

# Chapter 10

## Probability

*Probability theory is nothing but common sense reduced to calculation.*

Pierre Simon Laplace

In what follows we are going to cover the basics of probability. The ideas are reasonably straightforward, however as it involves counting it is very easy to make mistakes - as we shall see.

Suppose we perform an experiment whose outcome is not perfectly predictable e.g. roll a die or toss a coin. Imagine we make a list of all possible outcomes, call this list  $S$  the *sample space*. So

- If we toss a coin  $S$  consists of {Head, Tail}, we write  $S = \{\text{Head, Tail}\}$ ,
- If we roll a die  $S = \{1, 2, 3, 4, 5, 6\}$
- If a princess kisses a frog then we have two possibilities  $S = \{\text{we get a prince, we get an embarrassed frog}\}$
- When we roll two dice then  $S$  is the set of pairs

|       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|
| (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) |

An *event*  $A$  is a collection of outcomes of interest, for example rolling two dice and getting a double. In this case the event  $A$  is defined as

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

Suppose that the event B is that the sum is less than 4 when we roll two dice, then

$$B = \{ (1,1), (1,2), (2,1) \}.$$

If two events A and B have no elements in common then we say they are *mutually exclusive*. For example let A be the event {At least one 6} that is

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

Since A and B have no elements in common they are mutually exclusive. Define the event C as

$$C = \{ (2,3), (5,7) \}$$

Then A and C are also mutually exclusive. If  $D = \{ \text{sum exceeds } 10 \}$  then A and D are **not** mutually exclusive! *Check this yourself.*

### Combining events

- It is handy to have a symbol for not A, we use  $\sim A$  but we are not very picky and not A is acceptable.
- The event A and B, often written  $A \cap B$  is the set of outcomes which belong both to A and to B.
- The event A or B, often written  $A \cup B$  is the set of outcomes which belong either to A or to B or to both.

You will recognise the notation from the earlier discussion on sets.

Suppose  $S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$  then if we define  $A = \{1, 3, 5, 7, 9\}$  and  $B = \{4, 5, 7\}$  we have

- $A \cap B = A$  and  $B = \{5, 7\}$
- While  $A \cup B = A$  or  $B = \{1, 3, 4, 5, 7, 9\}$
- $\sim B = \text{not } B = \{1, 2, 3, 4, 8, 9\}$

### 10.0.2 Probability - the rules

Now to each event we are going to assign a measure ( in some way ) called the probability. We will write the probability of an event  $A$  as  $P[A]$ . We will set out some rules for probabilities, the main ones are as follows:

1.  $0 \leq P[A] \leq 1$ .
2.  $P[S] = 1$
3. For mutually exclusive events  $A$  and  $B$   $P[A \text{ or } B] = P[A] + P[B]$

We will add a few extra rules

- (i) For mutually exclusive events  $A_1$  and  $A_2$  and  $A_3 \cdots A_n$  then

$$P[A_1 \cup A_2 \cup A_3 \cdots \cup A_n \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots + P[A_n] + \cdots$$

or written differently

$$P[A_1 \text{ or } A_2 \text{ or } A_3 \cdots \text{ or } A_n \cdots] = P[A_1] + P[A_2] + P[A_3] + \cdots + P[A_n] + \cdots$$

- (ii) For an event  $A$

$$P[\text{not } A] = 1 - P[A]$$

- (iii) For events  $A$  and  $B$

$$P[A \text{ or } B] = P[A] + P[B] - P[A \text{ and } B]$$

All this is a bit fiddley but is not really very hard. If you were not too confused at this point you will have noticed that we do not have a way of getting the probabilities. This is a difficult point except in the case we are going to discuss.

### 10.0.3 Equally likely events

Suppose that every outcome of an experiment is equally likely. Then we can show from the rules above for any event  $A$

$$P[A] = \frac{\text{the number of outcomes in } A}{\text{the number of possible outcomes}}$$

This means we can do some calculations.

### examples

1. Suppose that the outcomes

- that a baby is a girl
- that a baby is a boy

are equally likely. Then as there are two possible outcomes we have  $P[\text{girl}] = 1/2 = P[\text{boy}]$ .

2. Suppose now a family has 3 children, the possibilities are

BB BG GB GG

and so  $P[\text{one boy and one girl}] = 2/4 = 1/2$  while  $P[\text{two girls}] = 1/4$

3. The famous statistician R A Fisher had seven daughters. If you count the possible sequences BBBB BBB to GGGGGGG you will find that there are  $2^7 = 128$ . Only one sequence is all girl so the probability of this event is  $1/128$ .
4. A pair of dice is thrown. What is the probability of getting totals of 7 and 11. Suppose now we throw the two dice twice. What is the probability of getting a total of 11 and 7 in this case?
5. We draw 2 balls from an urn containing 6 white and 5 black, What is the probability that we get one white and one black ball?

As you can see we really need some help in counting.

### Exercises

1. A poker hand consists of 5 cards drawn from a pack of 52. What is the probability that a hand is a *straight*, that is 5 cards in numerical order, but not all of the same suit.
2. What is the probability that a poker hand is a *full house*, that is a triple and a pair.
3. A and B flip a coin in turn. The first to get a head wins. Find the sample space. What is the probability that A wins?

4. The game of craps is played as follows: A player rolls two dice. If the sum is a 2, 3 or 12 he loses. If the sum is a seven or an 11 he wins. Otherwise the player rolls the dice until he gets his initial score, in which case he wins or gets a 7 in which case he loses. What is the probability of winning?
  
5. A man has  $n$  keys, one of which will open his door. He tries keys at random, discarding those that don't work until he opens the door. What is the probability that he is successful on the  $k$ th try.
  
6. **The birthday problem** How many people should be in a room to make the probability of two or more having the same birthday more than 0.5? This is quite difficult and a simpler approach is to consider the probability that no two people have the same birthday.

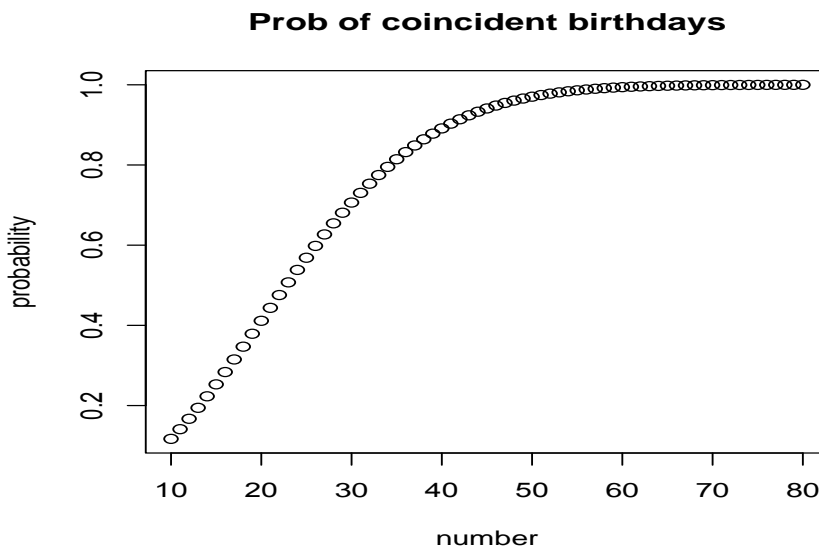
It is often a useful dodge in probability to look at  $P[\text{not } A]$  when  $P[A]$  is hard.

So  $P[\text{no coincidences}] =$

$$\frac{365 \times 364 \times 363 \times \dots \times (365 - n + 1)}{365 \times 365 \times \dots \times 365}$$

$$= 1 \times (1 - 364/365) \times (1 - 364/365) \times (1 - 363/365) \times \dots \times (1 - (365 - n + 1)/365)$$

| Number | Probability |
|--------|-------------|
| 15     | 0.74709868  |
| 16     | 0.71639599  |
| 17     | 0.68499233  |
| 18     | 0.65308858  |
| 19     | 0.62088147  |
| 20     | 0.58856162  |
| 21     | 0.55631166  |
| 22     | 0.52430469  |
| 23     | 0.49270277  |
| 24     | 0.46165574  |
| 25     | 0.43130030  |



### 10.0.4 Conditional Probability

Sometime it is natural to talk of the probability of an event A **given** some other event has occurred. We write the probability of A given B as  $P[A | B]$  and define it as

$$P[A | B] = \frac{P[A \cap B]}{P[B]}$$

Remember this is a fancy way of writing

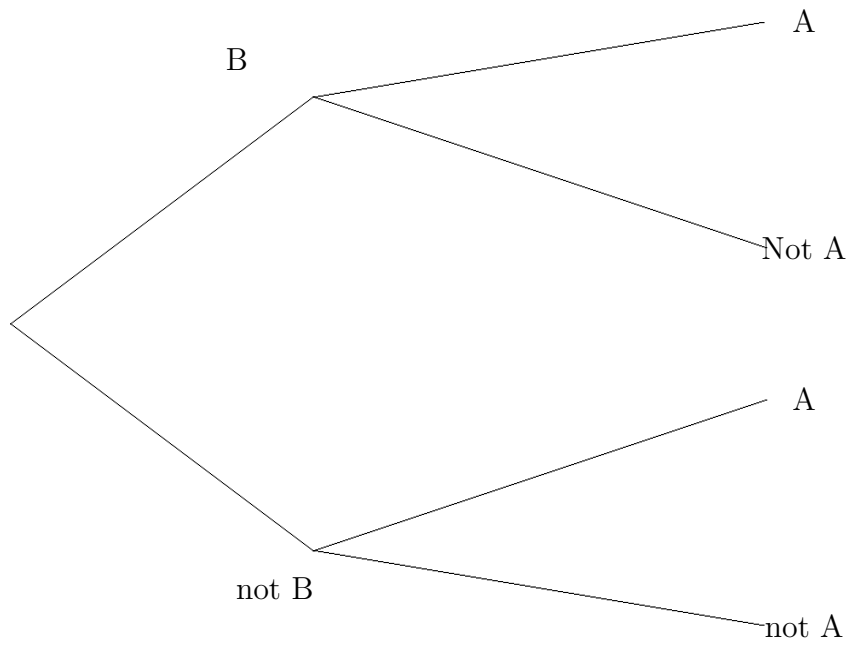
$$P[A | B] = \frac{P[A \text{ and } B]}{P[B]}$$

While conditional probabilities can have interesting philosophical implications they also allow one to do calculations. Thus

$$P[A] = P[A | B]P[B] + P[A | \sim B]P[\sim B]$$

or more generally if  $B_1, B_2, \dots$  are the only possibilities so  $\sum_{i=1}^n B_i = 1$  then

$$P[A] = \sum_{i=1}^n P[A | B_i]P[B_i]$$



360°  
thinking.

**Deloitte.**

Discover the truth at [www.deloitte.ca/careers](http://www.deloitte.ca/careers)

© Deloitte & Touche LLP and affiliated entities.



## Examples

1. Consider the table below.

|        | Employed | Unemployed | Total |
|--------|----------|------------|-------|
| Male   | 460      | 40         | 500   |
| Female | 140      | 260        | 400   |
| Total  | 600      | 300        | 900   |

Then

- $P[\text{Male}] = 500/900$
  - $P[\text{Male and Unemployed}] = 40/900$
  - $P[\text{Unemployed} \mid \text{Male}] = 40/500 = P[\text{Unemployed and Male}] / P[\text{Male}]$   
 $= 40/900 \div 500/900 = 40/500$ .
2. Suppose we buy widgets from 3 suppliers A,B and C. They supply all production and the number of defective items per batch as well as their share of our supply is given below.

|                      | A        | B    | C    |
|----------------------|----------|------|------|
|                      | Supplier |      |      |
| Proportion supplied  | 0.60     | 0.30 | 0.10 |
| Proportion defective | 0.03     | 0.05 | 0.07 |

What proportion of widgets are defective?

We know

- $P[\text{defective}|A] = 0.03$
- $P[\text{defective}|B]=0.03$
- $P[\text{defective}|C]=0.07$

so using the formula we have

$$P[\text{defective}] = P[\text{defective}|A] \times P[A] + P[\text{defective}|B] \times P[B] + P[\text{defective}|C] \times P[C]$$

$$\text{So } P[\text{defective}] = 0.03 \times 0.6 + 0.03 \times 0.3 + 0.07 \times 0.1 = 0.034$$



### 10.0.5 Bayes

We also have **Bayes Theorem**

$$P[A|B] = \frac{P[B|A]P[A]}{P[B]} \quad (10.1)$$

or

$$P[A|B] \propto P[B|A]P[A] \quad (10.2)$$

Here  $\propto$  means equal to but multiplied by a constant.

You will often find that you can compute  $P[A | B]$  when really you want  $P[B | A]$ . Bayes theorem gives you the means for turning one into the other.

#### Examples

1. Take the data in the example 2 above. We know that  $P[\text{defective} | A]=0.03$  and we found that  $P[\text{defective}]=0.034$ . Then suppose we pick up a defective component and ask what is the probability that it come from  $A$ . Thus we need  $P[A | \text{defective}]$ .

We can use Bayes to give

$$\begin{aligned} P[A | \text{defective}] &= P[\text{defective} | A]P[A]/P[\text{defective}] \\ &= 0.03 \times 0.6/0.34 = 9/17 = 0.529. \end{aligned}$$

2. Suppose that the probability that a person has a disease  $P[D] = 0.01$ . A test is available which is correct 90% of the time. If we use  $Y$  to denote that the test is positive and  $\sim Y$  negative we mean

$$P[Y|D] = P[\sim Y | \sim D] = 0.9$$

Now the probability of a yes is

$$P[Y] = P[Y|D]P[D] + P[Y | \sim D]P[\sim D] = 0.9 \times 0.01 + 0.1 \times 0.99 = 0.108.$$

The more interesting case is


$$P[D|Y] = \frac{P[Y|D]P[D]}{P[Y]} = 0.009/0.108 = 0.0833$$

## Exercises


1. An insurance broker believes that a quarter of drivers are accident prone. What is more the probability of an accident prone driver making a claim is  $\frac{1}{3}$  while for a non accident prone drive the probability is  $\frac{1}{5}$ . What is the probability of a claim? On his way home the broker sees that one of his customers has driven his car into a tree. What is the probability that this customer is accident prone?
2. An urn contains 4 red and 6 green balls. One ball is drawn at random and it's colour observed. It is then returned to the urn and 3 new balls of the same colour are added to the urn, which now contains 13 balls. A second ball is now drawn from the urn.
  - (a) What is the probability that the first ball drawn was green?
  - (b) What is the probability of getting a red ball *given* the first ball drawn was green
  - (c) What is the probability of getting a green ball in the second draw?
3. Sometime used by unscrupulous students of probability - We have 3 cards. The first card has two red sides, the second two black sides. The remaining card has one black and one red side. Otherwise the cards are identical.
 

The three cards are mixed in a hat and one card is selected at random an placed on a table. If the exposed side is red what is the probability that the hidden side is black?

SIMPLY CLEVER


**ŠKODA**  


We will turn your CV into  
an opportunity of a lifetime



Do you like cars? Would you like to be a part of a successful brand?  
We will appreciate and reward both your enthusiasm and talent.  
Send us your CV. You will be surprised where it can take you.

Send us your CV on  
[www.employerforlife.com](http://www.employerforlife.com)



### Independence

If  $P[A|B] = P[A]$  then we say  $A$  and  $B$  are independent. This is usually written in the equivalent form

$$P[A \cap B] = P[A]P[B]$$

Independent is very useful and plays a central role in statistics.

### 10.0.6 Random Variables and distributions

If we conduct an experiment and see an outcome we almost always code the outcome in some way, say H,T for head and tail or even 0,1. The coding is known as a random variable, usually written as a capital such as  $X$ . If we toss a coin we can say that the outcome is  $X$ . The actual values may be head, head, tail, giving the sequence of values of  $X$  as H, H, T, . . . We use random variables when we have probability distributions, that is lists of possible outcomes and probabilities, such as in the table

|            |     |     |     |     |
|------------|-----|-----|-----|-----|
| k          | 0   | 1   | 2   | 3   |
| $P[X = k]$ | 0.1 | 0.3 | 0.5 | 0.1 |

We point out that the sum of the probabilities must be one, that is  $\sum_{k=0}^3 P[X = k]$ . We define the *cumulative distribution function* (c.d.f.)  $F(x)$  as the cumulative sum of the probabilities

$$F(x) = \sum_{k=0}^x P[X = k]$$

So in the example above

|            |     |     |     |     |
|------------|-----|-----|-----|-----|
| k          | 0   | 1   | 2   | 3   |
| $P[X = k]$ | 0.1 | 0.3 | 0.5 | 0.1 |
| $F(x)$     | 0.1 | 0.4 | 0.9 | 1.0 |

It is more usual to give a formula for a random variable, for example

$$P[x = k] = 0.3 \times 0.7^{x-1} \quad x = 1, 2, 3, \dots$$

As the formula is commonly shorter you can see why.

## 10.1 Expectation

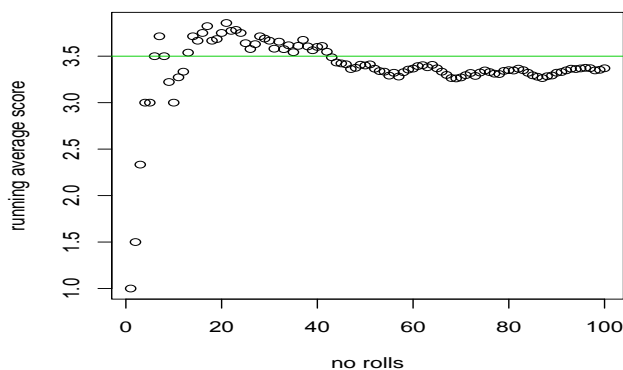
We can also view probability from the point of view of what happens in the long run. Given a random variable  $X$  define the expected value of  $X$  written  $E[X]$  as

$$E[X] = \sum_{\text{all } x} xP[X = x]$$

The expected value can be regarded as the *long run average*. So if we roll a fair die and the outcome is  $X$  then  $P[X = i] = 1/6 \quad i = 1, 2, \dots, 6$  and so

$$E[X] = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \dots + 6 \times \frac{1}{6} = 3.5$$

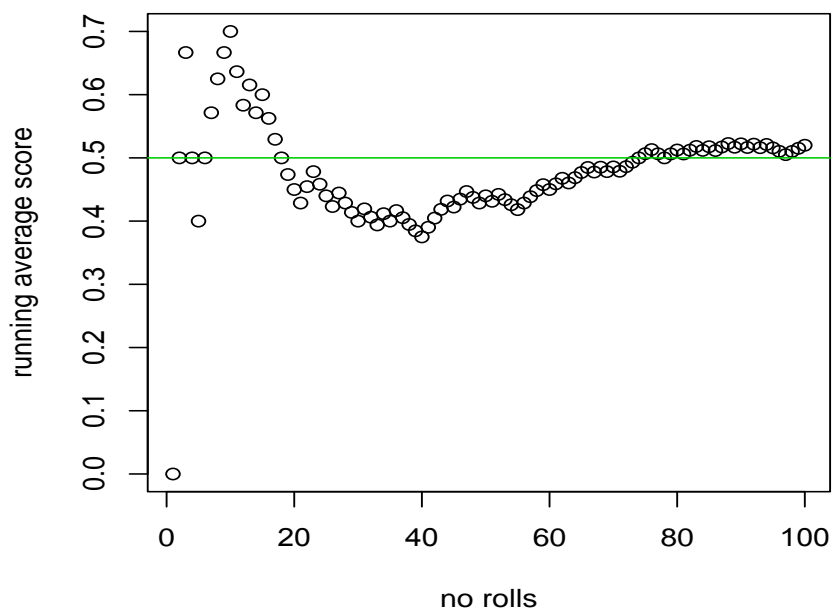
You can be sure that if you roll a die you will never get 3.5, *however* if you rolled a die and kept an average of the score you will find that this will approach 3.5, see the plot below



For a coin we have Head and Tail. Suppose we count head as 1 and tail as zero, then

$$P[X = 1] = 1/2 \text{ and } P[X = 0] = 1/2$$

and so  $E[X] = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$ . A similar experiment gives the following



### 10.1.1 Moments

Some important expected values in statistics are the *moments*

$$\mu_r = E[X^r] \quad r = 1, 2, \dots$$

since we can usually estimate these while probabilities are much more difficult. You will have met the

- *mean*  $\mu = E[X]$
- The *variance*  $\sigma^2 = E[(X - \mu)^2]$ .
- The parameter  $\sigma$  is known as the *standard deviation*.

The central moments are defined as

$$\mu_r = E[(X - \mu)^r] \quad r = 1, 2, \dots$$

The third and fourth moments  $E[(X - \mu)^3], E[(X - \mu)^4]$  are less commonly used. We can prove an interesting link between the mean  $\mu$  and the variance  $\sigma^2$ . The result is known as Chebyshev's inequality

$$P[|X - \mu| > \epsilon] \leq \left[ \frac{\sigma}{\epsilon} \right]^2 \quad (10.3)$$

This tells us that departure from the mean have small probability when  $\sigma$  is small.

### 10.1.2 Some Discrete Probability Distributions

We shall run through some of the most common and important discrete probability distributions.

#### The Discrete Uniform distribution

Suppose  $X$  can take one values  $1, 2, \dots, n$  with equal probability, that is

$$P[X = k] = \frac{1}{n} \quad k = 1, 2, \dots, n \quad (10.4)$$

- The mean is  $E[X] = \frac{n+1}{2}$
- the variance is  $\text{var}(X) = \frac{1}{3}n^3 + \frac{1}{4}n^2 + \frac{2}{3}n - \frac{1}{4}$

For example a die is thrown, the distribution of the score  $X$  is uniform on the integer 1 to 6.

#### The Binomial distribution

Suppose we have a series of trials each of which has two outcomes, success  $S$  and failure  $F$ . We assume that the probability of success,  $p$ , is constant, so for every trial

$$P[\text{Success}] = p \text{ and } P[\text{failure}] = 1 - p$$

the probability of  $X$  successes in  $n$  trials is given by

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n \quad (10.5)$$

- The mean is  $E[X] = np$
- the variance is  $\text{var}(X) = np(1 - p)$

The probability that a person will survive a serious blood disease is 0.4. If 15 people have the disease the number of survivors  $X$  has a Binomial  $B(15, 0.4)$  distribution.

- $P[X = 3] = \binom{15}{3} (0.4)^3 (0.6)^{12}$
- $P[X \leq 8] = \sum_{x=0}^8 \binom{15}{x} (0.4)^x (0.6)^{15-x}$
- $P[3 \leq X \leq 8] = P[X \leq 8] - P[X \leq 2] = \sum_{x=2}^8 \binom{15}{x} (0.4)^x (0.6)^{15-x}$

### Applying expectation using the Binomial

A more interesting use is:

Suppose we wish to test whether  $N$  people have a disease. It would seem that the only way to do this is to take a blood test, which will require  $N$  blood tests. Suppose we try the following:

1. We pool the blood of  $k < N$  people.
2. If the combined sample is negative we have  $k$  people without the disease.
3. If the pooled test is positive we then test all  $k$  people individually, resulting in  $k + 1$  tests in all.
4. Repeat until everyone is diagnosed

What does this save us?

I joined MITAS because  
I wanted **real responsibility**

The Graduate Programme  
for Engineers and Geoscientists  
[www.discovermitas.com](http://www.discovermitas.com)



**Month 16**

I was a construction  
supervisor in  
the North Sea  
advising and  
helping foremen  
solve problems

Real work  
International opportunities  
Three work placements





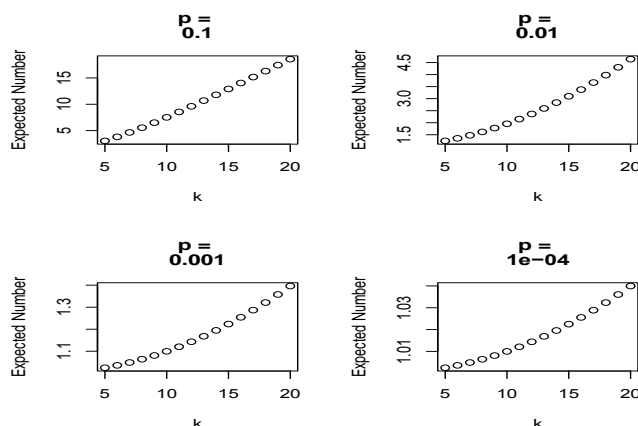
Assume the probability of a person having the disease is  $p$  and that we have a Binomial distribution for the number with the disease. Then for a group of  $k$

1.  $P[\text{just 1 test}] = (1 - p)^k$
2.  $P[k+1 \text{ tests}] = 1 - P[\text{just 1 test}] = 1 - (1 - p)^k$

So the expected number of tests is

$$E[\text{no. of tests}] = (1 - p)^k + (k + 1) (1 - (1 - p)^k) = k (1 - (1 - p)^k) + 1$$

This does give a considerable saving in the number of tests, see the diagram below



### The Hypergeometric distribution

Suppose we have  $N$  items and  $D$  of these are defective. I take a sample of size  $n$  from these items, then the probability that this sample contains  $k$  defectives is

$$P[X = k] = \frac{\binom{N-D}{n-k} \binom{D}{k}}{\binom{N}{n}} \quad k = 0, 1, 2, \dots, n \tag{10.6}$$

- The mean is  $E[X] = n \frac{D}{N}$
- the variance is  $\text{var}(X) = \frac{(N-n)}{(N-1)} n \frac{D}{N} (1 - \frac{D}{N})$

While situations involving the Hypergeometric are common it is common practice to approximate with the Binomial when  $N$  is large compared to  $D$ . We set  $p = D/N$  and sue

$$P[X = k] = \binom{n}{k} p^k (1 - p)^{n-k} \quad k = 0, 1, 2, \dots, n$$



### The Poisson distribution

Suppose events occur at random

$$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!} \quad k = 1, 2, \dots, n \quad (10.7)$$

- The mean is  $E[X] = \lambda$
- the variance is  $\text{var}(X) = \lambda$

The average number of oil tankers arriving per day at a port is 10. The facilities at the port can handle at most 15 arrivals in a day. What is the probability that the port will not be able to handle all the arrivals in a day? The variable  $X$  is Poisson  $\lambda = 10$  so

$$P[X \geq 16] = \sum_{x=16}^{\infty} \frac{10^x}{x!} \exp(-10) = 1 - \sum_{x=0}^{15} \frac{10^x}{x!} \exp(-10) = 1 - 0.9513$$

“I studied English for 16 years but...  
...I finally learned to speak it in just six lessons”  
Jane, Chinese architect

ENGLISH OUT THERE

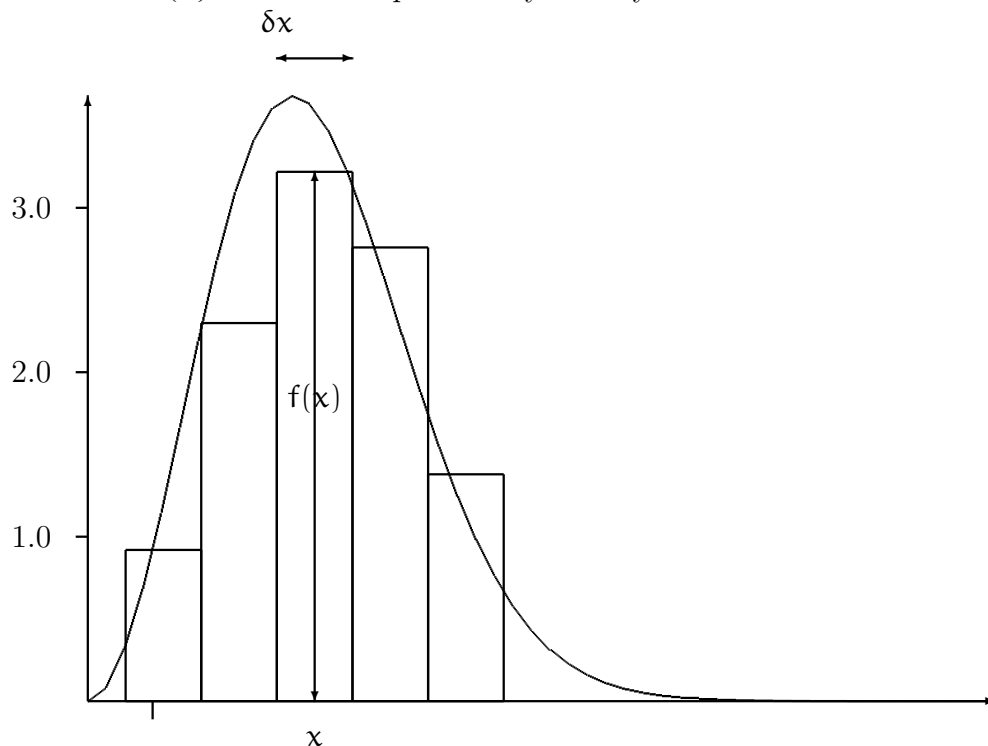
Click to hear me talking before and after my unique course download

### 10.1.3 Continuous variables

All the cases we have considered so far have been where  $X$  takes discrete values. This does not have to be true - we can imagine  $X$  taking a continuous set of values. Since we have thought of a probability at  $X=k$  we might think of the probability of  $X$  being in some small interval  $x, x + \delta x$ . This probability will be

$$P[x \leq X \leq x + \delta x] = f(x)\delta x$$

The function  $f(x)$  is called the probability density function.



The probability, as can be seen from the sketch is made up of boxes, and if we add these together we get a probability.

Personally I find it simpler to think of the cumulative distribution function  $F(x)$  which is defined as  $P[X \leq x] = F(x)$ . This is just a probability and is what you find in tables. We relate this to the density function by

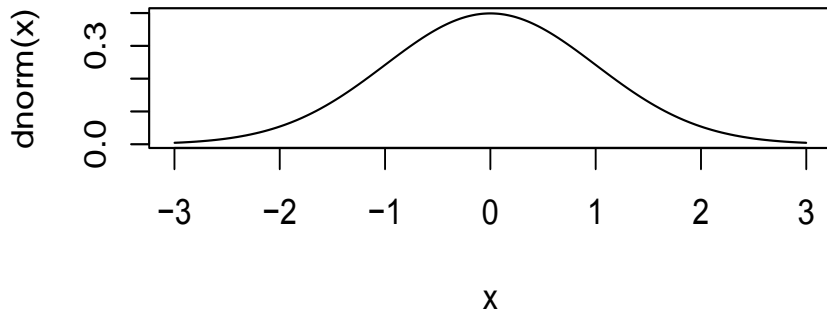
$$F(x) = \int_{-\infty}^x f(t) dt$$

It is then not difficult to show that

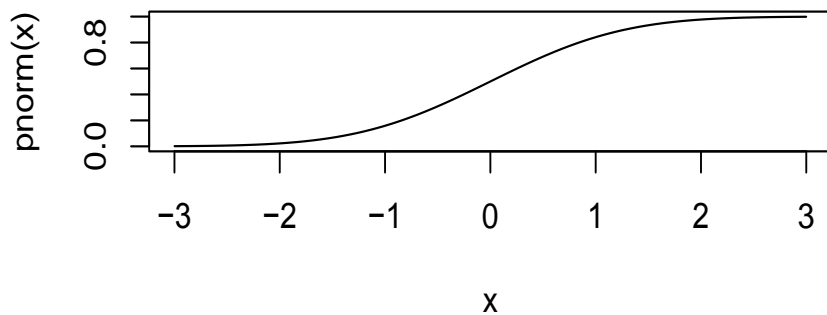
$$P[a \leq X \leq b] = \int_a^b f(t) dt$$

Typical shapes are

### density function



### distribution function



Excellent Economics and Business programmes at:



**university of  
 groningen**



**“The perfect start  
 of a successful,  
 international career.”**

**CLICK HERE**  
 to discover why both socially  
 and academically the University  
 of Groningen is one of the best  
 places for a student to be

[www.rug.nl/feb/education](http://www.rug.nl/feb/education)



## 10.1.4 Some Continuous Probability Distributions

### Uniform Distribution

Here  $X$  is uniformly distributed on a range, say  $(a, b)$  so

$$f(x) = \frac{1}{b-a} \quad (10.8)$$

It follows that  $F(x) = P[X \leq x] = \frac{x-a}{b-a}$  and  $P[c \leq X \leq d] = \frac{d-c}{b-a}$

- The mean is  $E[X] = \frac{a+b}{2}$
- the variance is  $\text{var}(X) = \frac{1}{12}(b-a)^2$

This is a useful model for a random choice in the interval from  $a$  to  $b$ .

### Exponential Distribution

Here  $X$  is distributed on the range  $(0, \infty)$  and

$$f(x) = \lambda \exp(-\lambda x) \quad (10.9)$$

where  $\lambda$  is a constant. It follows that  $F(x) = P[X \leq x] = 1 - \exp(-\lambda x)$  and  $P[c \leq X \leq d] = \exp(-\lambda c) - \exp(-\lambda d)$

- The mean is  $E[X] = \frac{1}{\lambda}$
- the variance is  $\text{var}(X) = \frac{1}{\lambda^2}$

### Normal Distribution

Here  $X$  is distributed on the range  $(-\infty, \infty)$  and

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \quad (10.10)$$

where  $\mu$  and  $\sigma$  are constants.

- The mean is  $E[X] = \mu$
- the variance is  $\text{var}(X) = \sigma^2$

The normal distribution crops up all over the place the problem is that there is no simple way of working out the probabilities. They can be computed but you either need the algorithm or tables.

### Normal Computation

Suppose  $X$  has a Normal distribution with mean  $\mu$  and Variance  $\sigma^2$ , often written  $N(\mu, \sigma^2)$ . We can show that  $X$  is related to a *Standard Normal* variable  $z$ , that is  $z$  is  $N(0,1)$  by

$$z = \frac{X - \mu}{\sigma} \quad (10.11)$$

And of course we have the reverse

$$X = \mu + \sigma \times z \quad (10.12)$$

Now the standard normal is what is given in the tables do we convert our problem into a standard one.

1. Suppose  $X$  is  $N(100, 9^2)$  Then

$$(a) \ P[X \leq 70] = P \left[ z = \frac{X-100}{9} \leq \frac{70-100}{9} \leq -3.33 \right] = 0.004$$

$$(b) \ P[X \leq 95] = P \left[ z = \frac{X-100}{9} \leq \frac{95-100}{9} \leq -5/9 \right] = 0.2893$$

$$(c) \ P[X \geq 109] = 1 - P \left[ z = \frac{X-100}{9} \geq \frac{109-100}{9} \geq 1 \right] = 1 - 0.2893$$

$$(d) \ P[70 \leq X \leq 109] = P[X \geq 109] - P[X \leq 70] = 0.7017 - 0.004$$

Suppose we wish to find the value  $a$  so that  $P[X \leq a] = 0.95$ . Then

$$P[X \geq a] = P \left[ z = \frac{X - 100}{9} \geq z = \frac{a - 100}{9} \right] = 0.9$$

From tables  $\frac{a-100}{9} = 1.645$  and so  $a = 100 + 1.645 \times 9$

### Another example

Suppose we know

$$1. \ P[X < 2] = 0.05$$

$$2. \ P[X > 14] = 0.25$$

So we have

$$P[X < 2] = P[z = (X - \mu)/\sigma < (2 - \mu)/\sigma] = 0.05$$

and so from tables  $(2 - \mu)/\sigma = -1.645$

We also have  $P[X > 14] = 0.25$  or

$$P[X < 14] = P[z < (X - \mu)/\sigma < (X - \mu)/\sigma] = 1 - 0.25 = 0.75$$

Hence  $(14 - \mu)/\sigma = 1.96$  We have a pair of equations

1.  $2 - \mu = -\sigma \times 1.645$
2.  $14 - \mu = \sigma \times 1.96$

Solving gives

$$(14 - \mu) - (2 - \mu) = 12 = 0.315\sigma$$

or  $\sigma = 3.32871$  and so  $\mu = 7.475728$

### The Normal approximation to the Binomial

A Binomial variable  $X$  which is  $B(n, p)$  can be approximated by a Normal variable  $Y$ , mean  $np$ , variance  $np(1 - p)$ . This can be very useful as the Binomial tables provided are not very extensive. *This is known as the Normal approximation to the Binomial.*

In this case

$$z = (Y - np) / \sqrt{np(1 - p)}$$

is standard Normal.

### Example

Suppose  $X$  is number of 6's in 40 rolls of a die. Let  $Y$  be  $N(\frac{40}{6}, \frac{40}{6}\frac{5}{6})$ . Then

$$P[X < 5] \simeq P[Y < 5] = P[z < \frac{5 - 20/3}{\sqrt{50/9}}] = \Phi(-0.7071068) = 0.2398$$

You can refine this approximation but we will settle for this at the moment.

### Exercises

1. A die is rolled, what is the probability that
  - (a) The outcome is even.
  - (b) The outcome is a prime.
  - (c) The outcome exceeds 2.
  - (d) The outcome is -1.
  - (e) The outcome is less than 12.
2. Two dice are rolled. What is the probability that
  - (a) The sum of the upturned faces is 7?
  - (b) The score on one die is exactly twice the score on the other.

- (c) You throw a double, that is the dice each have the same score.
3. Suppose we toss a coin 3 times. Find the probability distribution of
- (a)  $X$  = the number of tails.
- (b)  $Y$  = the number of runs. Here a run is a string of heads or tails. So for HTT  $Y=2$ .
4. The student population in the Maths department at the University of San Diego was made up as follows
- 10% were from California
  - 6% were of Spanish origin
  - 2% were from California and of Spanish origin.

If a student from the class was to be drawn at random what is the probability that they are

- (a) From California or of Spanish origin.
- (b) Neither from California nor of Spanish origin.
- (c) Of Spanish origin but not from California



**REGENT'S**  
UNIVERSITY LONDON

## Enhance your career opportunities

We offer practical, industry-relevant undergraduate and postgraduate degrees in central London

- > Accounting and finance
- > Business, management and leadership
- > Oil and gas trade management
- > Global banking and finance
- > Luxury brand management
- > Media communications and marketing

**Contact us to arrange a visit**  
**Apply direct for January or September entry**

T +44 (0)20 7487 7505   E [exrel@regents.ac.uk](mailto:exrel@regents.ac.uk)   W [regents.ac.uk](http://regents.ac.uk)



5. For two events A and B the following probabilities are known

$$P[A] = 0.52 \quad P[B] = 0.36 \quad P[A \cup B] = 0.68$$

Determine the probabilities

- (a)  $P[A \cap B]$
  - (b)  $P[\sim A]$
  - (c)  $P[\sim B]$
6. A hospital trust classifies a group of middle aged men according to body weight and the incidence of hypertension. The results are given in the table.

|                  | Overweight | Normal Weight | Underweight | Total |
|------------------|------------|---------------|-------------|-------|
| Hypertensive     | 0.10       | 0.08          | 0.02        | 0.20  |
| Not Hypertensive | 0.15       | 0.45          | 0.20        | 0.80  |
| Total            | 0.25       | 0.53          | 0.22        | 1.00  |

- (a) What is the probability that a person selected at random from this group will have hypertension?
  - (b) A person selected at random from this group is found to be overweight, what is the probability that this person is also hypertensive?
  - (c) Find  $P[\text{hypertensive} \cup \text{Underweight}]$
  - (d) Find  $P[\text{hypertensive} \cup \text{Not Underweight}]$
7. Two cards are drawn from an ordinary deck of 52 cards. What is the probability of drawing
- (a) Two aces.
  - (b) The two black aces.
  - (c) Two cards from the court cards K,Q,J
8. Five cards are drawn from a deck of cards. What is the chance that
- (a) Four cards are aces
  - (b) Four cards are the same i.e. 4 10's, 4 9's etc.
  - (c) All the cards are of the same suit.
  - (d) All the cards are of the same suit and are in sequence.



9. A student of statistics was told that there was a chance of 1 in a million that there was a bomb on an aircraft. The reasoned that there would be a one in  $10^{12}$  chance of being two bombs on a plane. He thus decided that he should take a bomb with him ( defused - he was not stupid) to reduce the odds of an explosion.

Assuming no security problems is this a sensible strategy?

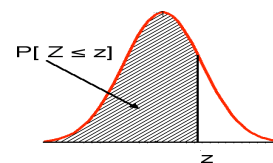
10. There are four tickets numbered 1,2,3,4. A two digit number is formed by drawing a ticket at random from the four and a second from the remaining three. So if the tickets were 4 and 1 the resulting number would be 41. What is the probability that
- The resulting number is even.
  - The resulting number exceeds 20
  - The resulting number is between 22 and 30.
11. Three production lines contribute to the total pool of parts used by a company.
- Line 1 contributes 20% and 15% of items are defective.
  - Line 2 contributes 50% and 5% of items are defective.
  - Line 3 contributes 30% and 6% of items are defective.
- What percentage of items in the pool are defective?
  - Suppose an item was selected at random and found to be defective, what is the probability that it came from line 1?
  - Suppose an item was selected at random and found not to be defective, what is the probability that it came from line 1?

## 10.2 The Normal distribution

This table gives the cumulative probabilities for the standard normal distribution, that is

$$P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$$

This is the shaded area in the figure.



| z    | 0.00   | -0.01  | -0.02  | -0.03  | -0.04  | -0.05  | -0.06  | -0.7   | -0.08  | -0.09  |
|------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| -3.4 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0002 |
| -3.3 | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0003 |
| -3.2 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 | 0.0005 |
| -3.1 | 0.0010 | 0.0009 | 0.0009 | 0.0009 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0007 | 0.0007 |
| -3.0 | 0.0013 | 0.0013 | 0.0013 | 0.0012 | 0.0012 | 0.0011 | 0.0011 | 0.0011 | 0.0010 | 0.0010 |
| -2.9 | 0.0019 | 0.0018 | 0.0018 | 0.0017 | 0.0016 | 0.0016 | 0.0015 | 0.0015 | 0.0014 | 0.0014 |
| -2.8 | 0.0026 | 0.0025 | 0.0024 | 0.0023 | 0.0023 | 0.0022 | 0.0021 | 0.0021 | 0.0020 | 0.0019 |
| -2.7 | 0.0035 | 0.0034 | 0.0033 | 0.0032 | 0.0031 | 0.0030 | 0.0029 | 0.0028 | 0.0027 | 0.0026 |
| -2.6 | 0.0047 | 0.0045 | 0.0044 | 0.0043 | 0.0041 | 0.0040 | 0.0039 | 0.0038 | 0.0037 | 0.0036 |
| -2.5 | 0.0062 | 0.0060 | 0.0059 | 0.0057 | 0.0055 | 0.0054 | 0.0052 | 0.0051 | 0.0049 | 0.0048 |
| -2.4 | 0.0082 | 0.0080 | 0.0078 | 0.0075 | 0.0073 | 0.0071 | 0.0069 | 0.0068 | 0.0066 | 0.0064 |
| -2.3 | 0.0107 | 0.0104 | 0.0102 | 0.0099 | 0.0096 | 0.0094 | 0.0091 | 0.0089 | 0.0087 | 0.0084 |
| -2.2 | 0.0139 | 0.0136 | 0.0132 | 0.0129 | 0.0125 | 0.0122 | 0.0119 | 0.0116 | 0.0113 | 0.0110 |
| -2.1 | 0.0179 | 0.0174 | 0.0170 | 0.0166 | 0.0162 | 0.0158 | 0.0154 | 0.0150 | 0.0146 | 0.0143 |
| -2.0 | 0.0228 | 0.0222 | 0.0217 | 0.0212 | 0.0207 | 0.0202 | 0.0197 | 0.0192 | 0.0188 | 0.0183 |
| -1.9 | 0.0287 | 0.0281 | 0.0274 | 0.0268 | 0.0262 | 0.0256 | 0.0250 | 0.0244 | 0.0239 | 0.0233 |
| -1.8 | 0.0359 | 0.0351 | 0.0344 | 0.0336 | 0.0329 | 0.0322 | 0.0314 | 0.0307 | 0.0301 | 0.0294 |
| -1.7 | 0.0446 | 0.0436 | 0.0427 | 0.0418 | 0.0409 | 0.0401 | 0.0392 | 0.0384 | 0.0375 | 0.0367 |
| -1.6 | 0.0548 | 0.0537 | 0.0526 | 0.0516 | 0.0505 | 0.0495 | 0.0485 | 0.0475 | 0.0465 | 0.0455 |
| -1.5 | 0.0668 | 0.0655 | 0.0643 | 0.0630 | 0.0618 | 0.0606 | 0.0594 | 0.0582 | 0.0571 | 0.0559 |
| -1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 | 0.0681 |
| -1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 | 0.0823 |
| -1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 | 0.0985 |
| -1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 | 0.1170 |
| -1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 | 0.1379 |
| -0.9 | 0.1841 | 0.1814 | 0.1788 | 0.1762 | 0.1736 | 0.1711 | 0.1685 | 0.1660 | 0.1635 | 0.1611 |
| -0.8 | 0.2119 | 0.2090 | 0.2061 | 0.2033 | 0.2005 | 0.1977 | 0.1949 | 0.1922 | 0.1894 | 0.1867 |
| -0.7 | 0.2420 | 0.2389 | 0.2358 | 0.2327 | 0.2296 | 0.2266 | 0.2236 | 0.2206 | 0.2177 | 0.2148 |
| -0.6 | 0.2743 | 0.2709 | 0.2676 | 0.2643 | 0.2611 | 0.2578 | 0.2546 | 0.2514 | 0.2483 | 0.2451 |
| -0.5 | 0.3085 | 0.3050 | 0.3015 | 0.2981 | 0.2946 | 0.2912 | 0.2877 | 0.2843 | 0.2810 | 0.2776 |
| -0.4 | 0.3446 | 0.3409 | 0.3372 | 0.3336 | 0.3300 | 0.3264 | 0.3228 | 0.3192 | 0.3156 | 0.3121 |
| -0.3 | 0.3821 | 0.3783 | 0.3745 | 0.3707 | 0.3669 | 0.3632 | 0.3594 | 0.3557 | 0.3520 | 0.3483 |
| -0.2 | 0.4207 | 0.4168 | 0.4129 | 0.4090 | 0.4052 | 0.4013 | 0.3974 | 0.3936 | 0.3897 | 0.3859 |
| -0.1 | 0.4602 | 0.4562 | 0.4522 | 0.4483 | 0.4443 | 0.4404 | 0.4364 | 0.4325 | 0.4286 | 0.4247 |
| 0.0  | 0.5000 | -      | -      | -      | -      | -      | -      | -      | -      | -      |

This table gives the cumulative probabilities for the standard normal distribution, that is

$$P[Z \leq z] = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) dx$$

This is the shaded area in the figure.

| z   | 0.00   | 0.01   | 0.02   | 0.03   | 0.04   | 0.05   | 0.06   | 0.07   | 0.08   | 0.09   |
|-----|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| 0.0 | 0.5000 | 0.5040 | 0.5080 | 0.5120 | 0.5160 | 0.5199 | 0.5239 | 0.5279 | 0.5319 | 0.5359 |
| 0.1 | 0.5398 | 0.5438 | 0.5478 | 0.5517 | 0.5557 | 0.5596 | 0.5636 | 0.5675 | 0.5714 | 0.5753 |
| 0.2 | 0.5793 | 0.5832 | 0.5871 | 0.5910 | 0.5948 | 0.5987 | 0.6026 | 0.6064 | 0.6103 | 0.6141 |
| 0.3 | 0.6179 | 0.6217 | 0.6255 | 0.6293 | 0.6331 | 0.6368 | 0.6406 | 0.6443 | 0.6480 | 0.6517 |
| 0.4 | 0.6554 | 0.6591 | 0.6628 | 0.6664 | 0.6700 | 0.6736 | 0.6772 | 0.6808 | 0.6844 | 0.6879 |
| 0.5 | 0.6915 | 0.6950 | 0.6985 | 0.7019 | 0.7054 | 0.7088 | 0.7123 | 0.7157 | 0.7190 | 0.7224 |
| 0.6 | 0.7257 | 0.7291 | 0.7324 | 0.7357 | 0.7389 | 0.7422 | 0.7454 | 0.7486 | 0.7517 | 0.7549 |
| 0.7 | 0.7580 | 0.7611 | 0.7642 | 0.7673 | 0.7704 | 0.7734 | 0.7764 | 0.7794 | 0.7823 | 0.7852 |
| 0.8 | 0.7881 | 0.7910 | 0.7939 | 0.7967 | 0.7995 | 0.8023 | 0.8051 | 0.8078 | 0.8106 | 0.8133 |
| 0.9 | 0.8159 | 0.8186 | 0.8212 | 0.8238 | 0.8264 | 0.8289 | 0.8315 | 0.8340 | 0.8365 | 0.8389 |
| 1.0 | 0.8413 | 0.8438 | 0.8461 | 0.8485 | 0.8508 | 0.8531 | 0.8554 | 0.8577 | 0.8599 | 0.8621 |
| 1.1 | 0.8643 | 0.8665 | 0.8686 | 0.8708 | 0.8729 | 0.8749 | 0.8770 | 0.8790 | 0.8810 | 0.8830 |
| 1.2 | 0.8849 | 0.8869 | 0.8888 | 0.8907 | 0.8925 | 0.8944 | 0.8962 | 0.8980 | 0.8997 | 0.9015 |
| 1.3 | 0.9032 | 0.9049 | 0.9066 | 0.9082 | 0.9099 | 0.9115 | 0.9131 | 0.9147 | 0.9162 | 0.9177 |
| 1.4 | 0.9192 | 0.9207 | 0.9222 | 0.9236 | 0.9251 | 0.9265 | 0.9279 | 0.9292 | 0.9306 | 0.9319 |
| 1.5 | 0.9332 | 0.9345 | 0.9357 | 0.9370 | 0.9382 | 0.9394 | 0.9406 | 0.9418 | 0.9429 | 0.9441 |
| 1.6 | 0.9452 | 0.9463 | 0.9474 | 0.9484 | 0.9495 | 0.9505 | 0.9515 | 0.9525 | 0.9535 | 0.9545 |
| 1.7 | 0.9554 | 0.9564 | 0.9573 | 0.9582 | 0.9591 | 0.9599 | 0.9608 | 0.9616 | 0.9625 | 0.9633 |
| 1.8 | 0.9641 | 0.9649 | 0.9656 | 0.9664 | 0.9671 | 0.9678 | 0.9686 | 0.9693 | 0.9699 | 0.9706 |
| 1.9 | 0.9713 | 0.9719 | 0.9726 | 0.9732 | 0.9738 | 0.9744 | 0.9750 | 0.9756 | 0.9761 | 0.9767 |
| 2.0 | 0.9772 | 0.9778 | 0.9783 | 0.9788 | 0.9793 | 0.9798 | 0.9803 | 0.9808 | 0.9812 | 0.9817 |
| 2.1 | 0.9821 | 0.9826 | 0.9830 | 0.9834 | 0.9838 | 0.9842 | 0.9846 | 0.9850 | 0.9854 | 0.9857 |
| 2.2 | 0.9861 | 0.9864 | 0.9868 | 0.9871 | 0.9875 | 0.9878 | 0.9881 | 0.9884 | 0.9887 | 0.9890 |
| 2.3 | 0.9893 | 0.9896 | 0.9898 | 0.9901 | 0.9904 | 0.9906 | 0.9909 | 0.9911 | 0.9913 | 0.9916 |
| 2.4 | 0.9918 | 0.9920 | 0.9922 | 0.9925 | 0.9927 | 0.9929 | 0.9931 | 0.9932 | 0.9934 | 0.9936 |
| 2.5 | 0.9938 | 0.9940 | 0.9941 | 0.9943 | 0.9945 | 0.9946 | 0.9948 | 0.9949 | 0.9951 | 0.9952 |
| 2.6 | 0.9953 | 0.9955 | 0.9956 | 0.9957 | 0.9959 | 0.9960 | 0.9961 | 0.9962 | 0.9963 | 0.9964 |
| 2.7 | 0.9965 | 0.9966 | 0.9967 | 0.9968 | 0.9969 | 0.9970 | 0.9971 | 0.9972 | 0.9973 | 0.9974 |
| 2.8 | 0.9974 | 0.9975 | 0.9976 | 0.9977 | 0.9977 | 0.9978 | 0.9979 | 0.9979 | 0.9980 | 0.9981 |
| 2.9 | 0.9981 | 0.9982 | 0.9982 | 0.9983 | 0.9984 | 0.9984 | 0.9985 | 0.9985 | 0.9986 | 0.9986 |
| 3.0 | 0.9987 | 0.9987 | 0.9987 | 0.9988 | 0.9988 | 0.9989 | 0.9989 | 0.9989 | 0.9990 | 0.9990 |
| 3.1 | 0.9990 | 0.9991 | 0.9991 | 0.9991 | 0.9992 | 0.9992 | 0.9992 | 0.9992 | 0.9993 | 0.9993 |
| 3.2 | 0.9993 | 0.9993 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9994 | 0.9995 | 0.9995 | 0.9995 |
| 3.3 | 0.9995 | 0.9995 | 0.9995 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9996 | 0.9997 |
| 3.4 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9997 | 0.9998 |