MTHS2007 Advanced Mathematics and Statistics for Mechanical Engineers

Chapter 7: probability

School of Mathematical Sciences

The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Background/Motivation

Most experimental work in Engineering involves the use of Statistics and Probability.

In many engineering situations it is impossible to be in possession of precise information about all relevant factors, yet appropriate decisions may still be needed that affect design, operations or management.

Such decisions should based on quantitative measures that take into account the known or measured variations or uncertainties in some principled way.

For example,

1. An environmental engineer might want to estimate the level of contaminant in a lake.

   Some questions which may arise are:

   (i) How many samples should be taken?

   (ii) How should we decide whether there has been any significant change since the last set of samples were taken?

2. A quality control engineer might periodically sample a few manufactured items and study the variation from some standard.

   How should the data be used to assess whether to adjust or even halt the process?

In this part of the module we will show how such questions can be answered using Probability and Statistics.

- Probability (Chapter 7)

  Looks at formulation and analysis of mathematical models for situations where random variations occur, and are known (or assumed) to satisfy some theoretical behaviour.

- Statistics (Chapter 8)

  Concerns decision making in the presence of uncertainties and involves extracting and dealing with information from real data. Probability provides the theoretical foundation for this.

Probability - Introduction

In probability theory, a quantitative measure is assigned to each of the possible *outcomes* of a situation, or experiment. This measure describes how likely each of the outcomes is.

Collections of outcomes define *events*, which are often of more interest than individual outcomes themselves. Probabilities can also be assigned to events.

Example: Rolling a standard die, the possible *outcomes* are the integers 1-6.

The *event* that the die shows up an even number occurs if the outcome is any of 2, 4 or 6.

The **probability of an impossible event is 0**.

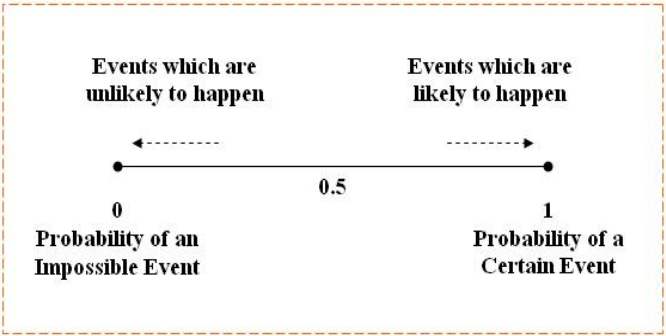
The **probability of a certain event is 1**.

Most events are neither impossible nor certain, they have varying degrees of likelihood.

The probability of an event lies between 0 and 1 (inclusive).

- Events which are very likely have probability close to 1.
- Events which are very unlikely have probability close to 0.
- Events which are about as likely to happen as to not happen have probability close to 0.5.

For example, the probability of getting a 'Head' when tossing a fair coin is 0.5.



Thus, if $P(A)$ denotes the probability of an event A, then

$$0 \leq P(A) \leq 1.$$

How to compute probability of an event ?

To begin with, suppose that every possible outcome of a random experiment is equally likely.

To compute the probability of an event A, we use the following classical definition of probability:

$$P(A) = \frac{r_A}{n},$$

where r_A = number of ways in which event A can occur

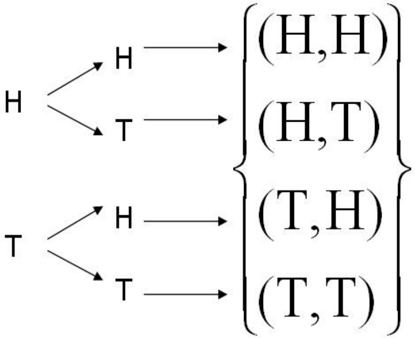
and n = total number of outcomes of the random experiment.

To understand how this works, let's consider some examples.

Example 1

In an experiment of tossing two fair coins simultaneously, the set of possible outcomes is: $U = \{ (H, H), (H, T), (T, H), (T, T) \}$

These outcomes can be obtained using the following tree diagram



Consider the following questions.

Q1 How many possible outcomes are there?

Answer The total number of outcomes is 4.

We denote this by writing $n = 4$.

Q2 What is the probability of the event that at least one 'Head' (H) is obtained?

Answer Let A denote the event that at least one H is obtained, then

$$A = \{(H, H), (H, T), (T, H)\} \quad \Rightarrow \quad r_A = 3.$$

Therefore

$$P(A) = \frac{r_A}{n} = \frac{3}{4} = 0.75.$$

Q3 What is the probability of the event that at most one 'Tail' (T) is obtained?

Answer Let B denote the event that at most one T is obtained, then

$$B = \{(H, H), (H, T), (T, H)\}$$

$$\therefore \; r_B = 3.$$

So

$$P(B) = \frac{r_B}{n} = \frac{3}{4} = 0.75.$$

Example 2

In an experiment of tossing two dice simultaneously, the set of possible outcomes is: $U = \{\ (x, y)\ :\ \ 1 \leq x, y \leq 6,\ \ x, y \in N\ \}$.

These outcomes can be tabulated as:

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

Consider the following questions.

Q1 How many possible outcomes?

Answer The total number of outcomes is 36.

We therefore write $n = 36$.

Q2 What is the probability of the event that the sum of the two dice (i.e. the numbers showing on the upper faces) is 0?

Answer Because this is impossible (the minimum sum is 2), the required probability is 0.

(We could say $A = \emptyset$, $r_A = 0$, so $P(A) = \frac{r_A}{n} = \frac{0}{36} = 0$.)

Q3 What is the probability of the event that the sum the two dice is 7?

Answer Let B denote the event that the sum is 7. Then

$$B = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$$

$$\therefore \; r_B = 6.$$

Therefore

$$P(B) = \frac{r_B}{n} = \frac{6}{36} = \frac{1}{6} \approx 0.1667.$$

Q4 What is the probability of the event that there is a 4 on at least one of the dice?

Answer Let C denote the event that at least one of the dice shows a 4. Then

$$C = \left\{ \begin{array}{cccccc} (1,4), & (2,4), & (3,4), & (4,2), & (4,3), & (4,4), \\ (5,4), & (6,4), & (4,1), & (4,5), & (4,6) & \end{array} \right\}$$

$$\therefore \; r_C = 11.$$

Thus

$$P(C) = \frac{r_C}{n} = \frac{11}{36} \approx 0.3056.$$

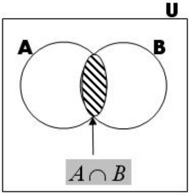
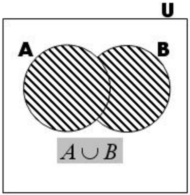
So far we have seen that probability is a numerical measure of the chance of an event and is defined as a proportion of all possible events.

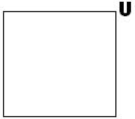
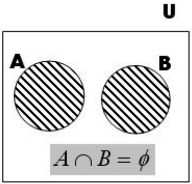
In the two examples above, we have used set notation to express an event as a collection of possible outcomes.

In fact, set notation is used extensively in probability theory.

Therefore, we first review the set theoretic notation/operations, and then see how these are used to formally describe probability in general.

Set Theoretic Operations/Concepts

Event	Notation	Meaning	Venn Diagram
A and B	$A \cap B$	Set with outcomes common to both A and B	
A or B	$A \cup B$	Set with outcomes in either A or B	

Event	Notation	Meaning	Venn Diagram
Impossible	ϕ	Empty set or Null set	
Certain	U	Universal set	
Mutually Exclusive	$A \cap B = \phi$	Disjoint Sets i.e. Nothing in common between A and B	

Axiomatic definition of Probability

Axiom 1: $P(U) = 1.$ (Probability of a certain event)

Axiom 2: $P(\phi) = 0.$ (Probability of an impossible event)

Axiom 3: For any event A, $0 \leq P(A) \leq 1.$

Axiom 4: If A and B are mutually exclusive events,

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

We can generalize the fourth property (which is called the *addition law* of probability) as:

If E_1, E_2, . . . , E_n are n mutually exclusive events, then

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

What if events are not mutually exclusive?

Then we use the following rule to evaluate probability of the event $(A \cup B)$.

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

Generalizing this rule for three events A, B and C, we have,

$$\begin{aligned} P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) \\ - P(B \cap C) - P(C \cap A) + P(A \cap B \cap C). \end{aligned}$$

Probability of the complementary event A^c

A^c (called 'A complement') is the event that A does not occur.

The probability of the event A^c is

$$P(A^c) = 1 - P(A).$$

Conditional Probability

There is another type of probability we need to consider. It arises from the situation where we gain some additional partial information about the outcome of an experiment. We use this information to update the probability of the event we are really interested in.

For example, suppose we are interested in the probability that a particular football match will end in a draw.

Now suppose we find out that the score at half time is $4 - 1$.

We would to revise our probability of the match ending in a draw.

This change of chance due to knowledge of some other event is captured by conditional probability.

Definition:

The conditional probability of event A, given that event B has occurred, is denoted by $P(A \mid B)$ and is defined as

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

From this definition of conditional probability, we have the *multiplication rule* for probabilities:

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

or

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

Example:

A box contains 10 bolts, 2 of which do not meet specification. If two bolts are chosen randomly from the box, what is the chance that both meet specification?

Consider the events A that the first bolt meets specification and B that the second bolt meets specification. We require $P(A \cap B)$.

We use the multiplication rule $P(A \cap B) = P(A)\,P(B \mid A)$.

First, we have $P(A) = 8/10$.

Then there are 9 bolts left, of which 7 meet specification, so $P(B \mid A) = 7/9$.

So $P(A \cap B) = P(A)\,P(B \mid A) = \frac{8}{10} \cdot \frac{7}{9} = \frac{28}{45} \approx 0.622$.

Law of Total Probability

An important application of conditional probability is when an event of interest is 'complicated', in the sense that its chance of occuring depends on other random events.

For example, suppose that a manufacturer buys widgets from 3 different suppliers, each supplier having a different proportion of faulty widgets. How can we find the chance that a randomly chosen widget is faulty?

If A is an event of interest and $B_1, B_2, \ldots, B_n$ are mutually exclusive and collectively exhuastive (i.e. they partition U) then

$$P(A) = \sum_{i=1}^{n} P(B_i)\, P(A \mid B_i).$$

Example:

A manufacturer buys 50% of its widgets from supplier 1, 40% from supplier 2 and 10% from supplier 3. The proportion of faulty widgets supplied by the suppliers are 0.015, 0.02 and 0.04, respectively. What is the probability that a randomly chosen widget is faulty?

We can write F for the event that the randomly chosen widget is faulty; and S_i for the event that the randomly chosen widget came from supplier i ($i = 1, 2, 3$).

Then we can find $P(F) = \sum_{i=1}^{3} P(S_i)\, P(F \mid S_i)$

$$= P(S_1)\, P(F \mid S_1) + P(S_2)\, P(F \mid S_2) + P(S_3)\, P(F \mid S_3)$$

$$= 0.5 \times 0.015 + 0.4 \times 0.02 + 0.1 \times 0.04 = 0.0195.$$

Independent Events (Statistical Independence)

It may be that knowledge of an event B has no effect on the probability of the event A, in which case A and B are independent.

Definition:

If for events A and B,

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

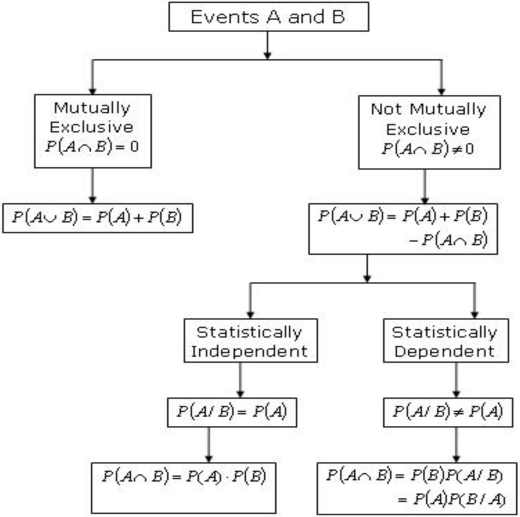
then events A and B are said to be *independent*.

Remark:

From the multiplication rule, it follows that independence of A and B (both with positive probability) is equivalent to

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

SUMMARY: Rules of Probability



Random variables

Often, we will be more interested in quantities which are derived from the outcomes of a random experiment, rather than the outcomes themselves.

A *random variable* is such a quantity — it could represent the outcome of the experiment itself (e.g. the outcome of a roll of a die), or a more complicated function of the outcomes (e.g. the sum of two dice).

We use capital letters to denote a random variable (usually X, Y or Z).

A *probability distribution* describes the possible values the random variable can take, together with the corresponding probabilities.

Probability Distributions

Probability distributions are of two types:

1. Discrete probability distribution

   In this type, the random variable (X say) takes only discrete values, e.g. number on a die, number of faulty components.

   (e.g. Binomial Distribution, Poisson Distribution)

2. Continuous probability distribution

   In this type, the random variable X takes values within a range, e.g. lifetime X of a component ($0 < X < \infty$).

   (e.g. Normal Distribution, Exponential Distribution).

To understand what a probability distribution is, re-consider the problem of throwing two dice simultaneously; and focus on the random variable which is the sum of the two dice.

The two-dice problem revisited

When two fair dice are thrown simultaneously, the outcomes are:

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

Example - Probability distribution

Let X denote the sum of the values on the faces of two dice.

Possible values for X and their associated probabilities are:

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12	Sum
$f(x_i)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	1

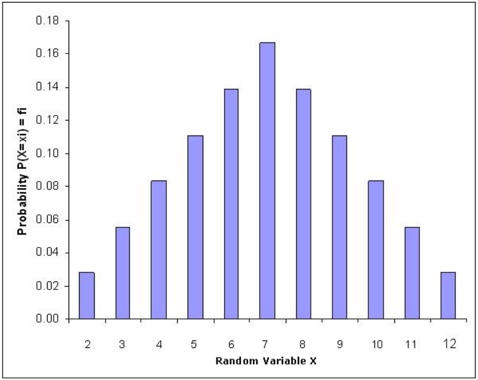
where $f(x_i) = p_i = P(X = x_i)$.

Note:

1. All probabilities are non-negative.
2. The sum of all probabilities must be 1.

Graphical representation of probability distribution function

The probability distribution (previous table) can be represented as a bar chart and the distribution of probabilities define a probability distribution function (pdf).

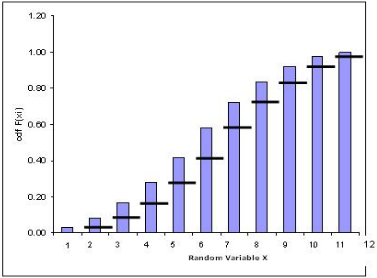


Cumulative Distribution Function (cdf)

It is also useful to define a function $F(x)$ that 'adds up' probability (from left to right) and is called the Cumulative Distribution Function (cdf)

$$F(x) = \sum_{y \leq x} f(y).$$

Note: The value of $F(x)$ starts at $F = 0$ and, as x increases, F increases to 1.



The cdf is useful for evaluating probabilities,

For example,

$$P(X \leq x) = F(x)$$

$$\begin{aligned} P(X > x) &= 1 - P(X \leq x) \\ &= 1 - F(x) \end{aligned}$$

$$\begin{aligned} P(x_1 < X \leq x_2) &= P(X \leq x_2) - P(X \leq x_1) \\ &= F(x_2) - F(x_1) \end{aligned}$$

Note:

For discrete values we need to be careful with '$<$' and '$\leq$'; and similarly for '$>$' and '$\geq$' .

Properties of probability distributions

It is useful to derive some standard analytic measures from probability distributions. Among these, the most important are:

- **Measure of Location**: a measure of the 'typical value'.
- **Measure of Spread**: a measure of the amount of variation.

Let X be a random variable with some probability distribution f. We define

1. Expected value (mean) = $E(X) = \mu = \sum_i x_i p_i$.
2. Variance = $V(X) = \sigma^2 = \sum_i (x_i - \mu)^2 p_i$.
3. Standard deviation = $\sigma = \sqrt{\text{Variance}}$.

Here the x_i's denote the possible values the random variable X can take, each with corresponding probability $p_i = f(x_i)$.

Binomial Distribution

The Binomial distribution models chance variation in repetitions of an experiment that has only two possible outcomes.

Consider a situation where an experiment will have one of two (mutually exclusive) outcomes.

For example, pass/fail, accept/reject, head/tail, working/faulty
0 / 1 (digital data), etc.

Such trials are called Bernoulli trials.

One of the outcomes is termed a 'Success' and the other outcome is termed a 'Failure'.

We define a random variable X by

$$P(X = 1) = P(\text{ success }) = p$$

$$P(X = 0) = P(\text{ failure }) = 1 - p = q$$

where $0 \leq p \leq 1$.

So the random variable X can take the values $x = 1$ (success) or $x = 0$ (failure).

Then, X has a pdf

$$f(x) = \begin{cases} p & \text{if } x = 1 \\ q & \text{if } x = 0. \end{cases}$$

It is often of interest to consider the total number of successes from a number of independent Bernoulli trials.

Example:

Consider the outcomes X_1, X_2, X_3 of 3 independent Bernoulli trials and calculate the chance of each possible sequence of outcomes:

$$\left.\begin{aligned} P(X_i = 1) &= p \\ P(X_i = 0) &= q \end{aligned}\right\} \quad \text{with} \quad p + q = 1.$$

As the trials are independent, we calculate the probabilities

$$\begin{aligned} &P\left(X_1 = x_1 \quad \text{and} \quad X_2 = x_2 \quad \text{and} \quad X_3 = x_3\right) \\ &\qquad = P\left(X_1 = x_1\right) \cdot P\left(X_2 = x_2\right) \cdot P\left(X_3 = x_3\right) \end{aligned}$$

So, for example,

$$\begin{aligned} &P\left(X_1 = 1 \text{ and } X_2 = 0 \text{ and } X_3 = 1\right) \\ &\qquad = P\left(X_1 = 1\right) \cdot P\left(X_2 = 0\right) \cdot P\left(X_3 = 1\right) \\ &\qquad = p \cdot q \cdot p \\ &\qquad = p^2\, q. \end{aligned}$$

We can calculate other probabilities similarly:

Trial 1	Trial 2	Trial 3	Probability
1	1	1	p^3
1	1	0	$p^2\,q$
1	0	1	$p^2\,q$
1	0	0	$p\,q^2$
0	1	0	$p\,q^2$
0	0	1	$p\,q^2$
0	1	1	$p^2\,q$
0	0	0	q^3

Let X denote the number of successes in 3 Bernoulli trials; then pdf of X is given by:

$X = x_i$	3	2	1	0	Sum
Probability $P(X = x_i) = f_i$	p^3	$3p^2q$	$3pq^2$	q^3	must be $\mathbf{1}$

In fact, $p^3 + 3p^2q + 3pq^2 + q^3 = (p+q)^3 = 1^3 = 1$. (Binomial theorem.)

Note:

The coefficients 1, 3, 3 and 1 of $p^3,\ 3p^2q,\ 3pq^2$ and q^3 are

the binomial coefficients $\binom{3}{0}, \binom{3}{1}, \binom{3}{2}$ and $\binom{3}{3}$

(which can also be obtained from Pascal's triangle).

Definition:

The probability distribution function of a binomial distribution $B(n, p)$ is given by

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

where $n = 0, 1, 2, \ldots$ and $p \in [0, 1]$ are specified parameters.

Here $n =$ number of Bernoulli trials,

$p =$ probability of *success*,

$q = 1 - p =$ probability of *failure*,

$x =$ number of successes,

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$, where $m! = m \cdot (m-1) \cdots 2 \cdot 1$ and $0! = 1$.

Note:

1. n and p are called *parameters* of the Binomial distribution.
2. The number of trials n is fixed.
3. There are only two possible outcomes: 'success' with probability p and 'failure' with probability q, where $q = 1 - p$.
4. The probability of success p in each independent trial is constant.

Example 1

Suppose it is known that a new treatment is successful in curing a muscular pain in 50% of cases. If it is tried on 15 patients, find the probability that

(a) At most 6 patients will be cured.

(b) The number of patients cured will be no fewer than 6 and no more than 10.

(c) 12 or more patients will be cured.

Example 1 Solution

Consider 'success' to be that a patient is cured

and let X denote the number cured patients (out of 15).

Then, X follows the Binomial distribution $B(n, p)$ with parameters

$$n = 15 \quad \text{and} \quad p = 0.5.$$

Therefore $\quad q = 1 - p = 1 - 0.5 = 0.5.$

$\therefore$ The probability distribution of X, the number cured, is

$$\begin{aligned} P(X = x) &= \binom{n}{x} p^x \, q^{n-x} \\ &= \binom{15}{x} (0.5)^x \, (0.5)^{15-x} \\ &= \binom{15}{x} (0.5)^{15}, \end{aligned}$$

for $x = 0, 1, 2, \ldots, 15$.

(a) Required probability is $P(X \leq 6)$:

$$= P(X=0) + P(X=1) + P(X=2) + P(X=3)$$

$$+P(X=4) + P(X=5) + P(X=6)$$

$$= \binom{15}{0}(0.5)^{15} + \binom{15}{1}(0.5)^{15} + \binom{15}{2}(0.5)^{15} + \binom{15}{3}(0.5)^{15}$$

$$+\binom{15}{4}(0.5)^{15} + \binom{15}{5}(0.5)^{15} + \binom{15}{6}(0.5)^{15}$$

$$= (\,1 + 15 + 105 + 455 + 1365 + 3003 + 5005\,)\ (0.5)^{15}$$

$$= 9949 \times (0.5)^{15}$$

$$\approx 0.3036.$$

(b) Required probability is $P(6 \leq X \leq 10)$:

$$= P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10)$$

$$= \binom{15}{6}(0.5)^{15} + \binom{15}{7}(0.5)^{15} + \binom{15}{8}(0.5)^{15}$$

$$+ \binom{15}{9}(0.5)^{15} + \binom{15}{10}(0.5)^{15}$$

$$= \left[\binom{15}{6} + \binom{15}{7} + \binom{15}{8} + \binom{15}{9} + \binom{15}{10}\right](0.5)^{15}$$

$$\approx 0.790.$$

(c) Required probability is $P(X \geq 12)$:

$$= \left[\binom{15}{12} + \binom{15}{13} + \binom{15}{14} + \binom{15}{15}\right] (0.5)^{15}$$

$$\approx 0.018.$$

Mean and Standard deviation of Binomial Distribution

1. Mean = $\mu = \sum_{x=0}^{n} x \binom{n}{x} p^x \, q^{n-x} \; = \; np.$

   (after some algebra . . .)

2. Variance = $\sigma^2 = \sum_{x=0}^{n} (x - np)^2 \binom{n}{x} p^x \, q^{n-x} \; = \; npq.$

3. Standard deviation = $\sqrt{\text{Variance}} = \sqrt{npq}.$

Example:

For the Binomial distribution $B(n, p) = B(3, 0.5)$,

Mean $= \mu = n\,p = 3 \times 0.5 = 1.5$

and

Standard deviation $= \sigma = \sqrt{n\,p\,q} = \sqrt{3 \times 0.5 \times 0.5} \approx 0.866.$

Example 2

If the mean of a Binomial distribution $B(n, p)$ is 3 and the variance is $\frac{3}{2}$, find the probability of obtaining at least 2 successes.

Solution:

Here, mean $= n\,p = 3$ and Variance $= n\,p\,q = \dfrac{3}{2}$.

So we have $q = \dfrac{n\,p\,q}{n\,p} = \dfrac{3/2}{3} = \dfrac{1}{2} \quad \Rightarrow \quad p = 0.5.$

Now, $n\,p = 3$, so $n = \dfrac{3}{p} = \dfrac{3}{0.5} = 6$.

$\therefore$ the required probability is: $P(X \geq 2)$

$$= \left[\binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6}\right] (0.5)^6 \approx 0.8906.$$

Alternatively:

Required probability is

$$
\begin{aligned}
P(X \geq 2) &= 1 - P(X \leq 1) \\
&= 1 - \left[\binom{6}{0} + \binom{6}{1} \right] (0.5)^6 \\
&= 1 - 0.109375 \\
&\approx 0.8906.
\end{aligned}
$$

Poisson Distribution

If the parameters n and p of a Binomial distribution are known, we can find the probability distribution and do calculations with it. But, in situations where n is large, this can be very laborious.

If we are in the situation where n is large and p is small we can simplify some of these calculations.

Mathematically, we assume that $n \to \infty$ and $p \to 0$ in such a way that $np \to \lambda \in (0, \infty)$.

Then it can be shown that

$$P(X = x) = \binom{n}{x} p^x \, q^{n-x} \to \frac{e^{-\lambda} \, \lambda^x}{x!}.$$

The probability distribution function for the Poisson distribution is

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

where λ is a fixed constant, called the *parameter* of the Poisson distribution. We write this as $X \sim Po(\lambda)$.

Note:

1. The Poisson distribution takes values in the non-negative integers, i.e. $x = 0, 1, 2, \ldots$.
2. The sum of probablities is, as required,

$$\sum_{x=0}^{\infty} \frac{e^{-\lambda}\,\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = e^{-\lambda}\, e^{\lambda} = 1.$$

We have seen that the Poisson distribution can be obtained from the Binomial(n, p) distribution as $n \to \infty, \quad p \to 0$.

In practice, this means that we can use the Poisson distribution to *approximate* the Binomial(n, p) distribution provided n is "large" and p is "small". This can be very useful for calculating approximate Binomial probabilities when n is large, which is precisely when calculating exact probabilities with the Binomial distribution becomes cumbersome.

Example:

The probability that a part produced by a certain machine is defective is known to be 0.1. What is the probability that in 10 items produced, at most 1 will be defective?

1. Solution using Binomial distribution

Consider success to be that the part selected at random is defective.

Then $X \sim B(n, p)$, where $n = 10, \quad p = 0.1 \quad \Rightarrow \quad q = 0.9$

Now,

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

$$= \binom{n}{x} (0.1)^x \, (0.9)^{10-x}, \qquad x = 0, 1, \ldots, 10.$$

The required probability is

$$P(X \leq 1) = \binom{10}{0}(0.1)^0(0.9)^{10} + \binom{10}{1}(0.1)^1(0.9)^9$$

$$= \cdots \approx 0.7360989.$$

2. Solution using Poisson distribution

Here $n = 10$ and $p = 0.1$, so $\lambda = np = 1$.

Hence, by Poisson distribution formula,

$$P(X = x) = \frac{e^{-1}\,1^x}{x!}, \qquad x = 0, 1, 2, \ldots, 10.$$

So the required probability is

$$P(X \le 1) = P(X = 0) + P(X = 1)$$

$$= \frac{e^{-1}\,1^0}{0!} + \frac{e^{-1}\,1^1}{1!}$$

$$= 2\,e^{-1} \approx 0.735758.$$

So the approximation is correct up to 3 decimal places in this example.

Mean and St Dev of Poisson distribution

1. Mean = $\mu = \sum_{x=0}^{\infty} x \, \frac{e^{-\lambda} \, \lambda^x}{x!}$

$$= \lambda \sum_{x=1}^{\infty} \frac{e^{-\lambda} \, \lambda^{x-1}}{(x-1)!}$$

$$= \lambda \sum_{m=0}^{\infty} \frac{e^{-\lambda} \, \lambda^{m}}{m!}$$

$$= \lambda.$$

2. Variance = $\sigma^2 = \sum_{x=0}^{\infty} (x-\lambda)^2 \, \frac{e^{-\lambda} \, \lambda^x}{x!}$

$$= \cdots = \lambda.$$

3. Standard deviation = $\sqrt{\text{Variance}} = \sqrt{\lambda}.$

Continuous Distributions

Deal with continuous measurements (height, weight, time, etc).

Many of their properties are analogous to those of discrete distributions, with integrals replacing summations.

The pdf is a continuous curve: Area under the pdf represents probability, and total area (probability) under the curve = 1.

So for a continuous random variable X with pdf $f(x)$ we have

$$\int f(x)\,\mathrm{d}x = 1,$$

(where the integral is over the range of possible values for X.)

Furthermore,

$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

(i.e the area under the pdf curve between a and b.)

The ideas of expected values derived for discrete probability distributions can be extended to continuous distributions.

This is done by replacing sums by integrals.

Thus

$$\mu = E(X) = \int_U x\, f(x) \mathrm{d}x$$

$$\sigma^2 = V(X) = E\left((X-\mu)^2\right) = \int_U (x-\mu)^2\, f(x) \mathrm{d}x$$

(where U is the set of all possible outcomes x.)

Normal Distribution - Introduction

We will concentrate on one in particular: the *normal* distribution. The normal distribution, also called the Gaussian distribution, is a very important probability distribution, applicable in many fields.

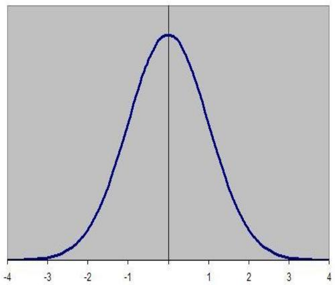
- It is the most important single distribution in Statistical/Probability methods.
- Each member $N(\mu, \sigma^2)$ of the family is defined by two parameters: **location** (mean/average) μ and **scale** (standard deviation/dispersion) σ. (So variance is σ^2.)
- The Probability Density Function (pdf) is defined as

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

Normal Curve

To indicate that a random variable X is normally distributed with mean μ and variance σ^2, we write $X \sim N(\mu, \sigma^2)$.

The pdf of a Normal distribution $N(\mu, \sigma^2)$ is a bell shaped curve called the Normal (or Gaussian) curve, shown in the figure below:

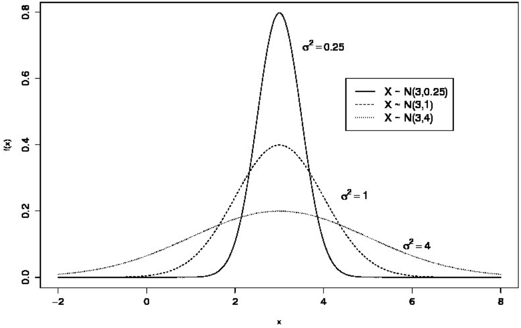


It is centred on and symmetric about the mean μ, and the variance σ^2 or standard deviation σ quantifies the spread.

Properties of Normal Curve

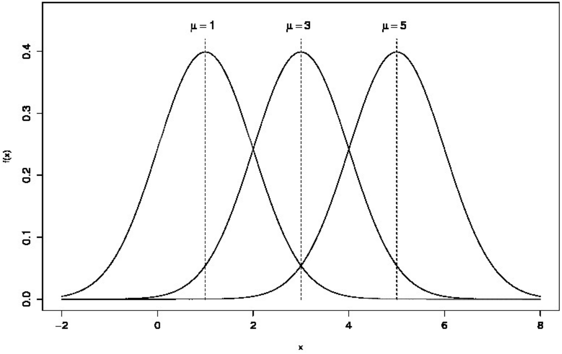
The pdf has a number of properties:

- The distribution is symmetric about the mean μ.
- Changing σ adjusts the spread (and height).



Properties of Normal Curve

- Changing μ adjusts the location of the distribution (as shown below).



- Mean = μ, Variance = σ^2.

Standard Normal Distribution

The standard normal distribution $N(0,1)$ is the normal distribution with a **mean of zero** and a **variance of one**.

That is,

Mean $\mu = 0$

and

Variance $\sigma^2 = 1$ OR Standard deviation $\sigma = 1$.

The symbol Z is often used for a random variable distributed as $N(0,1)$.

The probability density function for the Standard Normal distribution is given by

$$f(z) = \frac{1}{\sqrt{2\pi}}\, e^{-z^2/2} \qquad ; \quad -\infty < z < \infty$$

Standard Normal Table

The standard Normal table, defined for the standardised Normal variate $Z \sim N(0, 1)$, gives the area to the left of a specified value z as follows:

$P(Z \leq z_1) =$ Area under the curve to the left of z_1.

In other words, the table gives the *cumulative distribution function* (cdf) $F(z)$ for the standard normal distribution.

We can calculate other probabilities from these, e.g.

$P(z_1 \leq Z \leq z_2) = P(Z \leq z_2) - P(Z \leq z_1)$.

Standard Normal Table

In order to look at the tabulated values for the Normal distribution, we relate all normal random variables $X \sim N(\mu, \sigma^2)$ to the standard normal variate Z by

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

so that $Z \sim N(0, 1)$. This process is called *standardisation* of X.

So if $X \sim N(\mu, \sigma^2)$, then

$$\begin{aligned} P(X \leq x) &= P\left(\tfrac{X-\mu}{\sigma} \leq \tfrac{x-\mu}{\sigma}\right) \\ &= P\left(Z \leq \tfrac{x-\mu}{\sigma}\right). \end{aligned}$$

E.g.: If $X \sim N(5, 4)$, then $P(X \leq 7) = P\left(Z \leq \frac{7-5}{2}\right) = P(Z \leq 1)$,

which can be looked up in the tables. (**Ans:** 0.8413).

Standard Normal Table - Page 1

$$f(-z)=f(z),\quad F(-z)=1-F(z)$$

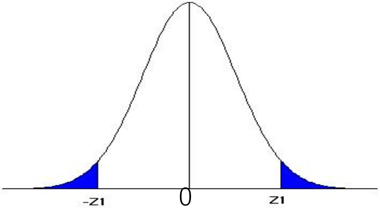
$f(z)$	z	$F(z)$									
0.00		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.399	0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.397	0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.391	0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.381	0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.368	0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.352	0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.333	0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.312	0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.290	0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.266	0.9	0.8159	0.8168	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
0.242	1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
0.218	1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
0.194	1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
0.171	1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
0.150	1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Standard Normal Table - Page 2

f(z)	z	F(z)									
0.00		0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.130	1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
0.111	1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
0.094	1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
0.079	1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
0.066	1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
0.054	2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
0.044	2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
0.035	2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
0.028	2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
0.022	2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
0.018	2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
0.014	2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
0.010	2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
0.008	2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
0.006	2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986

From the symmetry of the standard normal curve about its mean $\mu = 0$,

- $P(Z \leq 0) = 0.5$,
- $P(Z \leq z_1) = F(z_1)$
- $$\begin{aligned} F(-z_1) &= P(Z \leq -z_1) \\ &= P(Z \geq z_1) \\ &= 1 - P(Z \leq z_1) \\ &= 1 - F(z_1) \end{aligned}$$



Example

If $Z \sim N(0,1)$ is a standard normal variate, calculate

(a) $P(Z \leq 1.37)$

(b) $P(Z > 1.60)$

(c) $P(-1.37 < Z < 1.60)$

(d) $P(-1.62 < Z < -0.54)$

1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292

From the Standard Normal table, $P(Z \leq 1.37) = 0.9147.$

1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616

From the Standard Normal table,

$$P(Z > 1.60) = 1 - P(Z < 1.60) = 1 - 0.9452 = 0.0548.$$

Solution (c) & (d)

From the Standard Normal table (values obtained in (a) and (b)),

$$\begin{aligned} P(-1.37 < Z < 1.60) &= F(1.60) - F(-1.37) \\ &= 0.9452 - (1 - 0.9147) \\ &= 0.8599. \end{aligned}$$

From the Standard Normal table (check for yourself),

$$\begin{aligned} P(-1.62 < Z < -0.54) &= F(-0.54) - F(-1.62) \\ &= (1 - F(0.54)) - (1 - F(1.62)) \\ &= 1 - 0.7054 - 1 + 0.9474 \\ &= 0.2420. \end{aligned}$$

Notes for solving Examples

- $P(Z \leq z_1) = F(z_1)$
- $P(Z \leq -z_1) = F(-z_1) = 1 - F(z_1)$
- $P(z_1 \leq Z \leq z_2) = F(z_2) - F(z_1)$
- $P(Z \geq z_1) = P(Z \leq -z_1) = F(-z_1) = 1 - F(z_1)$

In all above formulae, F denotes the cumulative distribution function, values of which are available (for standarized normal variate Z) from the Normal distribution table.

Extra Examples

1. If Z is a standard normal variate (i.e. $Z \sim N(0, 1)$), then find
   (i) $P(0.6 < Z < 1.2)$ (Ans: 0.1592)
   (ii) $P(-2.1 < Z < -1.7)$ (Ans: 0.0267)
   (iii) $P(-0.8 < Z < 1.4)$ (Ans: 0.7074)
   (iv) $P(Z < -1.36)$ (Ans: 0.0869)
   (v) $P(X > 1.52)$ (Ans: 0.0643)
2. Given a population of birds, the wing spans are normally distributed with mean 14.1 cm and standard deviation 1.7 cm. Calculate the probability that a randomly selected bird has a wing span less than 17 cm.

   (Ans: 0.9560)

3. The number of calories in a salad on the lunch menu is normally distributed with mean $\mu = 200$ and standard deviation $\sigma = 5$. Find the probability that the salad you select will contain

   (i) More than 208 calories. (**Ans:** 0.0548)

   (ii) Between 190 and 200 calories. (**Ans:** 0.4772)

Functions of a random variable

If X is a random variable then so is $Y = g(X)$, for any reasonable function $g(\cdot)$.

From the definitions $\mu = E(X) = \sum_i \, x_i \, p_i$ and $\sigma^2 = V(X) = E\left(\, (X - \mu)^2 \, \right) = \sum_i \, (x_i - \mu)^2 \, p_i$ it can be shown that if a and b are constants,

$$E(aX + b) = a\,E(X) + b,$$

$$V(aX + b) = a^2\,V(X).$$

And also that $\quad V(X) = E(X^2) - [E(X)]^2.$

More generally, for any function $g(x)$,

$$[E(g(x))] = \begin{cases} \sum_i \; g(x_i)\; p_i & \text{discrete case} \\ \int_U g(x) f(x)\,\mathrm{d}x & \text{continuous case} \end{cases}.$$

Sums/Differences of random variables

Often one is interested in sums of random variables.

For example, $Y = X_1 + X_2$, where X_1 and X_2 are *independent* random variables with some distribution. Here Y is a *NEW* random variable: the sum of X_1 and X_2.

Generally, Y will not have the same distribution as X_1 or X_2.
(But see special cases on the next slide.)

It can be shown that

$$E(X_1 \pm X_2) = E(X_1) \pm E(X_2)$$

and

$$V(X_1 \pm X_2) = V(X_1) + V(X_2).$$

There are two important special cases where the sum/difference of two *independent* random variables does have the same type of distribution as the original variables:

Poisson If $X_1 \sim Po(\lambda_1)$ and $X_2 \sim Po(\lambda_2)$ then

$$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2).$$

Normal If $X_1 \sim N(\mu_1, \sigma_1^2)$ and $X_2 \sim N(\mu_2, \sigma_2^2)$ then

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

and

$$X_1 - X_2 \sim N(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2).$$

(These results generalise to more than two random variables.)

Poisson

In a certain factory the number of accidents that occur each month follows a Poisson distribution with mean $\frac{1}{2}$, independently for each month. What is the probability that there are exactly 2 accidents over a 3-month period?

Solution: Write X_i for the number of accidents in month i, so that $Y = X_1 + X_2 + X_3$ is the number of accidents over 3 months. The X_i variables are Poisson and independent, so Y is Poisson with mean $E(X_1) + E(X_2) + E(X_3) = \frac{3}{2}$.

So the probability we require is

$$P(Y = 2) = \frac{e^{-3/2}\,(\frac{3}{2})^2}{2!} = e^{-3/2} \cdot \frac{9}{8} \approx 0.251.$$

Normal

(a) The inside diameter X of a bearing supplied by a manufacturer can be approximated by a Normal distribution with mean 42.04 mm and a standard deviation of 0.06 mm. Obtain the probability that the diameter of a chosen bearing will fall within the tolerance range, $41.95 < X < 42.10$.

(b) The diameter Y of drive shafts supplied by an independent manufacturer may be taken as distributed according to a Normal distribution with mean 41.98 mm and standard deviation 0.1 mm. Obtain the probability that a bearing will not slide onto a shaft when selected. (The bearing must be able to slide onto the shaft in order to function.)

Solution:

(a) $X \sim N(\mu, \sigma^2) = N\left(42.04, (0.06)^2\right)$

Let $Z = \dfrac{X - 42.04}{0.06}$ so that $Z \sim N(0, 1)$.

$\therefore$ Required probability is

$$
\begin{aligned}
& P(41.95 < X < 42.10) \\
& = P\left(\frac{41.95 - 42.04}{0.06} < \frac{X - \mu}{\sigma} < \frac{42.10 - 42.04}{0.06}\right) \\
& = P\left(-1.5 < Z < 1\right) \\
& = F(1) - F(-1.5) \\
& = F(1) - (1 - F(1.5)) \\
& = 0.8413 - 1 + 0.9332 \qquad \text{(from Normal Table)} \\
& = 0.7745.
\end{aligned}
$$

(b) Bearing will not fit if $X \leq Y$.

Consider $D = X - Y$ (i.e. difference between diameters)

From property of Normal variates,

$$D \sim N\left(\mu_X - \mu_Y, \sigma_X^2 + \sigma_Y^2\right)$$

$$\Rightarrow \; D \sim N\left(42.04 - 41.98, 0.06^2 + 0.1^2\right)$$

$$\Rightarrow \; D \sim N\left(0.06, 0.1166^2\right)$$

So the required probability is $P(X - Y \leq 0) = P(D \leq 0)$

$$= P\left(\frac{D - 0.06}{0.1166} \leq \frac{0 - 0.06}{0.1166}\right) \quad \text{(Standardising } D\text{)}$$

$$= P\left(Z \leq -0.5146\right) \quad \text{where } Z \sim N(0,1)$$

$$= F(-0.5146) = 1 - F(0.5416)$$

$$= 1 - 0.6966 = 0.3034 \qquad \text{(from Normal Table)}$$