.27093	0.27425	-0.60
-0.50	0.27760	0.28096	0.28434	0.28774	0.29116	0.29460	0.29806	0.30153	0.30503	0.30854	-0.50
-0.40	0.31207	0.31561	0.31918	0.32276	0.32636	0.32997	0.33360	0.33724	0.3409	0.34458	-0.40
-0.30	0.34827	0.35197	0.35569	0.35942	0.36317	0.36693	0.37070	0.37448	0.37828	0.38209	-0.30
-0.20	0.38591	0.38974	0.39358	0.39743	0.40129	0.40517	0.40905	0.41294	0.41683	0.42074	-0.20
-0.10	0.42465	0.42858	0.43251	0.43644	0.44038	0.44433	0.44828	0.45224	0.45620	0.46017	-0.10
-0.00	0.46414	0.46812	0.47210	0.47608	0.48006	0.48405	0.48803	0.49202	0.49601	0.50000	-0.00

Standard Normal Table (continued)

Areas Under the Standard Normal Curve



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	z
0.00	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586	0.00
0.10	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535	0.10
0.20	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409	0.20
0.30	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173	0.30
0.40	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793	0.40
0.50	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240	0.50
0.60	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490	0.60
0.70	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524	0.70
0.80	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327	0.80
0.90	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891	0.90
1.00	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214	1.00
1.10	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298	1.10
1.20	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147	1.20
1.30	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774	1.30
1.40	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189	1.40
1.50	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408	1.50
1.60	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449	1.60
1.70	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327	1.70
1.80	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062	1.80
1.90	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670	1.90
2.00	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169	2.00
2.10	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574	2.10
2.20	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899	2.20
2.30	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158	2.30
2.40	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361	2.40
2.50	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520	2.50
2.60	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643	2.60
2.70	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736	2.70
2.80	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807	2.80
2.90	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861	2.90
3.00	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.9990	3.00
3.10	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929	3.10
3.20	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950	3.20
3.30	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965	3.30
3.40	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976	3.40
3.50	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983	3.50