Lecture Notes for Module MATH237

ENGINEERING MATHEMATICS AND STATISTICS

Lauren Ansell

School of Computing, Electronics and Mathematics
University of Plymouth
Drake Circus
Plymouth
Devon
PL4 8AA
Email: lauren.ansell@plymouth.ac.uk, room: 007, 2-5 Kirkby place

Reference Material

Mathematics

- James, G. (2011).
  Advanced Modern Engineering Mathematics.
  Addison Wesley.

- Croft, A., Davison, R. and Hargreaves, M. (1992).
  Engineering Mathematics. A modern foundation for electronic, electrical and control engineers.
  Addison Wesley.

- Bolton, W. (1994).
  Laplace & z-Transforms.
  Longman Scientific and Technical.

- Stroud, K. A. (1990).
  Further Engineering Mathematics.
  MacMillan.

- Bracewell, R.V. (1986)
  The Fourier Transform and its Applications
  McGraw-Hill

<u>You should also consult the library, specifically Section 510.2462</u>

Statistics

- Oakland, J.S. (2003)
  Statistical Process Control
  Butterworth-Heinemann

- Rosenkrantz, W.A. (2009)
  Introduction to Probability and Statistics for Science, Engineering and Finance
  Chapman and Hill

- Mann, P.S. (2007)
  Introductory Statistics
  Wiley

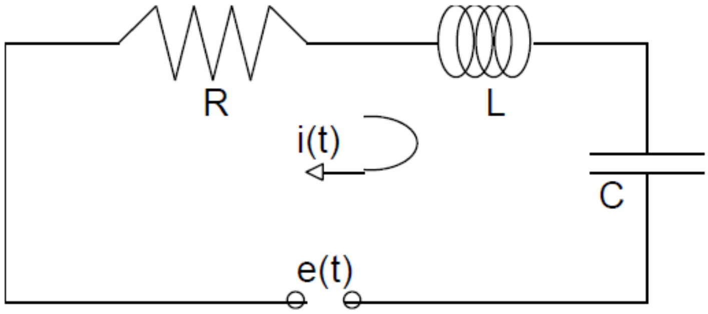
- Montgomery, D.C and Runger, G.C. (2007)
  Applied Statistics and Probability for Engineers
  Wiley

<u>You should also consult the library, specifically Section 519.5</u>

LAPLACE TRANSFORMS

Motivation (Why bother with Laplace transforms?)

The simplest answer is that they make the analysis of linear, time-invariant systems much easier. For example, suppose we want to find the output (the current) i(t) flowing in the circuit shown below if, at time t = 0, a voltage e(t) is applied. Assume that there is no current flowing for t < 0, and that there is no charge on the capacitor for t < 0.



The conventional method of solution is to start by writing down the equations governing the system (Kirchhoff's Laws):

$$L\frac{di}{dt} + Ri + \frac{q}{C} = e, \quad i = \frac{dq}{dt}.$$

Next, we have to solve this pair of simultaneous ordinary differential equations (not trivial) and find the two arbitrary constants by imposing conditions at t = 0. One problem with this approach is that conventional methods of solving the differential equations break down if e(t) is not 'smooth'. For example if e(t) jumps from zero volts for t < 0 to 10 volts for t > 0 then we can't differentiate e(t) at t = 0 and this can cause problems in the solution of the differential equation. Laplace transforms can easily overcome this problem. As a bonus, they also allow the initial conditions at t = 0 to be built in easily.

There are further benefits to be gained from using Laplace transforms when considering system response and stability, as we shall see later.

To proceed we need to build up a set of results so that a range of problems can be tackled. We do this by using some basic rules for Laplace transforms and by making use of tables of known results. In this course the emphasis will be placed on use of tables. You will be expected to be able to find the Laplace transforms of some simple signals directly from the integral definition but this is not what the course is really concerned with.

The main objective is for you to be able to solve problems arising elsewhere in your degree course using the Laplace transform.

The Laplace transform is a mathematical device which converts causal signals, i.e. functions of time t defined for t > 0, into new functions of a new variable s. For example if the signal is e^{-3t}, t > 0, then, as we shall see later, the Laplace transform turns out to be $\frac{1}{s+3}$.

Notation

If v(t) is the signal then its Laplace transform is denoted by V(s). The convention we shall stick to is that signals are denoted by lower case letters (e.g. "v") and their transforms by the corresponding upper case letter ("V"). In the example above we have $v(t) = e^{-3t}$, $V(s) = \frac{1}{s+3}$.
In general,

V(s) = Laplace transform of v(t),

= *L*[v(t)],

where the symbol *L* is short-hand for "the Laplace transform of the signal enclosed in brackets".

Definition

The Laplace transform of a signal v(t) is defined by

$$V(s) = \int_0^\infty v(t)e^{-st}dt.$$

This is an example of an integral transform. It transforms functions in the time domain to functions in the s domain.

Example

Find the Laplace transform of a causal d.c. signal v(t)=v_0 for t>0

$$L[v_0] = \int_0^\infty v_0 e^{-st}\,dt = v_0 \int_0^\infty e^{-st}\,dt = v_0\left[\frac{-e^{-st}}{s}\right]_0^\infty = v_0\left[\left(\frac{-e^{-s\times\infty}}{s}\right) - \left(-\frac{e^{-s\times 0}}{s}\right)\right]$$

$$= v_0\left[0 - \frac{-1}{s}\right] = \frac{v_0}{s}$$

Example

Find the Laplace transform of e^{-at}

Table of Standard Laplace Transforms

v(t)	t > 0	V(s)
$\delta(t)$	*unit impulse*	1
$\delta(t-T)$	*delayed impulse*	e^{-Ts}
e^{-at}		$\frac{1}{s+a}$
$\frac{1}{(n-1)!}t^{n-1}e^{-at}$	n = 1, 2, 3, . . .	$\frac{1}{(s+a)^n}$
H(t) or 1	*unit step*	$\frac{1}{s}$
$H(t-T)$	*delayed step*	$\frac{1}{s}e^{-Ts}$
$H(t) - H(t-T)$	*rectangular pulse*	$\frac{1}{s}\left(1-e^{-Ts}\right)$
$\frac{1}{a}\left(1-e^{-at}\right)$		$\frac{1}{s(s+a)}$
t	*unit ramp*	$\frac{1}{s^2}$
$\frac{1}{a^2}\left(at-1+e^{-at}\right)$		$\frac{1}{s^2(s+a)}$
$\frac{t^{n-1}}{(n-1)!} \quad (0!=1)$		$\frac{1}{s^n} \quad n=1,2,3,\ \ldots$
$\sin \omega t$		$\frac{\omega}{s^2+\omega^2}$
$\cos \omega t$		$\frac{s}{s^2+\omega^2}$
t^n (n a positive integer)		$n!/s^{n+1}$
sinh at		$a/(s^2-a^2)$
cosh at		$s/(s^2-a^2)$
$\sin(\omega t+\phi)$		$\frac{s\sin\phi+\omega\cos\phi}{s^2+\omega^2}$
$L\left[\frac{d^{(n)}v}{dt^{(n)}}\right]$		$s^nV(s)-s^{n-1}v(0)-s^{n-2}v'(0)\ \ldots -v^{(n-1)}(0)$
$e^{-at}v(t)$	*s-shift*	$V(s+a)$
$v(t-T)\,H(t-T)$	*time-shift*	$e^{-sT}V(s)$
v(at)	*time-scaling*	$(1/a)V(s/a)$
v(t) * g(t)	*convolution*	$V(s)\times G(s)$

Example

Use the Table of Standard Laplace Transforms to find the following:

i) $L[t^5]$

ii) $L[\sin(2t)]$

iii) $L[e^{-5t}]$

iv) $L[\cos(6t)]$

v) $L[\sinh(4t)]$

vi) $L[t^7]$

vii) $L[e^{2t}]$

viii) $L\left[\sin\left(t\sqrt{3}\right)\right]$

ix) $L[\mathrm{H}(t-4)]$

x) $L[\delta(t-5)]$

Properties of Laplace Transforms

In this section we will be looking at the linear property of the Laplace transform, the S-shift theorem (or first shift theorem) and the Laplace transform of a derivative.

Linearity

$$L[a_1 v_1(t) + a_2 v_2(t)] = a_1 L[v_1(t)] + a_2 L[v_2(t)].$$

<u>Example</u>

Find the LT of $v(t) = e^{-5t} + \cos(2t)$

$$\mathrm{L}[e^{-5t} + \cos(2t)] \;\; = L[e^{-5t}] + L[\cos(2t)] = \frac{1}{s+5} + \frac{s}{s^2 + 2^2}$$

<u>Example</u>

Find the LT of $v(t) = 3 + 7e^{-2t} + 4t$

S- Shift Theorem (or First Shift Theorem)

If $\quad L[v(t)] = V(s)$

then $\quad L[v(t)e^{-at}] = V(s+a).$

Proof

$$L[v(t)e^{-at}] = \int_0^\infty v(t)e^{-at}e^{-st}dt = \int_0^\infty v(t)e^{-t(s+a)}dt = V(s+a)$$

Here V(s + a) is the function obtained from V(s) when **<u>all</u>** occurrences of s are replaced by $(s\, +\, a)$. This theorem is very useful because it allows us to find the Laplace transform of a new signal (the original multiplied by an exponential) very easily. We shall see later that the theorem helps when it comes to <u>inverting</u> Laplace transforms, i.e. finding the time signal that corresponds to a given Laplace transform.

<u>Example</u>

What is $L[e^{-2t}t^3]$?

Table of Laplace transforms gives $\quad L[t^n] = \frac{n!}{s^{n+1}}$ so we have $\; L[t^3] = \frac{6}{s^4}$

We have then to replace s with $s + 2$

$$L[e^{-2t}t^3] = \frac{6}{(s+2)^4}$$

Example

What is $L[e^{-4t}\sin(3t)]$?

Laplace Transform of a Derivative

$$L\left[\frac{dv}{dt}\right] = \int_0^\infty e^{-st}\frac{dv}{dt}dt.$$

Proof

Using integration by parts:

$$u = e^{-st},\ \ \frac{du}{dt} = -se^{-st}, \qquad \frac{dv}{dt} = \frac{dv}{dt}\ ,\ \ v = v(t)$$

$$L\left[\frac{dv}{dt}\right] = [e^{-st}v(t)]_0^\infty - \int_0^\infty -v(t)se^{-st}dt$$

$$= \lim_{t\to\infty}[e^{-st}v(t)] - v(0) + s\int_0^\infty v(t)e^{-st}dt$$

$$= -v(0) + sV(s)$$

This result can be extended to second derivatives:

$$L\left[\frac{d^2v}{dt^2}\right] = s^2V(s) - sv(0) - v'(0)$$

and n[th] derivatives:

$$L\left[\frac{d^{(n)}v}{dt^{(n)}}\right] = s^nV(s) - s^{n-1}v(0) - s^{n-2}v'(0) \dots - v^{(n-1)}(0)$$

Example

Let v(t) = sin(3t). Find *L*[cos(3t)] using the above result.

$$\cos(3t) = \frac{1}{3}\frac{d}{dt}\sin(3t)$$

$$L[\sin(3t)] = \frac{3}{s^2+9}$$

$$L[\cos(3t)] = \frac{1}{3}\left(s \times \frac{3}{s^2+9} - \sin(0)\right) = \frac{s}{s^2+9}$$

Example

Solve the differential equation $v''(t) + v(t) = 1$, with the initial conditions: $v(0) = 0, v'(0) = 0$

$$s^2V(s) - sv(0) - v'(0) + V(s) = L[1] = \frac{1}{s}$$

$$s^2V(s) + V(s) = \frac{1}{s}$$

$$\mathrm{V}(s) = \frac{1}{s(s^2+1)}$$

To find the solution, v(t), we have to find the signal that has a Laplace transform equal to $\frac{1}{s(s^2+s)}$. This process is called <u>inverting the Laplace transform</u> and is probably the trickiest part of the method. We need to build up a technique for inverting Laplace transforms.

<u>Example</u>

Solve the differential equation $v''(t) + 5v'(t) = e^{-t}$, with the initial conditions: $v(0) = 1, v'(0) = 2$

<u>Key Points</u>

- V(s) can be found by evaluating the integral $\int_0^\infty v(t)e^{-st}dt$.

- We can evaluate the integral for simple signals but we mainly use tables of Laplace transforms.

- Laplace transforms simplify the analysis of LTI systems and can be used to determine system characteristics (e.g. system stability).

- Laplace transforms are useful for solving certain types of ordinary differential equations.
- Linearity – the Laplace transform of a linear combination of signals is the corresponding linear combination of the Laplace transforms of the individual signals.

- First Shift theorem (s Shift theorem): if L[v(t)] = V(s) then $L[e^{-at}$ v(t)] = V(s + a).

- Laplace transforms of derivatives:
$$L\left[\frac{dv}{dt}\right] = sV(s) - v(0), \qquad L\left[\frac{d^2v}{dt^2}\right] = s^2V(s) - sv(0) - v'(0).$$

- To solve problems using Laplace transforms we have to be able to invert the transforms, i.e. we have to be able to find the signal that corresponds to a given transform.

- The Laplace transform of a signal, v(t), is a new function, V(s), that depends on a new variable, s.

Inversion of Laplace Transforms

When we solve a problem using Laplace transforms we usually reach a point where we know the Laplace transform of the solution but want to find the signal to which it corresponds. We have to perform an inverse Laplace transform, from the s domain to the t domain. This process is denoted by L^{-1}:

$$v(t) = L^{-1}[V(s)].$$

The technique we use is to try to write V(s) in a form that can be found in the tables and then just read off the answer.

<u>Example</u>

i) $V(s) = \frac{3}{s^2+9}$

ii) $V(s) = \frac{8}{s+3}$

<u>Example</u>

If $V(s) = \frac{1}{s^2+5^2}$, what is v(t)?

We spot that $L[\sin(\omega t)] = \frac{\omega}{s^2+\omega^2}$

$$L[\sin(5t)] = \frac{5}{s^2+5^2}$$
$$V(s) = \frac{1}{5}\left(\frac{5}{s^2+5^2}\right)$$
$$v(t) = \frac{1}{5}\sin(5t)$$

<u>Example</u>

If $V(s) = \frac{3}{s^2+4}$, what is v(t)?

There are two mathematical techniques that are very useful when trying to invert Laplace transforms: (a) partial fractions, and (b) completing the square.

<u>Example</u>

Find $L^{-1}\left[\frac{1}{(s+3)(s-2)}\right]$.

Represent this function as sum of partial fractions

$$\frac{1}{(s+3)(s-2)} = \frac{A}{(s+3)} + \frac{B}{(s-2)}$$

Cover up rule:

$$1 = A(s-2) + B(s+3)$$

Let s=2 $\quad 1 = B(2+3) = 5B \quad B = \frac{1}{5}$

Let s=-3 $\quad 1 = A(-3-2) = -5A \quad A = \frac{-1}{5}$

$$\frac{1}{(s+3)(s-2)} = \frac{-1}{5}\frac{1}{(s+3)} + \frac{1}{5}\frac{1}{(s-2)}$$

$$x(t) = \frac{1}{5}(e^{2t} - e^{-3t})$$

Equate coefficients:

$$1 = A(s-2) + B(s+3)$$

Linear s: $0 = A + B \quad A = -B$

Constants: $1 = -2A + 3B$

Substitute in A=-B, $\quad 1 = 2B + 3B = 5B$

$$B = \frac{1}{5} \text{ and } A = \frac{-1}{5}$$

$$\frac{1}{(s+3)(s-2)} = \frac{-1}{5}\frac{1}{(s+3)} + \frac{1}{5}\frac{1}{(s-2)}$$

$$x(t) = \frac{1}{5}(e^{2t} - e^{-3t})$$

Example

Find $L^{-1}\left[\frac{1}{s^2(s^2+4)}\right]$.

Example

Find $L^{-1}\left[\frac{1}{(s-3)(s+2)^2}\right]$.

Example

Find inverse Laplace transform x(t) of $X(s) = \frac{s}{(s+3)^2+4}$.

The first step here is to rewrite s as $s = s + 3 - 3$, due to the bracket in the denominator which shows that the S-shift (or first shift) theorem has been used.

$$X(s) = \frac{s+3-3}{(s+3)^2+4}$$

The fraction is then split into two fractions, making sure that if we have $s+3$ in the denominator, we also have $s+3$ in the numerator.

$$X(s) = \frac{s+3}{(s+3)^2+4} - \frac{3}{(s+3)^2+4}$$

$$x(t) = L^{-1}[X(s)] = L^{-1}\left[\frac{s+3}{(s+3)^2+4}\right] - L^{-1}\left[\frac{3}{(s+3)^2+4}\right]$$

$$x(t) = e^{-3t}\left(\cos(2t) - \frac{3}{2}\sin(2t)\right)$$

Example

Find $L^{-1}\left[\frac{s}{s^2+8s+25}\right]$.

We have to complete the squares in the denominator.

$$s^2 + 8s + 25 = (s+4)^2 - 16 + 25 = (s+4)^2 + 9$$

$$x(t) = L^{-1}\left[\frac{s}{(s+4)^2+9}\right]$$

As in the previous example the denominator shows that the S-shift theorem has been used.

$$x(t) = L^{-1}\left[\frac{s+4-4}{(s+4)^2+9}\right]$$

The problem now is to find inverse LT of

$$L^{-1}\left[\frac{(s+4)-4}{(s+4)^2+9}\right] = L^{-1}\left[\frac{s+4}{(s+4)^2+9}\right] - L^{-1}\left[\frac{4}{(s+4)^2+9}\right]$$

$$x(t) = e^{-4t}\left(\cos(3t) - \frac{4}{3}\sin(3t)\right)$$

Example

Find $L^{-1}\left[\frac{5s}{s^2-16s+68}\right]$

N.B. All of the above solutions are valid for t > 0. We assume that they are zero for t < 0(this can be proved).

Key Points

- The process by which we find the signal v(t) corresponding to a Laplace transform V(s) is called “inverting the Laplace transform”, or “inversion” for short.

- We try to use tables to invert Laplace transforms but often the Laplace transform we are given doesn’t appear in the table.

- Two mathematical techniques are useful for manipulating Laplace transforms, namely (a) partial fractions and (b) completing the square.

- The inversion process often involves use of the first shift theorem.

Solution of Ordinary Differential Equations

We use the Laplace transform to solve Ordinary Differential Equations (ODEs). The method for solving ODE's is based on the Laplace transform for derivatives.

Recall: Laplace transforms of derivatives:

$$L\left[\frac{dv}{dt}\right] = sV(s) - v(0), \qquad L\left[\frac{d^2v}{dt^2}\right] = s^2V(s) - sv(0) - v'(0).$$

Example

Solve the differential equation $\frac{d^2y}{dt^2} + 4\frac{dy}{dt} + 5y = 10e^t$ with the initial conditions $y'(0) = 2,\ y(0) = 1$.

Apply the Laplace transform to all the terms (L[y]=Y(s))

$$s^2Y(s) - y'(0) - sy(0) + 4(sY(s) - y(0)) + 5Y(s) = 10L[e^t] = \frac{10}{s-1}$$

Substitute initial conditions

$$s^2Y(s) - 2 - s \times 1 + 4(sY(s) - 1) + 5Y(s) = \frac{10}{s-1}$$

$$s^2Y(s) - 2 - s + 4sY(s) - 4 + 5Y(s) = \frac{10}{s-1}$$

$$Y(s^2 + 4s + 5) - 6 - s = \frac{10}{s-1}$$

$$Y(s) = \frac{10}{(s-1)(s^2+4s+5)} + \frac{s+6}{s^2+4s+5}$$

Break the first fraction into partial fractions

$$\frac{10}{(s-1)(s^2+4s+5)} = \frac{A}{s-1} + \frac{Bs+C}{s^2+4s+5}$$

$$10 = A(s^2 + 4s + 5) + (Bs + C)(s - 1)$$

Quadratic in s: $0 = A + B \quad A = -B$
Linear in s: $0 = 4A - B + C$
Constants: $10 = 5A - C$

$$0 = 5A + C \quad C = -5A$$
$$10 = 5A + 5A = 10A \quad A = 1$$
$$B = -1, \qquad C = -5$$

$$\frac{1}{s-1} - \frac{s+5}{s^2+4s+5}$$

$$Y(s) = \frac{1}{s-1} - \frac{s+5}{s^2+4s+5} + \frac{s+6}{s^2+4s+5} = \frac{1}{s-1} + \frac{-s-5+s+6}{s^2+4s+5} = \frac{1}{s-1} + \frac{1}{(s+2)^2+1}$$

This now has a form which can be easily inversed.

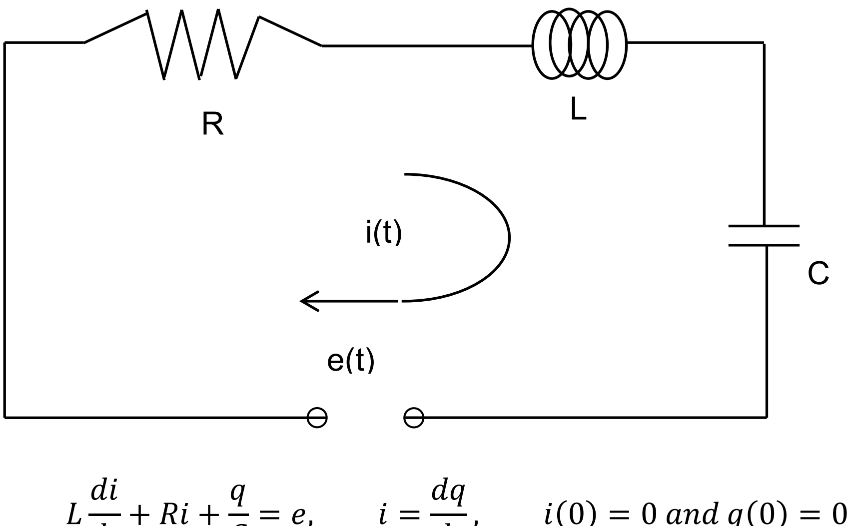
$$y(t) = e^t + e^{-2t}\sin t$$

Example

Solve the differential equation $\frac{d^2y}{dt^2} - 2\frac{dy}{dt} + y = 4$ with the initial conditions $y'(0) = 1,\ y(0) = 1.$

Example

Find the current flowing in the circuit below if there is no current flowing (or charge on the capacitor) for t < 0



$$L\frac{di}{dt} + Ri + \frac{q}{C} = e, \qquad i = \frac{dq}{dt}, \qquad i(0) = 0 \ and\ q(0) = 0$$

$$I(s) = \frac{sE(s)}{Ls^2 + Rs + \frac{1}{C}}$$

Suppose $e(t)=v_0$ (i.e. a constant voltage for $t>0$)

$$L[v_0] = \frac{v_0}{s} = E(s), \qquad I(s) = \frac{v_0}{Ls^2 + Rs + \frac{1}{C}}$$

Suppose: $v_0=10$ volts, $L=1$H, $C=10^{-4}$F and $R=160\Omega$

$$I(s) = \frac{10}{s^2 + 160s + 10^4}$$

$$I(s) = \frac{10}{(s+80)^2 + 60^2}$$

$$i(t) = L^{-1}[I(s)] = \frac{1}{6}e^{-80t}\sin(60t)$$

Example

Find the output position for a single mass with damper with input force F(t) and output position x(t), where $F(t) = \delta(t)$, m=1kg and d=4 N/m/s, subject to zero initial conditions.

Equations of motion based on Newton's second law

$$m\, x''(t) = F(t) - d\, x'(t)$$

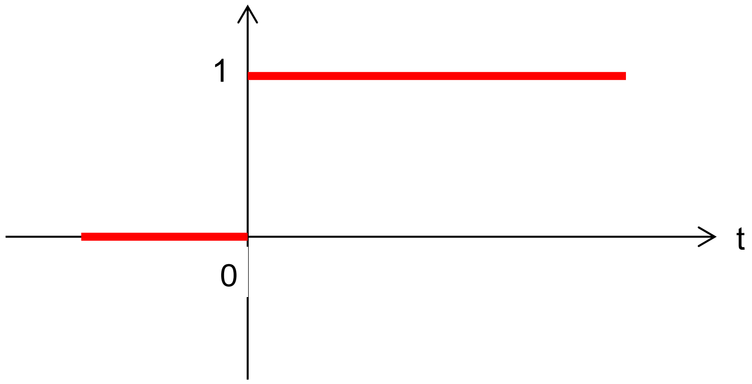
Key Points

- Laplace transforms convert linear ordinary differential equations into algebraic equations (i.e. equations where there are no derivatives).
- Initial conditions are easy to build in.
- Systems with jumps (discontinuities) in the input can be handled.
- Inversion is often the most difficult step in finding a solution.

Laplace Transform of Engineering Functions

Step Functions

The Heaviside function, H(t), is defined by $H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$

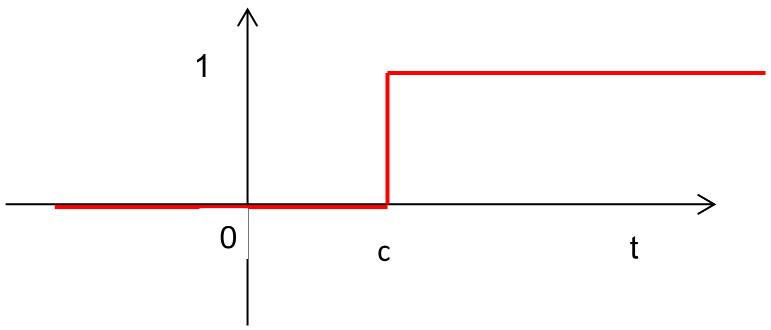


The Laplace transform of Heaviside function:

$$L[H(t)] = \int_0^\infty H(t)e^{-st}dt = \int_0^\infty 1 \times e^{-st}dt$$

$$= \left[\frac{e^{-st}}{-s}\right]_0^\infty = \lim_{t\to\infty}\frac{-e^{-st}}{s} - \frac{-1}{s} = \frac{1}{s}$$

The delayed Heaviside function H(t – c), where c is a constant, has the form shown below



and is defined by $H(t-c) = \begin{cases} 0, & t < c \\ 1, & t \geq c \end{cases}$

The Laplace transform of a delayed Heaviside function:

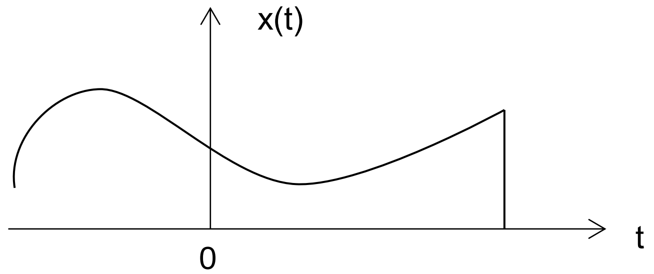
$$L[H(t-c)] = \int_0^\infty H(t-c)e^{-st}dt = \int_c^\infty 1 \times e^{-st}dt$$

$$= \left[\frac{e^{-st}}{-s}\right]_c^\infty = \lim_{t\to\infty}\frac{-e^{-st}}{s} - \frac{-e^{-sc}}{s} = \frac{e^{-sc}}{s}$$

Delayed Signals

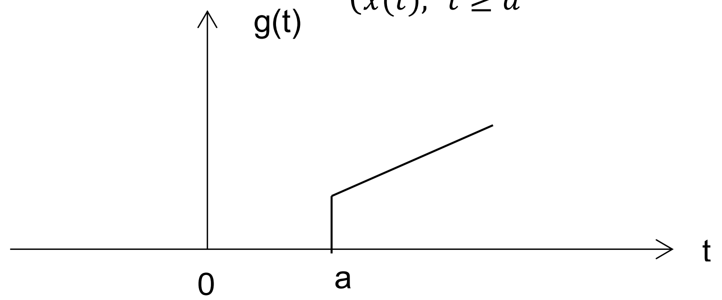
Use of H(t) to switch-on/off signals

Suppose we have a signal x(t) whose graph is shown below



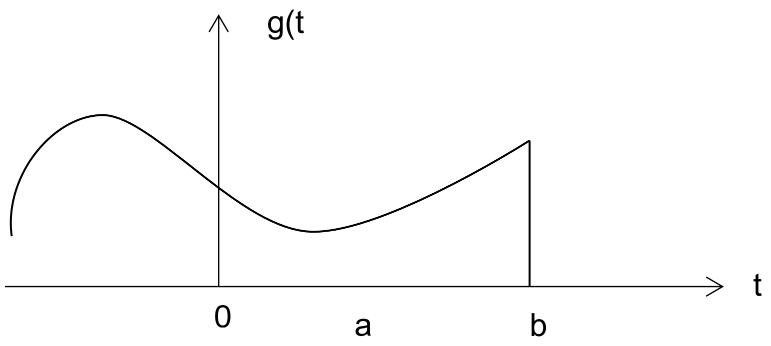
(a) Turn-on at t = a

$$g(t) = x(t)H(t-a) = \begin{cases} 0, & t < a \\ x(t), & t \geq a \end{cases}$$



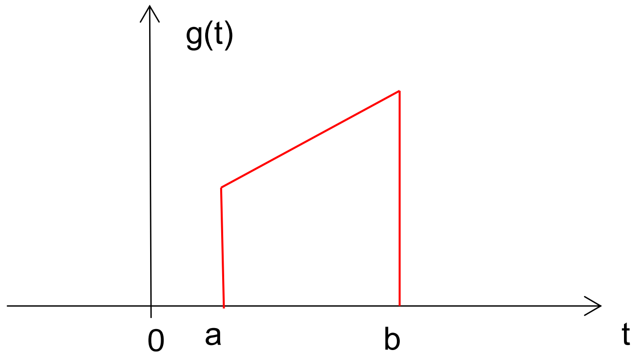
(b) Turn-off at t = b

$$g(t) = x(t)[1 - H(t-b)] = \begin{cases} x(t), & t < b \\ 0 & t \geq b \end{cases}$$



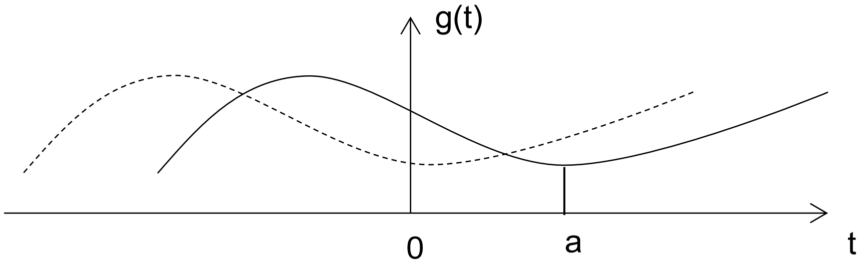
(c) Modulated pulse

$$g(t) = x(t)[H(t-a) - H(t-b)] = \begin{cases} 0 & t < a \\ x(t) & a \leq t < b \\ 0 & b \leq t \end{cases}$$



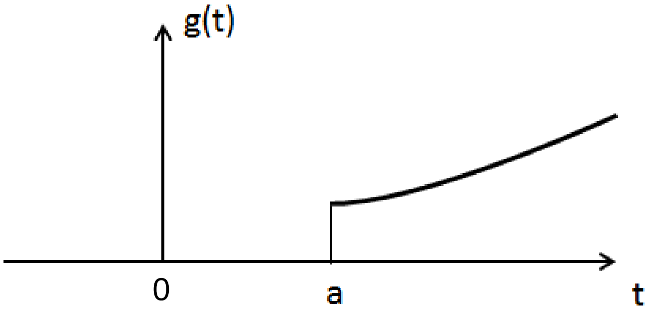
(d) Shift of v(t) along the t-axis

$g(t) = x(t-a)$



(e) Shift of v(t) along the t-axis and turn on at t = a

$g(t) = x(t-a)H(t-a)$



The new signal x(t – c) H(t – c) is a delayed and truncated version of the original signal x(t), the delay being c units. This type of signal occurs frequently in signal analysis and we need to know its Laplace transform.

Laplace transform of a delayed and truncated signal.

$$L[x(t-c)H(t-c)] = \int_0^\infty e^{-st}x(t-c)H(t-c)dt = \int_c^\infty e^{-st}x(t-c)dt$$

Put u = t – c so that dt = du :

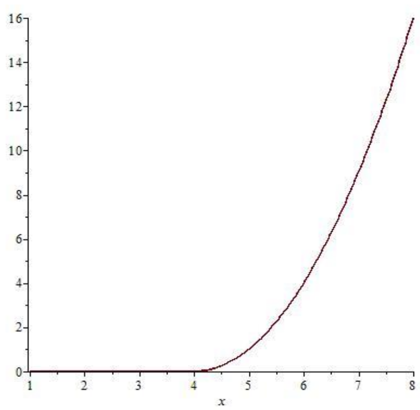
$$L[x(t-c)H(t-c)] = \int_{u=0}^\infty e^{-s(u+c)}x(u)du,$$

$$= e^{-sc}\int_{u=0}^\infty e^{-su}x(u)du = e^{-sc}X(s)$$

where X(s) is the Laplace transform of x(t).

Example

Sketch the signal H(t – 4) $(t-4)^2$ and find its Laplace transform.



$$y(t) = H(t-4)(t-4)^2$$

$$L[t^n] = \frac{n!}{s^{n+1}}$$

$$L[y(t)] = \frac{2e^{-4s}}{s^3}$$

Example

Sketch the signal H(t – 1) cos(t-1) and find its Laplace transform.

Sometimes we have to find the Laplace transform of signals of the form H(t – c) x(t), ie. where the combination (t – c) does not occur in x(t). It is fairly easy to overcome this problem, as the following example shows.

Example

Find the Laplace transform of (t^2 + 2t + 1) H(t – 3).

Need to replace all instances of t with $t-3$. Let $t = t - 3 + 3$

$$t^2 + 2t + 1 = ((t-3)+3)^2 + 2\big((t-3)+3\big) + 1$$

$$t^2 + 2t + 1 = (t-3)^2 + 6(t-3) + 9 + 2(t-3) + 6 + 1$$

$$t^2 + 2t + 1 = (t-3)^2 + 8(t-3) + 16$$

$$y(t) = H(t-3)[(t-3)^2 + 8(t-3) + 16]$$

$$y(t) = (t-3)^2 H(t-3) + 8(t-3)H(t-3) + 16H(t-3)$$

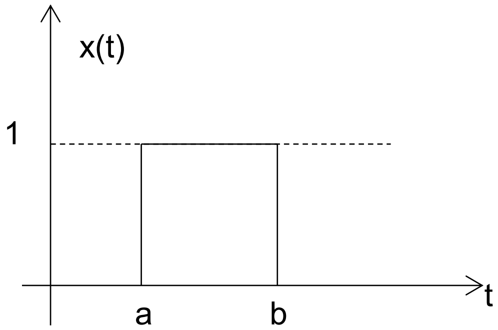
$$Y(s) = e^{-3s}\left(\frac{2}{s^3} + 8\frac{1}{s^2} + 16\frac{1}{s}\right)$$

Example

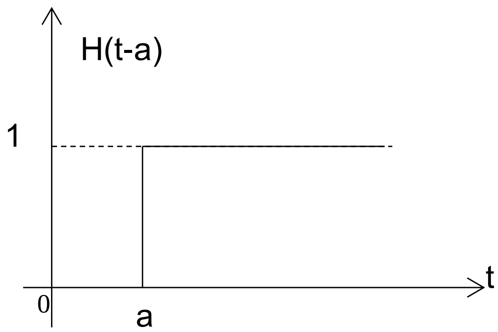
Find the Laplace transform of H(t+2)($3t^2$ + 7).

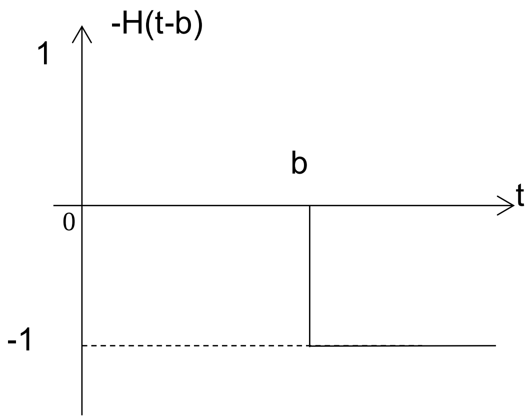
Piecewise Signals

Consider the 'top-hat' signal defined by $v(t) = \begin{cases} 0, & t < a \\ 1, & a \le t \le b \\ 0, & t > b \end{cases}$



We can express v(t) in terms of the Heaviside functions H(t – a) and H(t – b) because H(t – a) 'switches on' at t = a and – H(t – b) 'switches off' at t = b. Therefore v(t) = H(t – a) – H(t – b).





The combination [H(t – a) – H(t – b)] multiplied by a given signal selects the part of the signal corresponding to a ≤ t ≤ b and suppresses the rest. We can use this device to write down mathematical formulae for quite complicated signals <u>and</u> go on to find their Laplace transforms.

<u>Example</u>

Find the Laplace transform of the signal defined by $x(t) = \begin{cases} t^2, & 0 \le t < 3 \\ t+4, & 3 \le t < 5 \\ 9, & t \ge 5 \end{cases}$

Identify where each function is being turned on or off and multiply by the Heaviside function:

$$x(t) = t^2H(t) - t^2H(t-3) + (t+4)H(t-3) - (t+4)H(t-5) + 9H(t-5)$$

Using the Heaviside function as a common factor, factorise the expression:

$$x(t) = t^2H(t) - H(t-3)[t^2 - (t+4)\,] - H(t-5)[t+4-9]$$

And simplify

$$x(t) = t^2H(t) - (t^2 - t - 4)H(t-3) - (t-5)H(t-5)$$

Each occurrence of t needs to be replaced so that it matches the argument of the Heaviside function:

We write $t = (t-3) + 3$ and $t^2 = ((t-3)+3)^2 = (t-3)^2 + 6(t-3) + 9$

Substitute these expressions in to x(t):

$$x(t) = t^2H(t) - ((t-3)^2 + 6(t-3) + 9 - (t-3) - 3 - 4)H(t-3) - (t-5)H(t-5)$$

And simplify:

$$x(t) = t^2H(t) - ((t-3)^2 + 5(t-3) + 2\,)H(t-3) - (t-5)H(t-5)$$

Now the Laplace transform can be taken:

$$L[x(t)] = L[t^2H(t)] - L[(t-3)^2H(t-3)] - 5L[(t-3)H(t-3)] - 2L[H(t-3)] - L[(t-5)H(t-5)]$$

$$L[x(t)] = \frac{2}{s^3} - \frac{2e^{-3s}}{s^3} - \frac{5se^{-3s}}{s^3} - \frac{2s^2e^{-3s}}{s^3} - \frac{e^{-5s}}{s^2} = \frac{2 - 2e^{-3s} - 5se^{-3s} - 2s^2e^{-3s}}{s^3} - \frac{e^{-5s}}{s^2}$$

$$L[x(t)] = \frac{2 - e^{-3s}(2 + 5s + 2s^2)}{s^3} - \frac{e^{-5s}}{s^2} = \frac{2 - e^{-3s}(s+2)(2s+1)}{s^3} - \frac{e^{-5s}}{s^2}$$

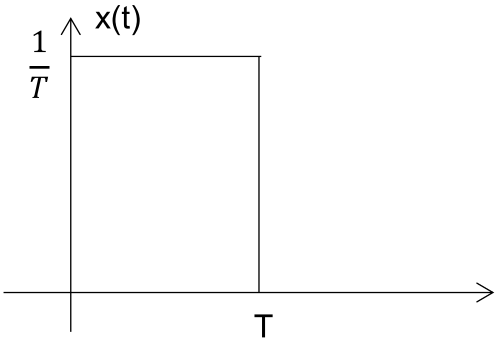
Example

Find the Laplace transform of the signal defined by $x(t) = \begin{cases} t, & 0 \le t < 1 \\ 3, & 1 \le t < 3 \\ 1 - 3t, & t \ge 3 \end{cases}$.

Note that whenever there are discontinuities (jumps) in the signal or its slope then exponential terms appear in the Laplace transform. It therefore seems reasonable to assume that if a Laplace transform contains exponentials then we should expect the corresponding time signal obtained by inverting the transform to include discontinuities. The following result states this idea more precisely.

Impulse Functions

Suppose we have a signal x(t) defined by $x(t) = \begin{cases} \frac{1}{T}, & 0 \le t \le T \\ 0, & t > T \end{cases}$



We can write

$$x(t) = \frac{1}{T}(H(t) - H(t-T)),$$

and

$$L[x(t)] = \frac{1}{T}\left(\frac{1}{s} - \frac{1}{s}e^{-sT}\right),$$

$$= \frac{1}{sT}(1 - e^{-sT}).$$

Now make T very small. The rectangle representing x(t) has a very small width but is very long. Also, since (for small T) we can use the Taylor expansion approximation about zero:

$$e^{-sT} \cong 1 - \frac{sT}{1!} + \frac{(sT)^2}{2!} - \frac{(sT)^3}{3!} + \cdots,$$

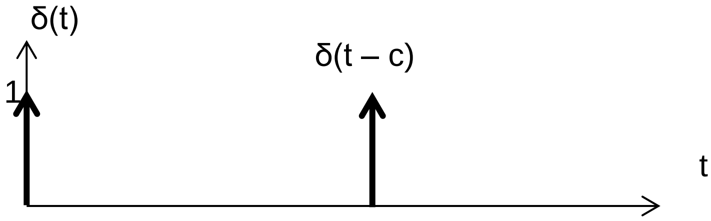
we have

$$L[x(t)] = \frac{1}{sT}\left(1 - 1 + sT - \frac{(sT)^2}{2} + - \frac{(sT)^3}{3!} \ldots\right),$$

$$= \frac{1}{sT}\left(sT - \frac{(sT)^2}{2} + \frac{(sT)^3}{3!} + \cdots\right),$$

$$= 1 - \frac{sT}{2} + \frac{(sT)^2}{3!} + \cdots.$$

As T $\rightarrow$ 0 the rectangle becomes a spike at t = 0 of infinite height. This spike is called a <u>delta function</u> and is denoted by δ(t). Its graph is an arrow of height 1 at the origin, as shown below.



The LT of the δ function is found by letting T → 0 in the above result:

$$L[\partial(t)] = \lim_{T\to 0} L[x(t)],$$

$$= \lim_{T\to 0}\left(1 - \frac{sT}{2} + \cdots\right)$$

$$= 1.$$

If the δ function is at t = c (we denote this by δ(t – c)) then, using the first shift theorem, we have

$$L[\partial(t-c)H(t-c)] = 1 \times e^{-sc} = e^{-sc}.$$

The Second Shift Theorem

If $L[v(t)] = V(s)$ then $L[v(t-c)H(t-c)] = e^{-sc}V(s)$.

It follows that

$$L^{-1}[e^{-sc}V(s)] = H(t-c)v(t-c), \text{ where } v(t) = L^{-1}[V(s)].$$

<u>Example</u>

Find $L^{-1}\left[\frac{3e^{-2s}}{s^2+9}\right]$.

Exponential function suggests that we are dealing with delayed piece-wise signal.
The delay is 2
Represent $X(s)$ as $X(s) = e^{-2s}\frac{3}{s^2+3^2}$

$$x(t) = H(t-2)\sin(3[t-2])$$

<u>Example</u>

Find $L^{-1}\left[\frac{7e^{9s}}{s^4}\right]$.

Key Points

- H(t - c) is zero for t < c, one for t $\geq$ c. It acts as a switch: off for t < c, on for t $\geq$ c. H(t) is called the Heaviside function (or unit step function).

- $L[H(t-c)] = \frac{e^{-sc}}{s}$. A special case is when c = 0, then $L[H(t)] = \frac{1}{s}$.

- H(t - c) v(t - c) is the signal v(t) moved c units right and then set to zero for t < c.

- $L[H(t-c)v(t-c)] = e^{-sc}V(s)$, and $L^{-1}[e^{-sc}V(s)] = H(t-c)v(t-c)$, where $V(s) = L[v(t)]$. This is the Second Shift theorem.

- Piecewise functions can be expressed in terms of Heaviside functions.

- The delta function, δ(t), is a spike at t = 0. δ(t - c) is a spike at t = c. The "area" under these δ functions is equal to 1.

- $L[\partial(t-c)] = e^{-cs}$. A special case is when c = 0 and then $L[\partial(t)] = 1$.

- To find the Laplace transform of a piecewise function first express it in terms of Heaviside functions and then use the second shift theorem.

Transfer Functions

Suppose we have a linear, time-invariant (LTI) system with input x(t) and output y(t) with y(t) related to x(t) by an ordinary differential equation of the form

$$a_n \frac{d^{(n)}x}{dt^{(n)}} + a_{n-1}\frac{d^{(n-1)}x}{dt^{(n-1)}} + \cdots + a_1\frac{dx}{dt} + a_0 x = b_m\frac{d^{(m)}y}{dt^{(m)}} + b_{m-1}\frac{d^{(m-1)}y}{dt^{(m-1)}} + \cdots + b_1\frac{dy}{dt} + b_0 y$$

If we take Laplace transforms of both sides of the equation and assume zero initial conditions then we obtain

$$a_n s^n X(s) + a_{n-1}s^{n-1}X(s) + \cdots + a_0 X(s) = b_m s^m Y(s) + b_{m-1}s^{m-1}Y(s) + \cdots + b_0 Y(s)$$

ie. $\quad Y(s) = G(s)X(s)$

where

$$G(s) = \frac{a_n s^n + a_{n-1}s^{n-1} + \cdots + a_1 s + a_0}{b_m s^m + b_{m-1}s^{m-1} + \cdots + b_1 s + b_0}$$

INPUT	$\Rightarrow$	SYSTEM	$\Rightarrow$	OUTPUT
x(t)		ODE		y(t)
X(s)		Transfer Function G(s)		Y(s)

The function G(s) is called the System Transfer Function. It relates the Laplace transform of the output to the Laplace transform of the input.

<u>Example</u>

A second order system is described by the equation $9\frac{d^2y}{dt^2} + 12\frac{dy}{dt} + 13y = 2\frac{dx}{dt} + 3x$.
Find the system transfer function.

$$9[s^2Y(s) - sy(0) - y'(0)] + 12[sY(s) - y(0)] + 13Y(s) = 2[sX(s) - x(0)] + 3X(s)$$

$$9s^2Y(s) + 12sY(s) + 13Y(s) = 2sX(s) + 3X(s)$$

$$Y(s)[9s^2 + 12s + 13] = X(s)[2s + 3]$$

$$G(s) = \frac{Y(s)}{X(s)} = \frac{2s+3}{9s^2 + 12s + 13}$$

<u>Example</u>

A second order system is described by the equation $2\frac{d^2y}{dt^2} - \frac{dy}{dt} = 5\frac{dx}{dt} + 17x$.
Find the system transfer function.

The system transfer function provides a "black-box" (ie. input-output) description of the system in terms of the Laplace variable s. It is often easier to work in the s-domain than in the time domain when analysing the response of a system. The system stability can be investigated in the s domain.
A transfer function is a property of the system, not of the input or output. It provides no information about how the components of the system evolve with time. It only relates input and output, hence it is a "black-box" description of the system. This means that two completely different systems, for example an LRC circuit and a mass-spring-damper system, can have the same transfer function.
Transfer functions are used for LTI systems where the initial conditions are zero.

Impulse Response

If we have a linear, time-invariant (LTI) system with a transfer function G(s) then we can relate the output to the input through the following equation:

$$Y(s) = G(s)X(s).$$

If we now make the input an impulse (ie. x(t) = δ(t)) then X(s) = 1. The output in this particular case is often denoted by h(t) (don't confuse this with the Heaviside function). We then have

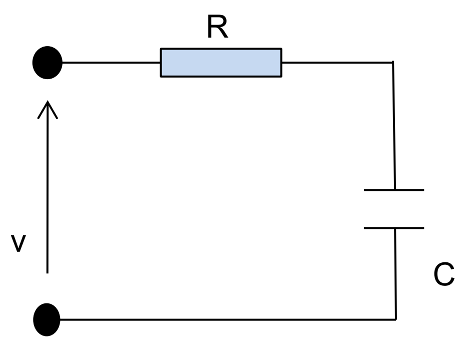
$$H(s) = G(s).1 = G(s),$$

$$\therefore h(t) = L^{-1}[G(s)].$$

Mathematically, if we know h(t) then we know G(s) and since G(s) enables the system output to be found for any input x(t), it follows that we can completely determine a system's behaviour if we know its response to an impulse. The signal h(t) is called the impulse response function.

Example

Find the transfer function of the simple low-pass CR network shown below. Find the impulse response function.



If i(t) is the current flowing then: $v(t) = iR + \frac{1}{C}q,\ i = \frac{dq}{dt}$

Taking the Laplace transform and applying zero initial conditions gives:

$$V(s) = RI(s) + \frac{Q(s)}{C}\ and\ I(s) = s(Qs)$$
$$V(s) = RI(s) + \frac{I(s)}{sC}$$

$$G(s) = \frac{I(s)}{V(s)} = \frac{1}{R + \frac{1}{sC}} = \frac{s/_R}{s + \frac{1}{RC}}$$

The impulse response is therefore:

$$h(t) = L^{-1}[G(s)] = L^{-1}\left[\frac{s}{R\left(s + \frac{1}{RC}\right)}\right]$$

$$\frac{s}{R\left(s + \frac{1}{RC}\right)} = \frac{A}{R} + \frac{B}{s + \frac{1}{RC}}$$

Linear s: $1 = A$

Constants: $0 = \frac{A}{RC} + BR, \quad BR = \frac{-A}{RC} = \frac{-1}{RC},\ B = \frac{-1}{R^2C}$

$$h(t) = L^{-1}\left[\frac{1}{R} - \frac{\frac{1}{R^2C}}{s + \frac{1}{RC}}\right] = \frac{1}{R}\delta(t) - \frac{1}{R^2C}e^{-t/_{RC}}$$

<u>Example</u>

Find the impulse response function for the system described by the transfer function $\frac{2s+3}{s^2+12s+52}$.

Poles and Zeros

Up to now, we have really assumed that the variable s in the Laplace transform is just like the variable t in that it is real. However, the numerators and denominators of transfer functions are polynomials which may have complex roots. To allow for this possibility we have to allow s to be complex. There are also deeper mathematical reasons why s has to be complex. We write

$$s = \sigma + 2j\pi f$$

where σ and $2\pi f$ are real. This, in turn, leads to the conclusion that the Laplace transform itself must also be allowed to be complex.
A graphical description of $V(s)$ is therefore difficult since four dimensions are needed. We would have to plot $|V(s)|$ and $arg\ (V(s))$, for example, against σ and $2\pi f$, where $s = \sigma + 2\,j\pi f$. In the case of a transfer function $G(s)$ a simpler graphical interpretation is obtained by simply marking in the complex s plane the points at which $G(s)$ is zero (the "zeros") and at which $G(s)$ is infinite (the "poles").
More generally, if

$$G(s) = \frac{P(s)}{Q(s)},$$

where $P(s)$ and $Q(s)$ are polynomials in s then the zeros of $G(s)$ are the roots of $P(s) = 0$ and the poles of $G(s)$ are the roots of $Q(s) = 0$.

Example

Find the poles and zeros of $H(s) = \frac{s-5}{s^2+s+1}$.

Zeros are found from: $s - 5 = 0, \quad s = 5$

Poles are found from: $s^2 + s + 1 = 0$

Using the quadratic formula: $s = \frac{-1\pm\sqrt{1^2-4\times1\times1}}{2\times1} = \frac{-1\pm\sqrt{1-4}}{2} = \frac{-1\pm\sqrt{-3}}{2} = \frac{-1}{2}(1 \pm j\sqrt{3})$

Example

Find the poles and zeros of $H(s) = \frac{s^2-4}{s(s^2-6s+2)}$

System Stability

A system is stable if it remains at rest when there is no external input and it returns to rest when any external source is removed. Mathematically, this means that the system output will decay to zero as time increases once any external sources are turned off. The stability of a system is not dependent on the input or output and is a property of the system itself. Stability can therefore be determined from the transfer function. Since a simple pole of the form $\frac{1}{s-a}$ corresponds to a time response e^{+at} it is seen that if a < 0 all such terms will decay to zero as t $\rightarrow \infty$. Thus a simple criterion for system stability is that all poles should be negative. Since a pole may be a complex number, this condition generalises to: **a system is stable only if all the poles have negative real parts, ie. if all the poles lie in the left half of the complex s plane. This condition is both necessary and sufficient.**

<u>Example</u>

Investigate the stability of the low-pass CR network.

The transfer function for the low-pass CR network was $G(s) = \frac{s/_R}{s+\frac{1}{RC}}$

The poles are found by: $s + \frac{1}{RC} = 0,\ \ s = \frac{-1}{RC}$

This system is stable if **both** R and C are greater than zero or less than zero.

<u>Example</u>

Investigate the stability of the system whose transfer function is $G(s) = \frac{2s+3}{9s^2+12s+13}$

Key Points

- A linear, time invariant system is described by a linear, ordinary differential equation (ODE).

- The Laplace transform of the ODE, assuming zero initial conditions, gives rise to the ratio of the Laplace transform of the output to the Laplace transform of the input, X(s)/Y(s). This ratio is called the transfer function of the system and is denoted by G(s).

- The transfer function gives no information about the internal state of the system – it is a “black-box” description that relates input to output.

- The output from the system when the input is a delta function is called the impulse response. It is often denoted by h(t). Note that the impulse response is the inverse of the Laplace transform of the transfer function.

- An LTI system is stable if, and only if, the poles of the system lie in the left half of the complex s plane.

Convolution

Definition

The convolution of two continuous signals $g_1(t)$ and $g_2(t)$ is another signal, f(t) say, defined by

$$f(t) = \int_{-\infty}^{\infty} g_1(t-\tau)\, g_2(\tau)\, d\tau .$$

The symbol "*" is used to denote convolution and so we can write

$$f(t) = g_1(t) * g_2(t).$$

It can be shown (by a change of variable) that

$$f(t) = \int_{-\infty}^{\infty} g_1(\tau)\, g_2(t-\tau)\, d\tau,$$

$$= g_2(t) * g_1(t).$$

It is often the case that we are dealing with causal signals for which the argument t of $g_1(t)$and $g_2(t)$ must be non-negative. We assume that $g_1(\tau)$ is zero for $\tau < 0$ and $g_2(t - \tau)$ is zero for $(t - \tau) < 0$ i.e. $\tau > t$. In this case the definition of convolution reduces to

$$f(t) = \int_0^t g_1(t-\tau)\, g_2(\tau)\, d\tau = \int_0^t g_1(\tau)\, g_2(t-\tau)\, d\tau.$$

Important: The variable of integration in the above expressions is τ. The time variable t is treated as a constant.

Convolution and the Laplace transform

One of the main reasons why convolution is so important in the analysis of LTI systems is that it can provide the system response to any input provided that the impulse response is known. This very important result relies on the following general result, known as the **Convolution Theorem**.

Transform of convolution = product of transforms.

In Laplace transform notation, for example, it can be shown that

$$L[g_1(t) * g_2(t)] = G_1(s)G_2(s),$$

where

$$G_1(s) = L[g_1(t)] \text{ and } G_2 = L[g_2(t)].$$

Example

Use the convolution theorem to find inverse Laplace transform of $\frac{1}{(s+2)(s+3)}$.

$$Let\ X(s) = \frac{1}{s+2}\ and\ H(s) = \frac{1}{s+3}$$

$$Then\ x(t) = L^{-1}[X(s)] = e^{-2t}\ and\ h(t) = L^{-1}[H(s)] = e^{-3t}$$

$$x(t) * h(t) = \int_{\tau=0}^{t} e^{-3(t-\tau)} e^{-2\tau} d\tau = e^{-3t} \int_{\tau=0}^{t} e^{\tau} d\tau = e^{-3t} [e^{\tau}]_0^t = e^{-3t}[e^t - 1] = e^{-2t} - e^{-3t}$$

Example

Use the convolution theorem to find inverse Laplace transform of $\frac{3}{s(s^2+4)}$.

Key Points

- In the case of continuous signals, the convolution is an integral. For discrete signals the convolution is a summation.
- The transform (Laplace, Fourier or z) of the convolution of two signals is the product of their individual transforms.
- The output from a linear, time invariant system is the convolution of the impulse response and the input.

FOURIER TRANSFORMS

Introduction

The Fourier transform is an example of an integral transform that takes a signal $x(t)$ and transforms it to a new function, $X(f)$, in the frequency domain. We denote the Fourier transform by the symbol "*F*" and so

$$X(f) \ = \ F[x(t)],$$

i.e. $X(f)$ is the Fourier transform of $x(t)$. The formal definition is as follows:

$$X(f) = \int_{-\infty}^{\infty} e^{-2j\pi ft} x(t) dt$$

This definition looks very similar to the Laplace transform but there are important differences. The signal $x(t)$ must be defined for all times t, not just for $t \geq 0$, and the range of integration is $-\infty < t < \infty$.

In the Laplace transform we have the term e^{-st} where $s = \sigma + 2j\pi f$. So

$$e^{-st} = e^{-(\sigma+2j\pi f)t} = e^{-\sigma t - 2j\pi ft} = e^{-\sigma t} e^{-2j\pi ft}$$

If we set $\sigma = 0$, then (apart from the range of integration) the Laplace and Fourier transform look the same. Putting $\sigma = 0$ resrticts the type of signal $x(t)$ that we can look at. Not all signals have a Fourier transform, for example, the simple dc signal $x(t) = 1$ does not have a Fourier transform, as shown below.

$$F[1] = \int_{-\infty}^{\infty} e^{-2j\pi ft} x(t) dt = \int_{-\infty}^{\infty} e^{-2j\pi ft} \times 1 dt$$

$$= \left[-\frac{e^{-2j\pi ft}}{2j\pi f} \right]_{-\infty}^{\infty} = \lim_{t \to \infty} \left(-\frac{e^{-2j\pi ft}}{2j\pi f} \right) - \lim_{t \to -\infty} \left(-\frac{e^{-2j\pi ft}}{2j\pi f} \right)$$

Now $e^{-2j\pi ft} = \cos(2\pi ft) - j\sin(2\pi ft)$

and so $\lim_{t\to\infty}\left(e^{-2j\pi ft}\right)$ does not exist.

Neither does the other limit as $t \to -\infty$. So, the Fourier transform doesn't exist. (Note that we shall show later how to give an interpretation of the Fourier transform of some signal which, strictly, do not have Fourier transforms.)

The Dirichlet Conditions

If the signal x(t) satisfies the following conditions (the so-called Dirichlet conditions) then X(f) does exist.

- $\int_{-\infty}^{\infty} |x(t)| dt$ must exist – this condition is sometimes stated "x(t) must be absolutely integrable".

- x(t) must have only a finite number of maxima and minima in any finite time interval.

Example

In the case x(t) = 1 it is the first condition that is violated because $\int_{-\infty}^{\infty}|x(t)|dt = \int_{-\infty}^{\infty} 1dt = [t]_{-\infty}^{\infty}$, therefore this integral is not finite. Hence, strictly, its Fourier transform does not exist.

The total energy associated with a signal is defined to be

$$E = \int_{-\infty}^{\infty}\{x(t)\}^2 dt$$

This integral must be finite for it to be possible to find the Fourier transform of $x(t)$. However, even if the above integral is infinite (for example if $x(t) = 10\, cos(3t)$ then E is infinite) it may still be possible to find a quantity P defined by

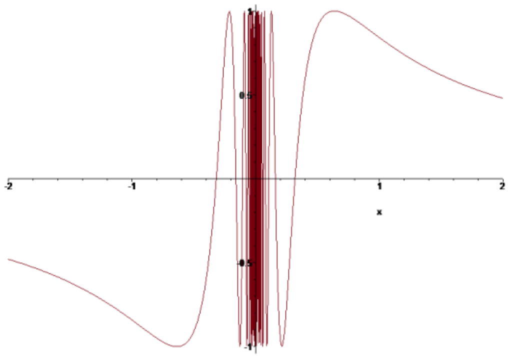
$$P = \lim_{T\to\infty}\frac{1}{T}\int_{\frac{-T}{2}}^{\frac{T}{2}}\{x(t)\}^2 dt$$

P is the average power in the signal. Signals for which E is finite are called energy signals, those for which P is finite are called power signals. Generally, Fourier transforms are applied to energy signals but there are exceptions, some of which we shall consider later.

NB. There are many functions for which it <u>is</u> possible to find the Laplace transform even when the Fourier transform does not exist.

Example

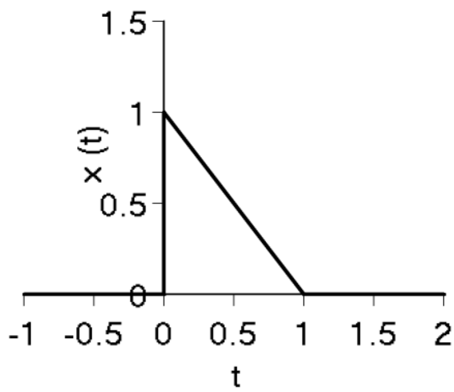
Consider the function $\sin(x^{-1})$, the graph of this function is shown to the right.



This function oscillates with increasing frequency as x→ 0 and so violates the second condition.

Example

Find the Fourier transform of the signal defined as $x(t) = \begin{cases} 0, & t < 0 \\ 1 - t, & 0 \le t \le 1 \\ 0, & t > 1 \end{cases}$



$$\boldsymbol{X(f) = \int_0^1 e^{-2j\pi ft}(1-t)dt = \int_0^1 e^{-2\pi jft} - te^{-2\pi jft}dt}$$

Integration by parts is required:

$$let\ u = t,\ \frac{du}{dt} = 1,\ \frac{dv}{dt} = e^{-2\pi jft},\ v = \frac{-e^{-2\pi jft}}{2\pi jf}$$

$$\left[\frac{-e^{-2\pi jft}}{2\pi jf}\right]_0^1 - \left[\frac{-te^{-2\pi jft}}{2\pi jf} - \int_0^1 \frac{-e^{-2\pi jft}}{2\pi jf} dt\right]_0^1 = \left[\frac{-e^{-2\pi jft}}{2\pi jf} + \frac{te^{-2\pi jft}}{2\pi jf} + \frac{e^{-2\pi jft}}{(2\pi jf)^2}\right]_0^1$$

$$\left[\frac{-e^{-2\pi jft}}{2\pi jf} + \frac{te^{-2\pi jft}}{2\pi jf} - \frac{e^{-2\pi jft}}{(2\pi f)^2}\right]_0^1$$

$$\left(\frac{-e^{-2\pi jf}}{2\pi jf} + \frac{e^{-2\pi jf}}{2\pi jf} - \frac{e^{-2\pi jf}}{(2\pi f)^2}\right) - \left(\frac{-1}{2\pi jf} + 0 - \frac{1}{(2\pi f)^2}\right)$$

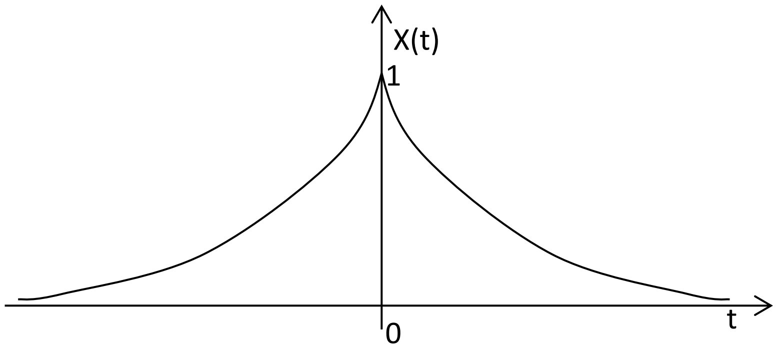
$$\frac{-e^{-2\pi jf}}{(2\pi f)^2} + \frac{1}{2\pi jf} + \frac{1}{(2\pi f)^2} = -\frac{j}{2\pi f} - \frac{1}{(2\pi f)^2}(e^{-2\pi jf} - 1)$$

<u>Example</u>

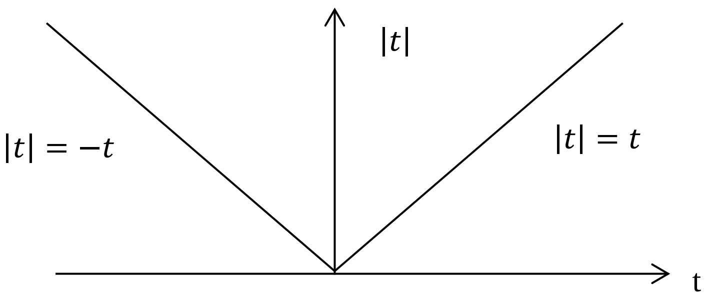
Find the Fourier transform of the signal defined as $\boldsymbol{x(t)} = \begin{cases} \mathbf{0}, & \boldsymbol{t} < \mathbf{0} \\ \mathbf{4}\boldsymbol{t}, & \mathbf{0} \le \boldsymbol{t} \le \mathbf{4} \\ \mathbf{0}, & \boldsymbol{t} > \mathbf{4} \end{cases}$

<u>Example</u>

Find the Fourier transform of the two-sided exponential x(t) = $e^{-c|t|}$, where c is a positive constant.



Note that |t| is never negative – its graph is shown below.



If t > 0, |t| = t,
if t < 0, |t| = –t.

This means that we have to evaluate the integral in two 'pieces':

$$F[x(t)] = \int_{-\infty}^{0} e^{-j2\pi ft+ct}dt + \int_{0}^{\infty} e^{-j2\pi ft-ct}dt = \int_{-\infty}^{0} e^{t(-j2\pi f+c)}dt + \int_{0}^{\infty} e^{-t(j2\pi f+c)}dt$$

$$= \left[\frac{e^{t(-j2\pi f+c)}}{(c-2\pi jf)}\right]_{-\infty}^{0} + \left[\frac{-e^{-t(j2\pi f+c)}}{(c+2\pi jf)}\right]_{0}^{\infty}$$

$$= \left(\frac{e^{0}}{(c-2\pi jf)} - \frac{e^{-\infty(-j2\pi f+c)}}{(c-2\pi jf)}\right) + \left(\frac{-e^{\infty(j2\pi f+c)}}{(c+2\pi jf)} - \frac{-e^{0}}{(c+2\pi jf)}\right)$$

$$= \left(\frac{1}{(c-2\pi jf)} - 0\right) + \left(0 - \frac{-1}{(c+2\pi jf)}\right) = \frac{1}{(c-2\pi jf)} - \frac{-1}{(c+2\pi jf)} = \frac{1}{(c-2\pi jf)} + \frac{1}{(c+2\pi jf)}$$

$$= \frac{1}{(c-2\pi jf)} \times \frac{(c+2\pi jf)}{(c+2\pi jf)} + \frac{1}{(c+2\pi jf)} \times \frac{(c-2\pi jf)}{(c-2\pi jf)} = \frac{(c+2\pi jf)}{c^2+4(\pi f)^2} + \frac{(c-2\pi jf)}{c^2+4(\pi f)^2}$$

$$= \frac{c+2j\pi f+c-2j\pi f}{c^2+4(\pi f)^2} = \frac{2c}{c^2+4(\pi f)^2}$$

Fourier Transforms of odd and even function

A function is named an even function if $f(t) = f(-t)$, examples of even functions are: cos(t), t^2, $e^t + e^{-t}$.

A function is named an odd function if $f(t) = -f(-t)$, examples of odd functions are: sin(t), t^3, $e^t - e^{-t}$.

Note that any function can be represented as a sum of odd and even functions.
The integral of an odd function in symmetrical limits is zero.

$$\int_{-a}^{a} f(t)dt = \int_{-a}^{0} f(t)dt + \int_{0}^{a} f(t)dt = -\int_{a}^{0} f(t)d(-t) + \int_{0}^{a} f(t)dt = 0$$

Using the Euler formula, we find that the Fourier Transform of an even function is real and is represented as:

$$X(f) = 2\int_0^\infty \cos 2\pi f t \;\; x(t)dt$$

The Fourier transform of an odd function is imaginary and is represented as:

$$X(f) = -2j\int_0^\infty \sin\; 2\pi f t \;\; x(t)dt$$

Inverse Fourier Transform

Like all functions, the Fourier transform has an inverse. The notation for this is $x(t) = F^{-1}[F(f)]$.

The formal definition is: $x(t) = \int_{-\infty}^{\infty} e^{2j\pi f t} X(f)df$.

<u>Example</u>

Find the inverse Fourier transform of $X(f) = \frac{-j}{\pi f}$.

$$x(t) = \int_{-\infty}^{\infty} e^{2j\pi f t}\; \frac{-j}{\pi f} df \; = \int_{-\infty}^{\infty} (\cos(2\pi f t) + j\sin(2\pi f t))\; \frac{-j}{\pi f} df$$

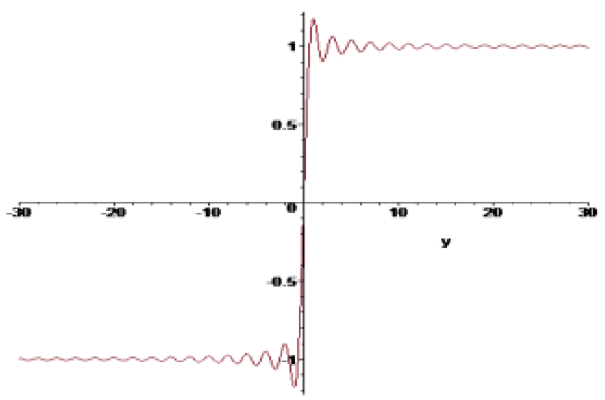
The first integral is zero. Therefore

$$x(t) = \int_{-\infty}^{\infty} \frac{\sin(2\pi f t)}{\pi f} df = 2\int_0^\infty \frac{\sin(2\pi f t)}{\pi f} df$$

We will use the property that $\int_0^\infty \frac{\sin \pi f}{\pi f} df = \frac{1}{2}$

And will make a change of variables: $y = 2ft,\;\; f = y/2t, \quad df = dy/2t$

$$x(t) = 2\int_0^{\infty \cdot 2t} \frac{\sin \pi y}{\pi y(\frac{1}{2t})} \frac{1}{2t} dy = 2\int_0^{\infty \cdot 2t} \frac{\sin \pi y}{\pi y} dy$$



If t>0, x(t)=1

If t<0, x(t)=-1

Example

Using the integral definition, find the inverse Fourier transform for the function $X(f) = f + 2$ for $-2 \leq f \leq 2$.

Key points

- The Fourier transform converts a signal x(t), i.e. a function of time t, defined for all times, not just t > 0, into a new function X(f) of the frequency f. It is a transform from the time domain to the frequency domain.

- Energy signals are those for which the total energy in the signal is finite. Fourier transforms are normally associated with energy signals.

- Power signals are those for which the total power in the signal is finite.

- We find the Fourier transform by evaluating the integral $X(f) = \int_{-\infty}^{\infty} e^{-2j\pi ft} x(t)dt$

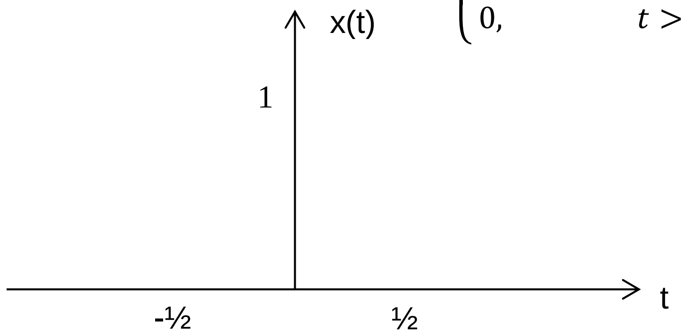
- We find the inverse transform by evaluating the integral $x(t) = \int_{-\infty}^{\infty} e^{2j\pi ft} X(f)df$

- Noting if a function is odd or even can make the calculations easier.

Engineering Functions

The Rectangular Pulse, rect(t)

The Rectangular Pulse is defined by $x(t) = \begin{cases} 0, & t < \frac{-1}{2} \\ 1, & \frac{-1}{2} \le t \le \frac{1}{2} \\ 0, & t > \frac{1}{2} \end{cases}$



The Fourier transform of rect(t).

$$\mathrm{rect}(f) = \int_{\frac{-1}{2}}^{\frac{1}{2}} e^{-2j\pi f t} dt = \left[\frac{e^{-2j\pi f t}}{-2j\pi f}\right]_{\frac{-1}{2}}^{\frac{1}{2}} = \left[\frac{j}{2\pi f} e^{-2j\pi f t}\right]_{\frac{-1}{2}}^{\frac{1}{2}} = \frac{j}{2\pi f}\left(e^{-j\pi f} - e^{j\pi f}\right)$$

Using Euler's formula: $e^{j\pi f} = \cos(\pi f) + \mathrm{j}\sin(\pi f)$

$$\mathrm{rect}(f) = \frac{j}{2\pi f}\left(\cos(\pi f) - j\sin(\pi f) - \cos(\pi f) - j\sin(\pi f)\right) = \frac{\sin(\pi f)}{\pi f}$$

The function sinc(x) (sine-cardinal).

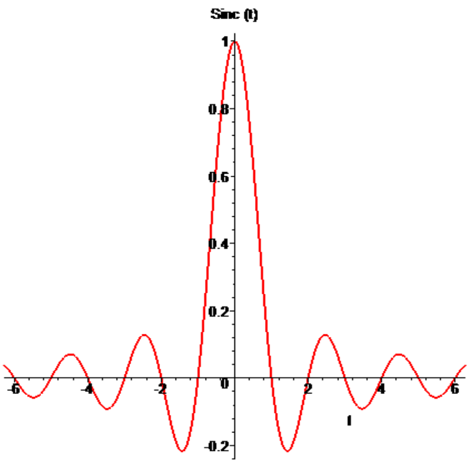
We define sinc(t) as follows: $\mathrm{sinc}(t) = \begin{cases} \frac{\sin(\pi t)}{\pi t}, & t \ne 0 \\ 1, & t = 0 \end{cases}$

sinc(t) = 0 whenever sin(πt) = 0 , (except t = 0) ,

i.e. $\pi t = n\pi$, (n = integer, n $\ne$ 0) ,

i.e. t = n , n $\ne$ 0.

So, the graph of sinc(t) is as shown below.



A more concise way of writing the Fourier transform of the rectangular pulse is therefore $F[\mathrm{rect}(t)] = \mathrm{sinc}(f)$.

The Fourier transform of sinc(t)

We just found that $F[\mathrm{rect}(t)] = \mathrm{sinc}(f)$

We are going to use the inverse Fourier transform to represent the rectangular pulse in the form

$$\mathrm{rect}(t) = F^{-1}[\mathrm{sinc}(f)] = \int_{-\infty}^{\infty} e^{2j\pi f t}\mathrm{sinc}(f)df$$

Change the argument t → -t

$$\mathrm{rect}(-t) = \int_{-\infty}^{\infty} e^{-2j\pi f t}\mathrm{sinc}(f)df$$

The integral form now looks like a direct Fourier transform.

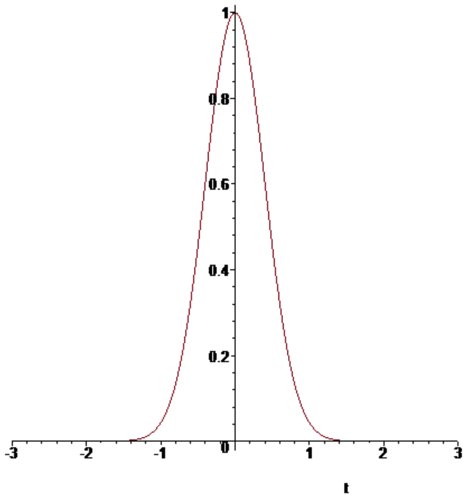
Change the variable f → t, and use the property of an even function rect(t)=rect(-t)

$$\mathrm{rect}(t) = \int_{-\infty}^{\infty} e^{-2j\pi f t}\mathrm{sinc}(f)df$$

Therefore we have:

$$F[\mathrm{sinc}(t)] = \mathrm{rect}(f)$$

The Gaussian function



We define the Gaussian function as $g(t) = e^{-\pi t^2}$.

The Fourier transform of the Gaussian function:

Use the value of the normalised Gaussian integral: $\int_{-\infty}^{\infty} e^{-\pi t^2} dt = 1$

$$G(f) = \int_{-\infty}^{\infty} e^{-2j\pi f t} e^{-\pi t^2} dt$$

Rearrange the integrand: $e^{-2j\pi f t} e^{-\pi t^2} = e^{-\pi(2jft + t^2)} = e^{-\pi((t+jf)^2 - (jf)^2)}$

$$G(f) = \int_{-\infty}^{\infty} e^{-\pi f^2} e^{-\pi (t-jf)^2} dt$$

We can also change the integration variable $\tilde{t} = (t + jf)\ and\ d\tilde{t} = dt$

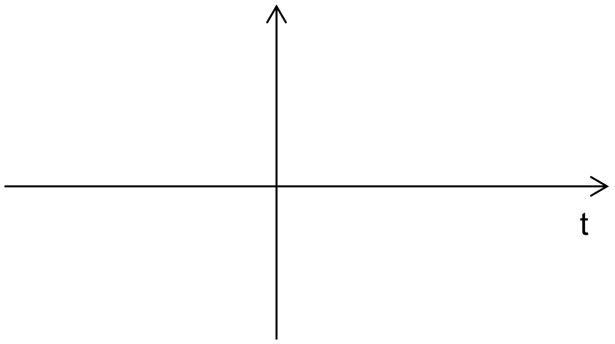
We write the transform as:

$$G(f) = e^{-\pi f^2} \int_{-\infty + jf}^{\infty + jf} e^{-\pi \tilde{t}^2} d\tilde{t}$$

As we have used the normalised Gaussian function, we know the integral takes the value 1, so therefore we have $(f) = F[g(t)] = e^{-\pi f^2}$.

The Signum function

We define this function as $\mathrm{sgn}(t) = \begin{cases} -1, & t < 0 \\ 1, & t \geq 0 \end{cases}$.



The Signum function takes the sign of t but not its magnitude. What is the Fourier transform for $\mathrm{sgn}(t)$?

We found that the inverse Fourier transform of $\frac{-j}{\pi f}$ was $2\int_0^{\infty \cdot 2t} \frac{\sin(\pi y)}{\pi y} dy$, which tended to 1 if t>0 and -1 if t<0. We use this to define the Fourier transform of the Signum function.

Using frequency image $\frac{-j}{\pi f}$ and taking its inverse Fourier transform it is found that:

$$\mathrm{sgn}(t) = F^{-1}\left[\frac{-j}{\pi f}\right]$$
$$F[\mathrm{sng}(t)] = \frac{-j}{\pi f}$$

<u>Key points</u>

- The Rectangular Pulse is defined as $\mathrm{rect}(t) = \begin{cases} 0, & t < \frac{-1}{2} \\ 1, & \frac{-1}{2} \leq t \leq \frac{1}{2} \\ 0, & t > \frac{1}{2} \end{cases}$ and has the Fourier transform $F[\mathrm{rect}(t)] = \mathrm{sinc}(f)$.
- The Sinc function is defined as $\mathrm{sinc}(t) = \begin{cases} \frac{\sin(\pi t)}{\pi t}, & t \neq 0 \\ 1, & t = 0 \end{cases}$ and has the Fourier transform $F[\mathrm{sinc}(t)] = \mathrm{rect}(f)$.
- The Signum function is defined as $\mathrm{sgn}(t) = \begin{cases} 1, & t > 0 \\ -1, & t < 0 \end{cases}$ and has the Fourier transform $F[\mathrm{sgn}(t)] = \frac{-j}{\pi f}$.
- We shall evaluate the Fourier transform of a few simple signals directly from the integral definition but, in general, we will make use of tables of Fourier transforms.

Table of Standard Fourier Transforms

Description	Function	Transform
definition	v(t)	$V(f) = \int_{-\infty}^{\infty} v(t)e^{-2j\pi ft}dt$
Scaling	$v\left({}^{t}/_{T}\right)$	$\lvert T\rvert . V(fT)$
Time shift	v(t – T)	$V(f) . e^{-j2\pi fT}$
Frequency shift	$v(t)e^{2j\pi f_0 t}$	$V(f - f_0)$
Complex conjugate	v*(t)	V*(–f)
Reciprocity	V(t)	v(–f)
Addition	A . v(t) + B . w(t)	A . V(f) + B . W(f)
Multiplication	v(t) . w(t)	V(f) * W(f)
Convolution	v(t) * w(t)	V(f) . W(f)
Delta function	$\delta(t)$	1
Constant	1	$\delta(f)$
Rectangular function	rect(t)	$\mathrm{sinc}(f) = \frac{\sin(\pi f)}{\pi f}$
Sinc function	sinc(t)	rect(f)
Unit step function	u(t) or H(t)	$\frac{1}{2}\delta(f) - \frac{j}{2\pi f}$
Signum function	sgn(t)	$-\frac{j}{\pi f}$
Decaying exponential, two-sided	$e^{-c\lvert t\rvert}$	$\frac{2c}{c^2 + (2\pi f)^2}$
Decaying exponential, one-sided	$e^{-c\lvert t\rvert} . H(t)$	$\frac{c - 2j\pi f}{c^2 + (2\pi f)^2}$
Gaussian function	$e^{-\pi t^2}$	$e^{-\pi f^2}$

Amplitude and Phase Spectra

We have seen from the above examples that the Fourier transform of a real signal $x(t)$ is a function of f, $X(f)$, which may be real or complex. In general, it can be split into its real and imaginary parts, $\mathcal{R}(f)$ and $\mathcal{J}(f)$ where

$$X(f) = \mathcal{R}(f) + j\mathcal{J}(f).$$

Alternatively, it could be represented in the polar from through an amplitude, $A(f) = |X(f)|$ and a phase, $\theta(f)$. Remember that the phase angle changes within $-\pi \le \theta \le \pi$ and the phases $\theta + 2\pi n$ are identical.

If we want to 'plot the graph' of $X(f)$ against f then we actually need <u>two</u> graphs,

(a) $|X(f)|$ against f – this is called the amplitude spectrum

(b) $\theta(f)$ against f – this is called the phase spectrum.

In the very simple case when $X(f)$ is real and positive then we only need <u>one</u> graph, $X(f)$ against f, and this is called the frequency spectrum.

<u>Example</u>

Find the amplitude spectrum and the phase spectrum for the signal $x(t) = \begin{cases} 0, & t < 0 \\ e^{-2t}, & t \ge 0 \end{cases}$.

$$X(f) = \int_0^{\infty} e^{-2t} e^{-2j\pi f t} dt = \frac{1}{(2+2j\pi f)} = \frac{1-j\pi f}{2(1+(\pi f)^2)} = \frac{1}{2(1+(\pi f)^2)} - j\,\frac{\pi f}{2(1+(\pi f)^2)}$$

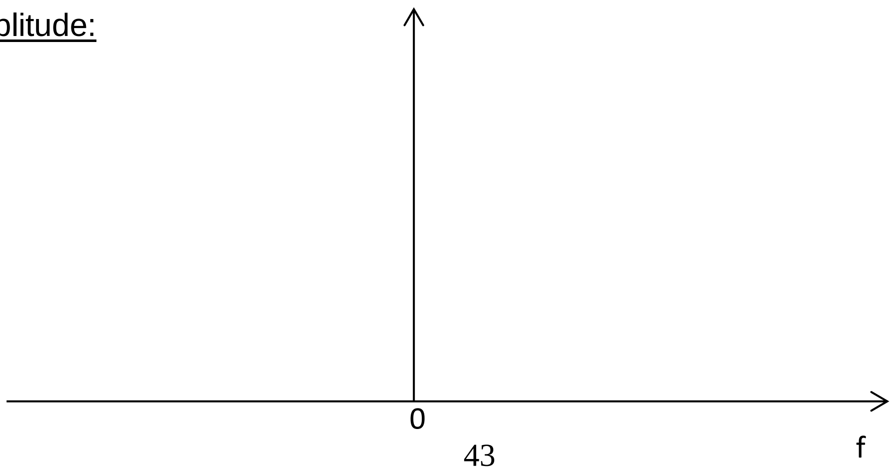
$$A(f) = \sqrt{\mathcal{R}(f)^2 + \mathcal{J}(f)^2} = \sqrt{\left(\frac{1}{2(1+(\pi f)^2)}\right)^2 + \left(\frac{\pi f}{2(1+(\pi f)^2)}\right)^2} = \sqrt{\frac{1+\pi f^2}{4(1+(\pi f)^2)^2}} = \sqrt{\frac{1}{4(1+(\pi f)^2)}}$$

$$= \frac{1}{\sqrt{4(1+(\pi f)^2)}} = \frac{1}{2\sqrt{(1+(\pi f)^2)}}$$

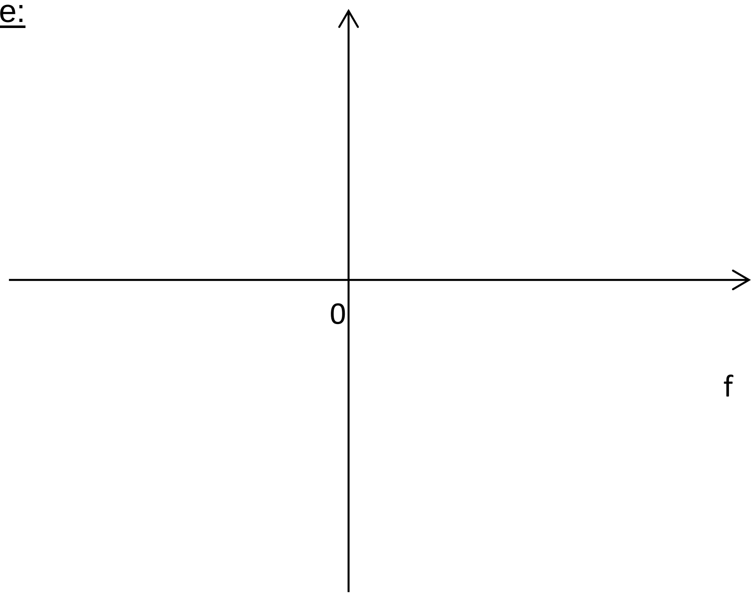
$$\tan^{-1}\left(\frac{I}{R}\right) = \tan^{-1}\left(\frac{\frac{-\pi f}{2(1+(\pi f)^2)}}{\frac{1}{2(1+(\pi f)^2)}}\right) = \tan^{-1}\left(\frac{-\pi f}{2(1+(\pi f)^2)} \times \frac{2(1+(\pi f)^2)}{1}\right)$$

$$= \tan^{-1}(-\pi f) = -\tan^{-1}(\pi f)$$

<u>Amplitude:</u>



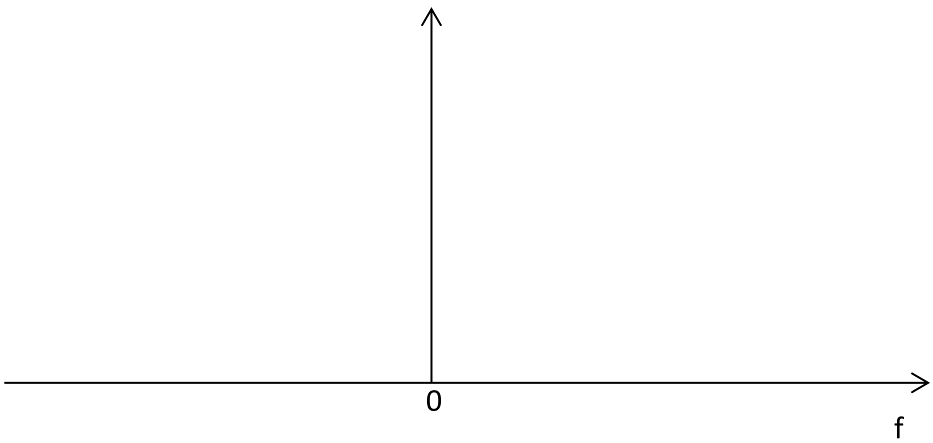
Phase:



The amplitude and phase spectra of the rectangular pulse.

The Rectangular Pulse is defined by $x(t) = \begin{cases} 0, & t < \frac{-1}{2} \\ 1, & \frac{-1}{2} \leq t \leq \frac{1}{2} \\ 0, & t > \frac{1}{2} \end{cases}$ and we know the Fourier transform is $F[\mathrm{rect}(t)] = \frac{\sin(\pi f)}{\pi f}$. This function only has real part and therefore only one graph, the frequency spectrum, will need to be plotted.
The amplitude is:

Amplitude:

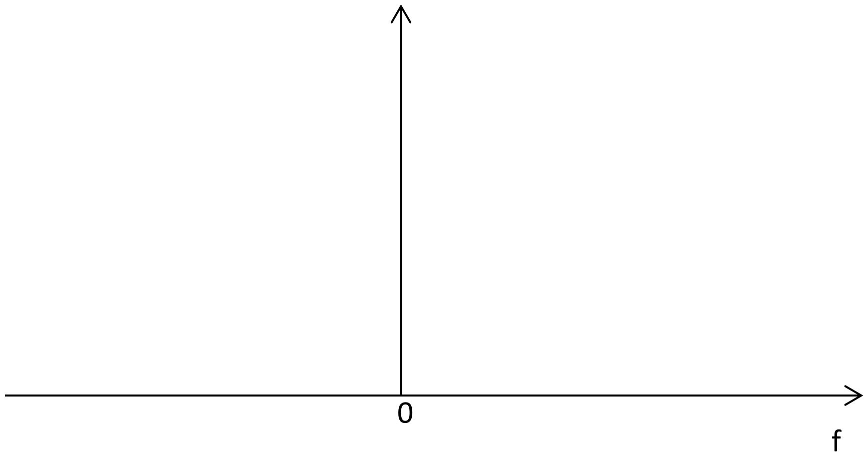


Example

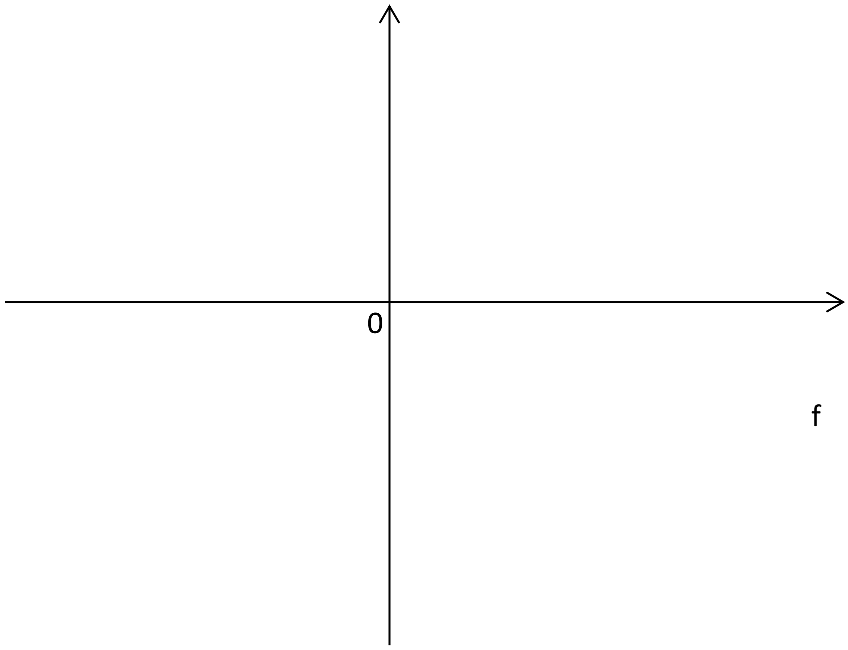
Find the amplitude and phase spectrum of the signal defined as $x(t) = \begin{cases} 0, & t < 0 \\ 1 - t, & 0 \leq t \leq 1. \\ 0, & t > 1 \end{cases}$

We found that $X(f) = \frac{-j}{2\pi f} - \frac{1}{(2\pi f)^2}\left(e^{-2j\pi f} - 1\right)$.

Amplitude:



Phase:



Key Points

- The Fourier transform $X(f)$ often turns out to be complex. It therefore has a real part, $\mathcal{R}(f)$, and an imaginary part, $\mathcal{J}(f)$. Equivalently, $X(f)$ has an amplitude, $|X(f)|$, and a phase, $\theta(f)$.

- The function $|X(f)|$ is called the amplitude spectrum and the function $\theta(f)$ is called the phase spectrum.

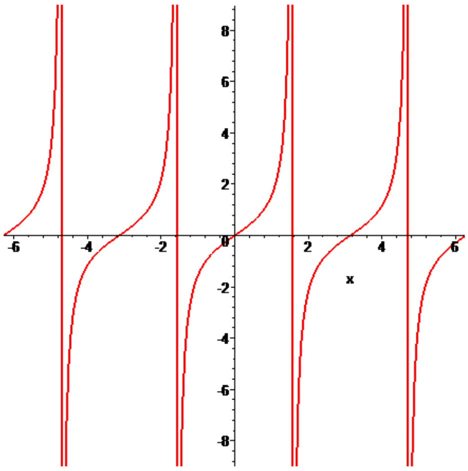
- If $X(f)$ is real, i.e. its imaginary part $\mathcal{J}(f)$ is zero, then the phase $\theta(f)$ is zero for all frequencies f and $|X(f)|$ is called the frequency spectrum.

Additional Note:

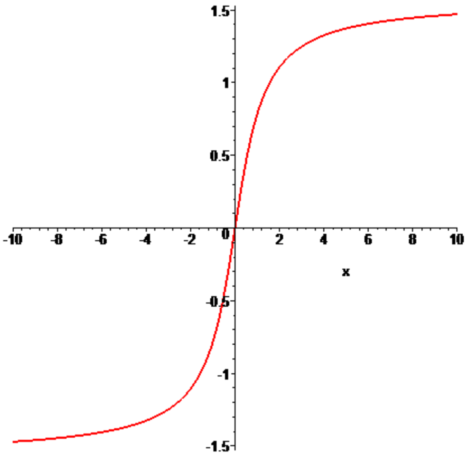
Amplitude and phase representation of a complex number.

First let us recall some properties of pair of functions $\tan(x)$ and $\tan^{-1}(x)$. Their plots are shown below:

tan(x)



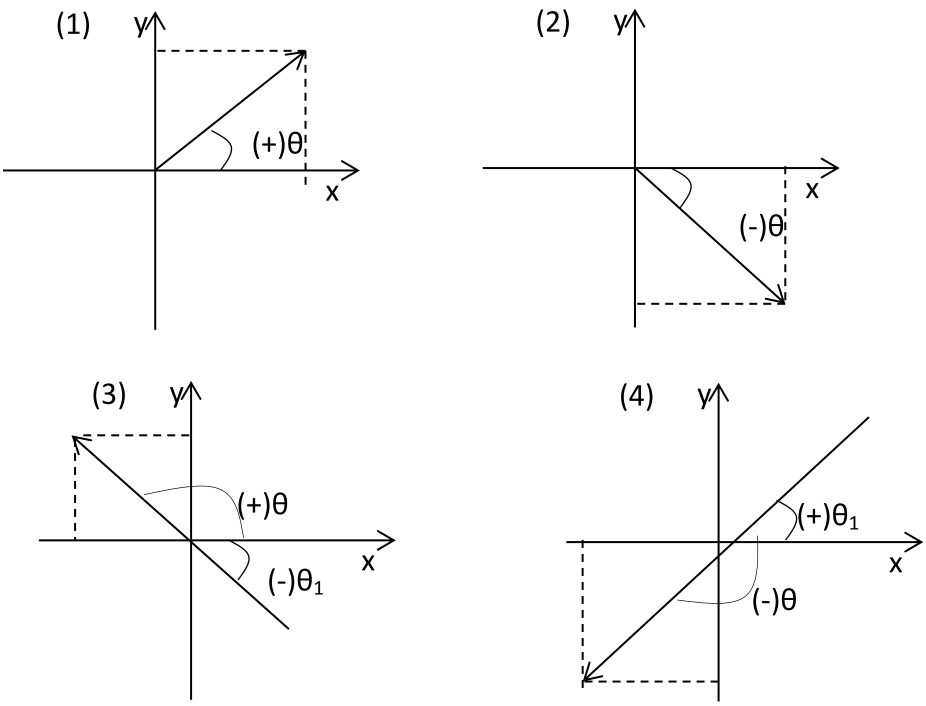
$\tan^{-1}(x)$



The function $\tan(x)$ is a periodic function with a period π since $\tan(x) = \tan(x+\pi)$, it tends to $\pm\infty$ when $x \to \pm\frac{\pi}{2} \pm n\pi$. The interval of $\frac{-\pi}{2} < x \leq \frac{\pi}{2}$ is known as the prime interval. The inverse function $\tan^{-1}(x)$ gives the value of angle in this prime interval, it tends to $\pm\frac{\pi}{2}$ when $x \to \pm\infty$.

Now we will consider a complex number, $z = x + jy$. It can be represented in a polar form: $z = Ae^{j\theta}$. Here A is known as amplitude of a complex number $A = \sqrt{x^2 + y^2}$ and θ is the phase $\theta = \tan^{-1}\left(\frac{y}{x}\right)$.

The phase is defined in the interval $-\pi < \theta \leq \pi$. The angle $\theta = 2\pi$ corresponds to $\theta = 0$. Therefore, $\theta = \pi + \alpha = \pi + \alpha - 2\pi = -\pi + \alpha$. The angles $-\pi$ and π define the same phase vector. Let us consider a graphical representation of a complex number in the complex plane.



Since the function $\tan^{-1}(x)$ is restricted to the interval $\pm\frac{\pi}{2}$, and the phase of the complex number changes within $\pm\pi$, the phase definition will depend on in which quadrant of the complex plane the phase vector is.
For cases (1) and (2), that is when $Re(z) > 0, \theta = \tan^{-1}\left(\frac{y}{x}\right)$. The phase angle is positive for $Im(z) > 0$ and negative for $Im(z) < 0$, which corresponds to the definition of $\tan^{-1}$.
For case (3), when $Re(z) < 0$ and $Im(z) > 0$, $\theta = \theta_1 + \pi = \tan^{-1}\left(\frac{y}{x} + \pi\right)$.
For case (4), when $Re(z) < 0$ and $Im(z) < 0$, $\theta = \theta_1 - \pi = \tan^{-1}\left(\frac{y}{x} - \pi\right)$.

The rule for defining the phase θ of a complex number $z = x + jy$ could be formulated as following:

$$\theta = \begin{cases} \tan^{-1}\left(\dfrac{y}{x}\right), & x = Re(z) > 0 \\ \tan^{-1}\left(\dfrac{y}{x}\right) + sgn(y)\pi, & x = Re(z) < 0 \end{cases}$$

Some Properties of Fourier Transforms

Linearity

$$F[a_1x_1(t) + a_2x_2(t)] = a_1F[x_1(t)] + a_2F[x_2(t)]$$

The Fourier transform of a sum of a number of functions is equal to the sum of the Fourier transforms of those functions.

Example

Find the Fourier transform $x_r(t)+x_l(t)$ for the following functions:

$$x_r(t) = \begin{cases} 0, & -\infty < t \le 0 \\ e^{-2t}, & 0 \le t < \infty \end{cases} \; and \; x_l(t) = \begin{cases} e^{2t}, & -\infty < t \le 0 \\ 0, & 0 \le t < \infty \end{cases}$$

$$F[x_r + x_l] = F[x_r] + F[x_l] = \int_0^\infty e^{-2t}e^{-j2\pi ft}dt + \int_{-\infty}^0 e^{2t}e^{-j2\pi ft}dt$$

$$= \int_0^\infty e^{-2t-j2\pi ft}dt + \int_{-\infty}^0 e^{2t-j2\pi ft}dt = \int_0^\infty e^{-t(2+j2\pi f)}dt + \int_{-\infty}^0 e^{t(2-j2\pi f)}dt$$

$$= \left[\frac{e^{-t(2+j2\pi f)}}{2+j2\pi f}\right]_0^\infty + \left[\frac{e^{t(2-j2\pi f)}}{2-j2\pi f}\right]_{-\infty}^0 = \frac{1-j\pi f}{2(1+(\pi f)^2)} + \frac{1+j\pi f}{2(1+(\pi f)^2)}$$

$$= \frac{1-j\pi f+1+j\pi f}{2(1+(\pi f)^2)} = \frac{1}{1+(\pi f)^2}$$

Example

Using the table of standard transforms, find the Fourier transform for the signal $x(t) = \frac{1}{5}\mathrm{rect}(t) + e^{-\pi t^2} + 4\mathrm{sinc}(t)$

Time-shift Property (Translation Property)

If $F[x(t)] = X(f)\ \ then\ \ F[x(t-a)] = e^{-2j\pi af}X(f)$

Proof:

By definition, $F[x(t)] = \int_{-\infty}^{\infty} x(t)e^{-2j\pi ft}dt$

So, $F[x(t-a)] = \int_{-\infty}^{\infty} x(t-a)e^{-2j\pi ft}dt$

Change the variables, let $u = t - a$, so $t = u + a$ and $\frac{du}{dt} = 1$ and the limits will remain infinite.

Hence, $F[x(t-a)] = \int_{-\infty}^{\infty} x(u)e^{-2j\pi f(u+a)}du$

$$= \int_{-\infty}^{\infty} x(u)e^{-2j\pi fu}e^{-2j\pi a}du = e^{-j2\pi fa}\int_{-\infty}^{\infty} x(u)e^{-2j\pi fu}du$$

The integration variable is arbitrary and so can be changed from u to t which gives us the desired result.

<u>Example</u>

Find the Fourier transform for $x(t) = \begin{cases} 0, & t < 0 \\ 1, & 0 \le t \le 1 \\ 0, & t > 1 \end{cases}$

The function can be represented via $\mathrm{rect}(t)$ function

$$x(t) = \mathrm{rect}\left(t - \frac{1}{2}\right)$$

The rule we need to apply is: $F[x(t-a)] = e^{-2j\pi af}X(f)$

The value a is: $a = \frac{1}{2}$

Apply the rule to the signal: $e^{-2j\pi f\times\frac{1}{2}}\frac{\sin(\pi f)}{\pi f}$

Remember to cancel fractions if possible.

The integral represents the Fourier transform for the normalised rectangular pulse, therefore:

$$X(f) = e^{-j\pi f}\frac{\sin(\pi f)}{\pi f}$$

<u>Example</u>

Find the Fourier transform of a rectangular pulse of total duration 1, amplitude 3 and centred at t=4.

Example

Find the Fourier transform of $x(t) = -\mathrm{sgn}(t-7)$

How does a time shift affect the amplitude and phase spectra?

Remember that $X(f)$ will, in general, be a complex function of f and so it will have a modulus, $|X(f)|$, and phase, $\theta(f)$.
Now

$$\left|e^{-2j\pi f a}X(f)\right| = \left|e^{-2j\pi f a}\right||X(f)|$$

because $\left|e^{-2\pi j f a}\right| = 1$. This means that $X(f)e^{-2\pi j f a}$ has the <u>same</u> amplitude spectrum as $X(f)$.

However

$$\begin{aligned}\arg\left(e^{-2j\pi f a}X(f)\right) &= arg\left(e^{-2j\pi f a}\right) + \arg(X(f)) \\ &= \theta(f) - 2\pi f a\end{aligned}$$

So there is a (frequency dependent) phase shift in the new spectrum.

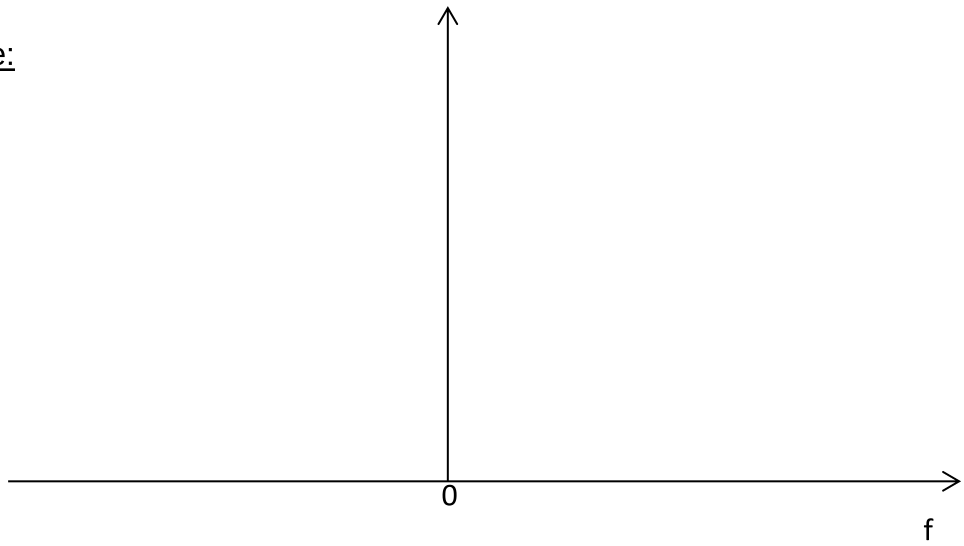
Example

What is the amplitude and phase $x(t) = \begin{cases} 0, & t < 0 \\ 1, & 0 \le t \le 1 \\ 0, & t > 1 \end{cases}$

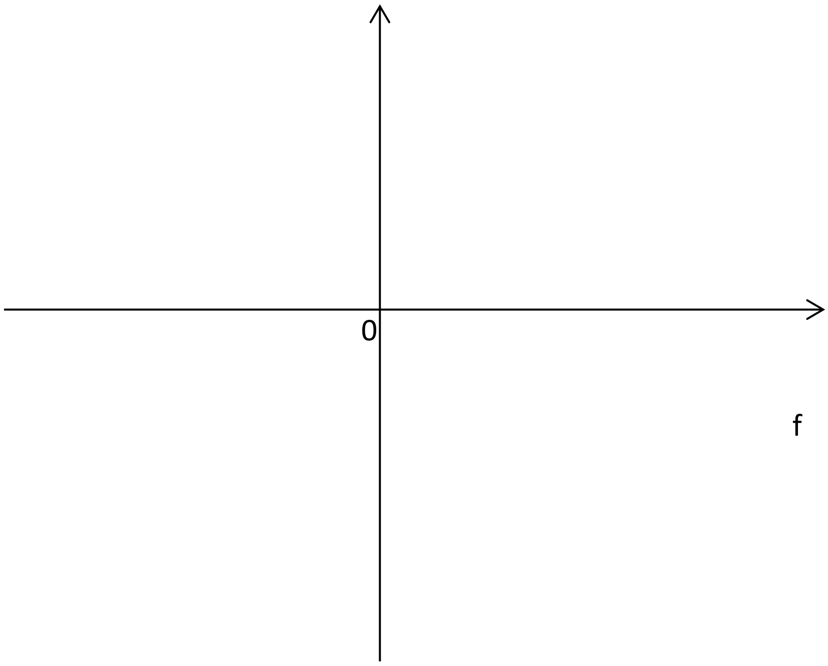
$$X(f) = e^{-j\pi f}\,\frac{\sin(\pi f)}{\pi f}$$

$$X(f) = \left|\frac{\sin(\pi f)}{\pi f}\right| \begin{cases} e^{-j\pi f}, & sinc(f) > 0 \\ e^{-j\pi f + j\pi}, & sinc(f) < 0 \end{cases}$$

<u>Amplitude:</u>



Phase:



Frequency Shift Property

If $F[x(t)] = X(f)\ then\ F\left[e^{2j\pi f_0 t}x(t)\right] = X(f - f_0)$

This result resembles very closely the time shift property 'in reverse'. Its importance is related to the concept of <u>modulation</u>.

Suppose we have a signal $x(t)$ and multiply it by a cosine wave at some carrier frequency $2\pi f_c$ The cosine wave is then $\cos(2\pi f_c t)$ and the new, modulated signal is $y(t) = x(t)\cos(2\pi f_c t)$.

What is the Fourier transform of the new signal? What is its amplitude spectrum?

We first write

$$\cos(2\pi f_c t) = \frac{1}{2}\left(e^{2j\pi f_c t} + e^{-2j\pi f_c t}\right)$$

Then

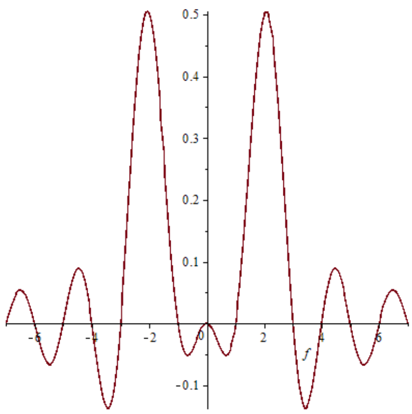
$$F[y(t)] = F\left[\frac{1}{2}\left(e^{2j\pi f_c t}x(t) + e^{-2j\pi f_c t}x(t)\right)\right]$$

$$= \frac{1}{2}F\left[\left(e^{2j\pi f_c t}x(t)\right)\right] + \frac{1}{2}F\left[\left(e^{-2j\pi f_c t}x(t)\right)\right]$$

$$= \frac{1}{2}X(f - f_c) + \frac{1}{2}X(f + f_c)$$

We see that the effect of modulation on the amplitude spectrum is to produce two replicas, of half the amplitude, centred on f = f_c and f = –f_c.

Modulated rectangular pulse



Scaling Property

If $F[x(t)] = X(f)$ then $F[x(at)] = \frac{1}{|a|} X\left(\frac{f}{a}\right)$

The modulus signs around the a allow for the possibility that a is negative. We shall only consider cases when $a > 0$.

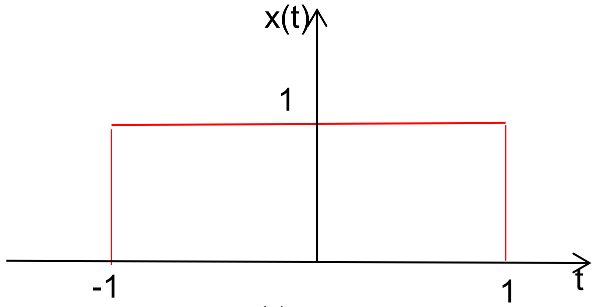
$$F[x(at)] = \int_{-\infty}^{\infty} x(at) e^{-2j\pi f t} dt$$

If we assume $a > 0$ and change the variables as follows, $u = at$ so $\frac{du}{dt} = a$

$$F[x(u)] = \frac{1}{a} \int_{-\infty}^{\infty} x(u) e^{-2j\pi \frac{f}{a} u} du = \frac{1}{a} X\left(\frac{f}{a}\right)$$

<u>Example</u>

Find the Fourier transform of the function $x(t)$ depicted below



$$x(t) = rect\ (t/2)$$

$$F\left[\text{rect}\left(\frac{t}{2}\right)\right] = 2X(2f),\ where \qquad X(f) = F[\text{rect}(t)]$$

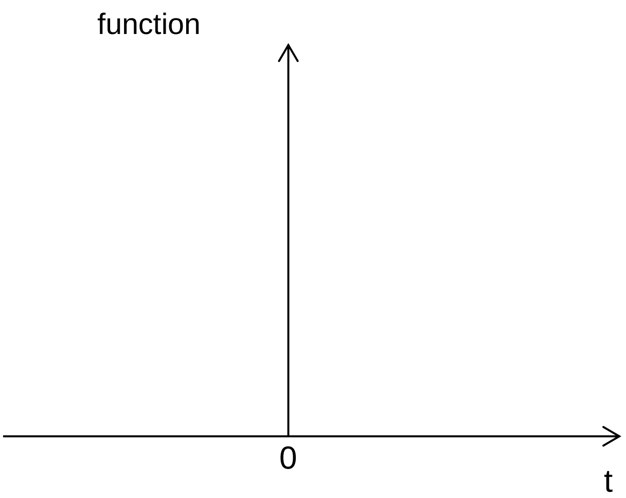
$$F\left[\text{rect}\left(\frac{t}{2}\right)\right] = 2\frac{\sin(\pi 2f)}{2\pi f} = \frac{\sin(2\pi f)}{\pi f}$$

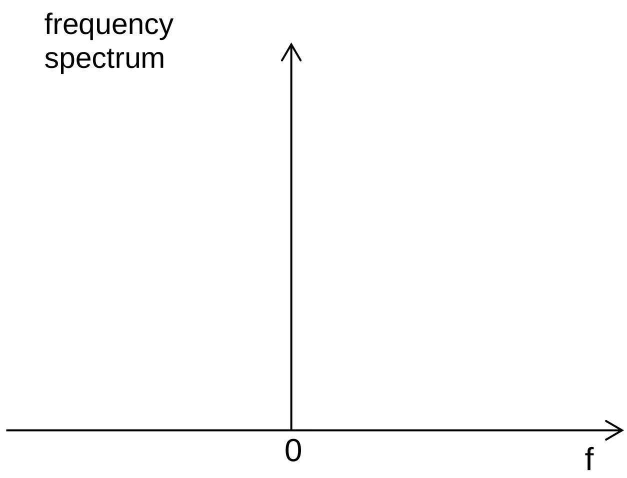
The effect of Scaling

Consider the following function and its Fourier transform:

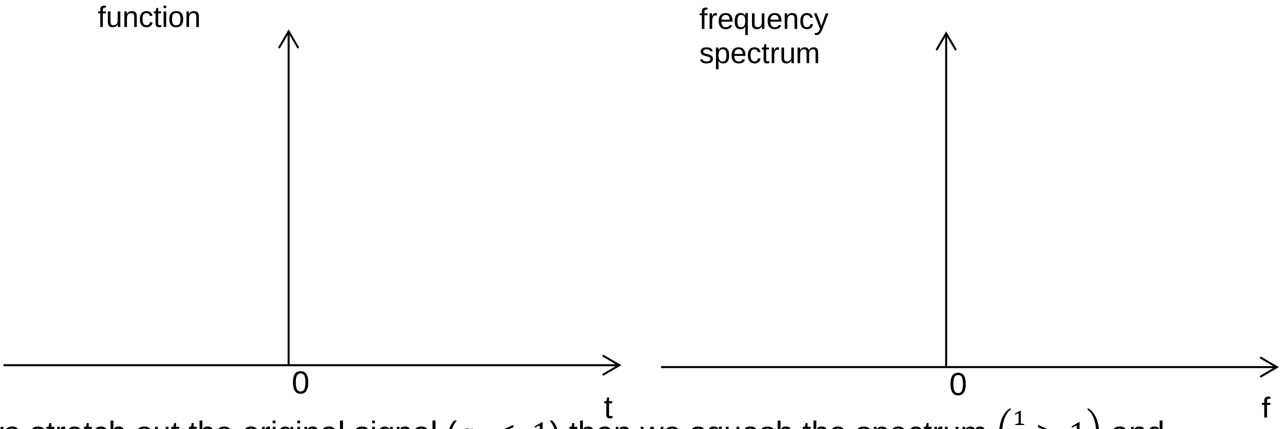
$$F\left[e^{-c|t|}\right] = \frac{2c}{c^2 + (2\pi f)^2}$$

When c=1





When c=10



If we stretch out the original signal ($a < 1$) then we squash the spectrum $\left(\frac{1}{a} > 1\right)$ and increase the amplitude. Conversely, if we squash the signal ($a > 1$) then we stretch the spectrum $\left(\frac{1}{a} < 1\right)$ and reduce the amplitude.

The Reciprocal Property

This is a very important result that allows a new Fourier transform to be written down for every Fourier transform that can be found. It is stated formally as follows.

$$If\ X(f) = F[x(t)] \text{ then } F[X(t)] = x(-f)$$

Example

Let $x(t) = \mathrm{rect}(t)$ then $X(f) = \mathrm{sinc}(f)$

Using the reciprocal property:

$F[\mathrm{sinc}(t)] = \mathrm{rect}(-f)$

$= \mathrm{rect}\ (f)$ as rect is an even function.

Example

Let $x(t) = e^{-t}$ then $X(f) = \frac{2}{1+(2\pi f)^2}$

Key Points

- The Fourier transform of a linear combination of signals is the corresponding linear combination of Fourier transforms.

- The time shift property: $if\ F[x(t)] = X(f)\ then\ F[x(t-a)] = e^{-2j\pi af} X(f)$

- Time shifting does not affect the amplitude spectrum but it does change the phase spectrum.

- The frequency shift property: $if\ F[x(t)] = X(f)\ then\ F\left[e^{2j\pi f_0 t} x(t)\right] = X(f - f_0)$

- The frequency shift result is used when dealing with modulated signals, i.e. signals which are the product of a signal $v(t)$ and a sine wave or cosine wave at some “carrier” frequency f_c.

- The scaling property: $if\ F[x(t)] = X(f)\ then\ F[x(at)] = \frac{1}{|a|} X\left(\frac{f}{a}\right)$

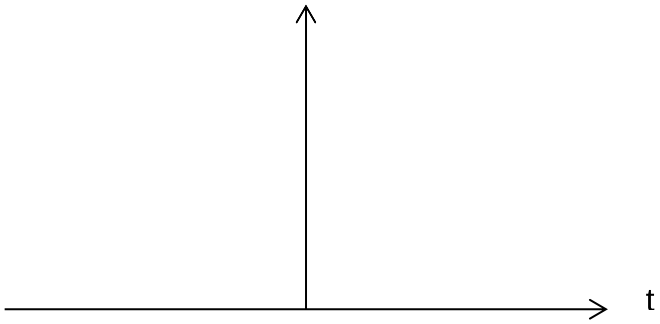
- The reciprocal property allows a new Fourier transform to be written down from a given Fourier transform as follows: if $X(f) = F[x(t)]$ then $F[X(t)] = x(-f)$.

Further Engineering Functions

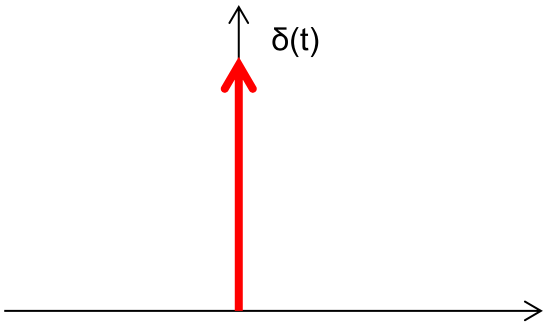
The Impulse Function (δ function)

Let's start with a rectangular pulse of amplitude $\frac{1}{T}$ and width T centred at t = 0

i.e. $\frac{1}{T}\mathrm{rect}\left(\frac{1}{T}\right)$.

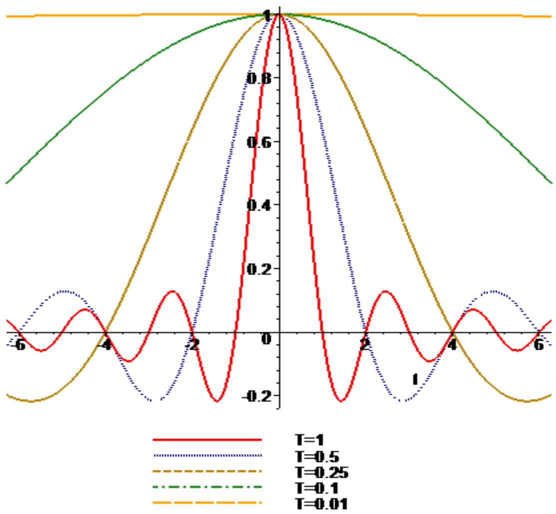


The area of the rectangle is always 1 no matter what value T takes. Now let $T \to 0$. The signal represented by $\frac{1}{T}\mathrm{rect}\left(\frac{1}{T}\right)$ becomes more and more concentrated around t = 0 and its amplitude grows indefinitely. In the limit we have an infinite 'spike' at t = 0. This 'function' is called the Impulse Function, or delta function, and is denoted by the symbol $\delta(t)$. We illustrate its graph by an arrow:



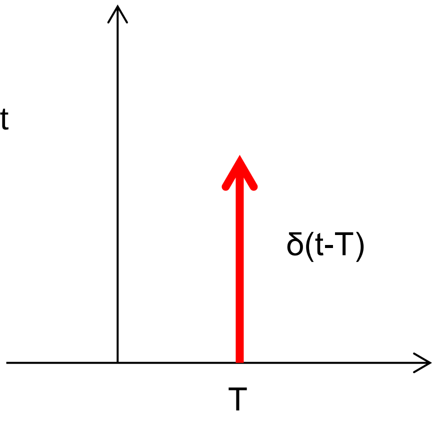
To find the FT of this impulse we start with the FT of $\frac{1}{T}\mathrm{rect}\left[\frac{t}{T}\right]$ and see what happens as $T \to 0$.

The FT of $\frac{1}{T}\mathrm{rect}\left(\frac{t}{T}\right)$ is sinc(fT) and its graph is as shown below



As $T \to 0$ the distance between successive zeros becomes infinitely large and so we may deduce (although this is not a proof) that $F[\delta(t)] = 1$.

Fourier transform of a delayed delta function



From this, if we use the time shift theorem, it follows that

$$F[\delta(t-T)] = 1 \text{ x } e^{-2j\pi fT} = e^{-2j\pi fT}.$$

Example

Use the reciprocal property to show that F[1]=δ(f)

$$F[\delta(t)] = 1$$

Using reciprocal property,

$$F[1] = \delta(f)$$

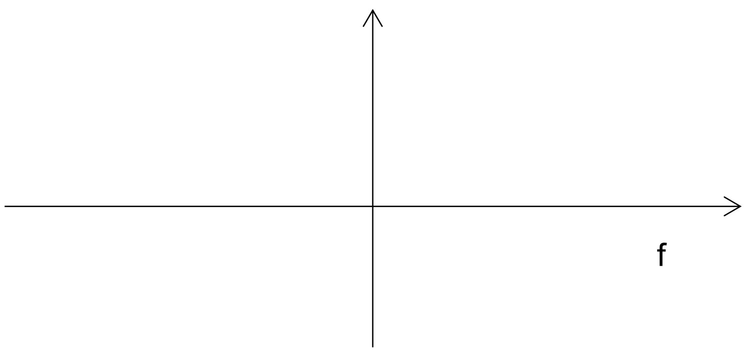
Example

Use the frequency shift property to show that $F\left[e^{2j\pi f_c t}\right] = \delta(f - f_c)$.

Example

Show that $F[\sin 2\pi f_c t] = \frac{1}{2j}[\delta(f - f_c) - \delta(f + f_c)]$.

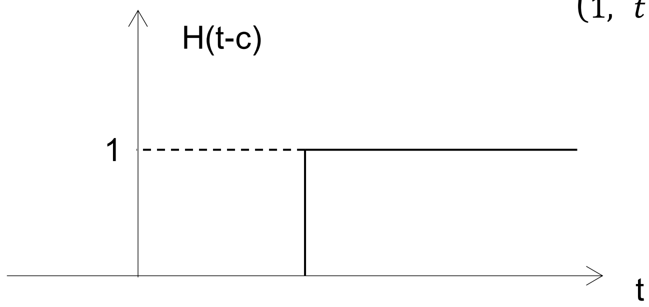
So, the FT of a sine is a pair of δ functions (of half the amplitude of the sine) at $f = \pm f_c$ where f_c is the frequency of the sine wave.



The Heaviside Step Function or Unit Step Function

This signal is usually denoted by H(t) and is defined by $H(t) = \begin{cases} 0, & t < 0 \\ 1, & t \geq 0 \end{cases}$

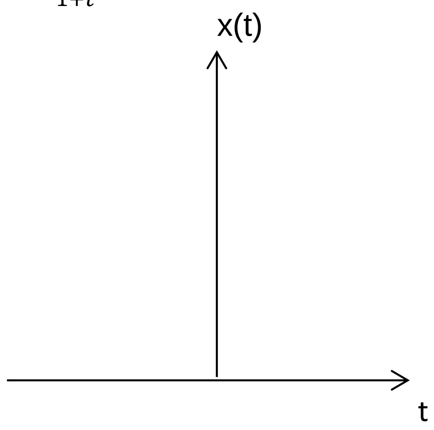
A simple extension of the definition gives $H(t) = \begin{cases} 0, & t < c \\ 1, & t \geq c \end{cases}$

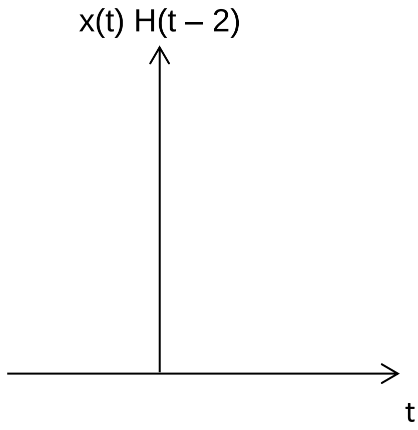


Recall: The effect of multiplying a given signal v(t) by H(t – c) is to 'switch-off' v(t) until time t = c.

<u>Example</u>

If $x(t) = \frac{1}{1+t^2}$, sketch x(t) H(t – 2).





Fourier Transform of H(t)

It can be shown that $F[H(t)] = \frac{1}{2}\delta(f) - \frac{j}{2\pi f}$

$$H(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \end{cases}$$

Compare with $\operatorname{sgn}(t) = \begin{cases} -1, & t < 0 \\ 1, & t > 0 \end{cases}$

$$\operatorname{sgn}(t) + 1 = \begin{cases} 0, & t < 0 \\ 2, & t > 0 \end{cases}$$

Therefore, $\frac{1}{2}(\operatorname{sgn}(t) + 1) = H(t)$

We can easily find the Fourier Transform: $F[H(t)] = \frac{1}{2}\left(\delta(f) - \frac{j}{\pi f}\right)$

The Fourier transform for H(t-c) is $F[H(t-c)] = \frac{1}{2}\delta(f) - \frac{j}{2\pi f}e^{-2j\pi f c}$

Key Points

- The impulse function, or delta function, δ(t), is the mathematical representation of a "spike" occurring at t = 0. A spike at t = c is written δ(t-c). By considering a limiting form of a rectangular pulse we find *F*[δ(t)] = 1 and *F*[δ(t-T)] = $e^{-2j\pi fT}$.

- The Heaviside step function H(t), jumps from zero to one at t = 0: it acts as a "switch on" at t = 0. U(t-c) switches on at t = c.

- The Fourier transform of the step function is: $F[H(t)] = \frac{1}{2}\delta(f) - \frac{j}{2\pi f}$

- The Fourier transform of the delayed step function is: $F[H(t-c)] = \frac{1}{2}\delta(f) - \frac{j}{2\pi f}e^{-2j\pi fc}$

- By combining Heaviside step functions it is possible to modify a given signal in a variety of ways. You need to understand how these modifications are obtained.

Transfer function and Convolution

The transfer function and impulse response

One of the uses of transforms is in the analysis of systems. By definition, the transfer function for a particular system is the ratio of the transform of output to that of the input ie. If the input signal is x(t) and the output signal is y(t), the transfer function is $H(f) = \frac{Y(f)}{X(f)}$
If the transform of the input signal is known then the transform of the output can be found by multiplication, as shown in the following diagram. A final inverse transform then gives the system output.

$$\begin{tikzcd}
x(t) \arrow[r, "\text{Transform}"] \arrow[d, dashed, "\text{System}"'] & X(f) \arrow[d, "\text{Transfer Function } H(f)"] \\
y(t) & Y(f) \arrow[l, "\text{Inverse Transform}"']
\end{tikzcd}$$

<u>Example</u>

What is the response of a system y(t) whose transfer function is $H(f) = \frac{-j}{\pi f}$ to the input $x(t) = \delta(t)$?

$$Y(f) = H(f)X(f)$$

$$\text{Since } F[\delta(t)] = 1$$

$$Y(f) = H(f) = \frac{-j}{\pi f}$$

$$y(t) = F^{-1}[Y(f)] = F^{-1}\left[\frac{-j}{\pi f}\right] = \operatorname{sgn}(t)$$

There is, however, another method of evaluating the response of a system more directly and it involves a process called convolution.

The Convolution Theorem

One of the main reasons why convolution is so important in the analysis of systems is that it can provide the system response to any input provided that the transfer function is known. This very important result relies on the following general result, known as the Convolution Theorem which applies to a number of transforms not just Fourier transforms.

Transform of Convolution = Product of Transforms

In the case of Fourier transforms: $F[x(t) * h(t)] = F[x(t)] \cdot F[h(t)]$

And $F^{-1}[X(f) * H(f)] = F^{-1}[X(f)] \cdot F^{-1}[H(f)]$

Example

What is the output due to an input signal $x(t) = \sin(2\pi f_0 t)$ to a system whose impulse response is h(t)=H(t)?

i) Method 1, finding the transfer function via the transforms

$$X(f) = \frac{1}{2j}\big(\delta(f-f_0) - \delta(f+f_0)\big), \qquad H(f) = \frac{1}{2}\left(\delta(f) - \frac{j}{\pi f}\right)$$

$$y(t) = F^{-1}[X(f)H(f)] = \frac{1}{4j}\int_{-\infty}^{\infty}\big(\delta(f-f_0) - \delta(f-f_0)\big)\left(\delta(f) - \frac{j}{\pi f}\right)e^{2j\pi f t}df$$

Answer

$$y(t) = -\frac{1}{4j\pi f_0}\left(e^{2j\pi f_0 t} - e^{-2j\pi f_0 t}\right) = \frac{1}{2\pi f_0} - \frac{1}{2\pi f_0}\cos(2\pi f_0 t)$$

ii) Method 2, find the output using convolution $Y(f) = X(f)H(f)$

This means that

$$y(t) = \int_{-\infty}^{\infty} x(t-\tau)h(\tau)d\tau = \int_{-\infty}^{\infty}\sin\big(2\pi f_0(t-\tau)\big)H(\tau)d\tau$$

$$= \int_0^t \sin\big(2\pi f_0(t-\tau)\big)d\tau = \int_0^t \frac{1}{2j}\left(e^{2\pi f_0(t-\tau)} - e^{-2\pi f_0(t-\tau)}\right)d\tau$$

$$= \frac{1}{2j}\left[\frac{-e^{2\pi f_0(t-\tau)}}{2j\pi f_0} - \frac{e^{-2\pi f_0(t-\tau)}}{2j\pi f_0}\right]_0^t = \frac{1}{4\pi f_0}\left[e^{-2\pi f_0(t-\tau)} + e^{2\pi f_0(t-\tau)}\right]_0^t$$

$$= \frac{1}{4\pi f_0}\left[\left(e^{-2\pi f_0(t-t)} + e^{2\pi f_0(t-t)}\right) - \left(e^{-2\pi f_0(t-0)} + e^{2\pi f_0(t-0)}\right)\right]$$

$$= \frac{1}{4\pi f_0}\left[(e^0 + e^0) - (e^{-2\pi f_0 t} + e^{2\pi f_0 t})\right] = \frac{2}{4\pi f_0} - \frac{1}{4\pi f_0}\left[e^{2\pi f_0 t} + e^{-2\pi f_0 t}\right]$$

$$= \frac{1}{2\pi f_0} - \frac{1}{2\pi f_0}\cos(2\pi f_0 t)$$

Example

Find the convolution of the following signals $u(t) = \begin{cases} 1-t, & 0 \le t < 1 \\ 0, & otherwise \end{cases}$ and $v(t) = \begin{cases} e^{-t}, & 0 \le t \\ 0, & t < 0 \end{cases}$.

$$w(t) = u(t) * v(t) = \int_{-\infty}^{\infty} u(\tau)v(t-\tau)d\tau$$

The lower limit for both these functions is zero, but the upper limits differ, this needs to be accounted for in the upper limit of the integration

$$\int_0^{\min(t,1)} (1-\tau)e^{-(t-\tau)}d\tau = \int_0^{\min(t,1)} e^{\tau-t} - \tau e^{\tau-t}d\tau$$

$$= \int_0^{\min(t,1)} e^{\tau-t} d\tau - \int_0^{\min(t,1)} \tau e^{\tau-t} d\tau = [e^{\tau-t} - \tau e^{\tau-t} + e^{\tau-t}]_0^{\min(t,1)}$$
$$= [2e^{\tau-t} - \tau e^{\tau-t}]_0^{\min(t,1)} = [(2-\tau)e^{\tau-t}]_0^{\min(t,1)}$$

Evaluating at the limit t first

$$[(2-\tau)e^{\tau-t}]_0^t = (2-t)e^{t-t} - (2-0)e^{0-t}) = 2 - t - 2e^{-t}$$

Now evaluate at the limit 1

$$[(2-\tau)e^{\tau-t}]_0^1 = (2-1)e^{1-t} - \left((2-0)e^{0-t}\right) = e^{1-t} - 2e^{-t} = (e^1 - 2)e^{-t}$$

$$w(t) = \begin{cases} 0, & t < 0 \\ 2 - t - 2e^{-t}, & 0 \le t < 1 \\ (e^1 - 2)e^{-t}, & t \ge 1 \end{cases}$$

Example

Show that the convolution between two rectangular pulses is a triangular pulse using the multiplication of the Fourier transform.

Key Points

- We use transforms in the analysis of signals.
- Transfer function of a particular system is the ration of the transform of the output to that of the input: $H(f) = \frac{Y(f)}{X(f)}$.
- Transform of convolutions =product of transforms.
- For Fourier transforms: $F[x(t) * h(t)] = F[x(t)] \cdot F[h(t)] and F^{-1}[X(f) * H(f)] = F^{-1}[X(f)] \cdot F^{-1}[H(f)]$.

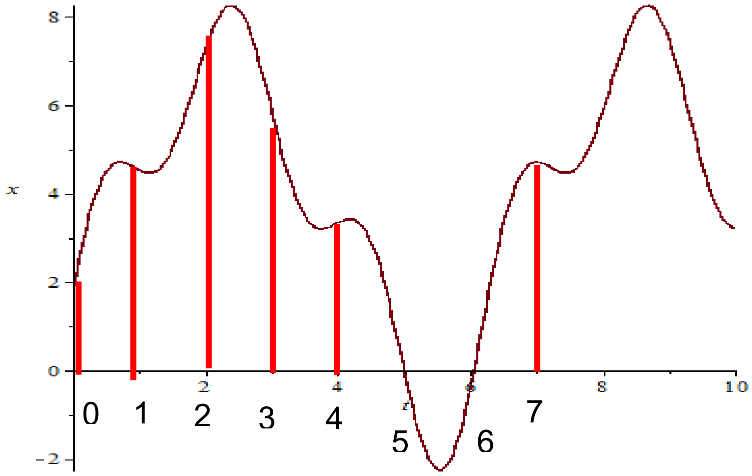
z TRANSFORMS

Introduction

The signals we have dealt with so far are continuous functions of time, $x(t)$. However, it is often the case with engineering systems that the signal we are dealing with is discrete, i.e. it is known at individual instants of time but not for all times. The most common reason for the occurrence of a discrete signal is that a continuous signal has been sampled at discrete time intervals. This is the situation we shall consider in order to introduce the concept of the z transform. The key point to grasp at this stage is that the z transform is used in the analysis and synthesis of discrete systems whereas both the Fourier and Laplace transforms are concerned with continuous time signals.

Sampled Signals and the Definition of the z transform

Suppose we have a continuous signal $x(t)$ that is sampled every T seconds. The term sampling is taken here to mean the process of measuring $x(t)$ over a short time interval Δt centred on t = 0, T, 2T, … and in this way obtaining a sequence of discrete samples v(0), v(T), v(2T), v(3T) and so on.



The sequence x(0), x(T), x(2T), x(3T) … can be written in several different ways. If, for example, it is known that sampling takes place every T seconds then the samples are sometimes written x(0), x(1), x(2) x(3) … or x_0, x_1, x_2, x_3 … or {x(n)} or $\{x_n\}$, where curly brackets indicate a sequence and n denotes the n-th member of the sequence.

Definition

The z transform of the sequence x_0, x_1, x_2 … is

$$Z[\{x_n\}] = X(z) = \sum_{n=0}^{\infty} \frac{x_n}{z^n}.$$

If the sequence originates from sampling a continuous signal $x(t)$ at intervals of length T then we can, loosely, speak of the 'z transform of x(t)'. This is defined in the obvious way:

$$Z[x(t)] = X(z) = \sum_{n=0}^{\infty} \frac{x(nT)}{z^n}.$$

Notice that the z transform operates on a discrete set of values but produces a continuous function of z, X(z).

Example

Find the z transform of the ramp function $x(t) = t$.

We can write $x(nT) = 0, T, 2T, 3T, \dots, nT, \dots$

$$X(z) = T\sum_{n=0}^{\infty}\frac{n}{z^n}$$

$$X(z) = T(\frac{1}{z} + \frac{2}{z^2} + \frac{3}{z^3} + \cdots + \frac{n}{z^n} + \cdots)$$

$$X(z) = \frac{T}{z}\left(1 + \frac{2}{z^1} + \frac{3}{z^2} + \cdots + \frac{n}{z^{n-1}} + \cdots)\right)$$

$$X(z) = \frac{T}{z}\left(\frac{1}{\left(1-\frac{1}{z}\right)^2}\right) \ where \ \left|\frac{1}{z}\right| < 1$$

$$X(z) = \frac{T}{z}\left(\frac{z^2}{(z^2-1)^2}\right) = T\frac{z}{(z-1)^2}$$

Notice the following points.

1. In the above example we needed to know the sum of a certain series. The following results will prove to be useful.

$$1 + x + x^2 + x^3 + \ldots = \frac{1}{1-x}, \qquad |x| < 1,$$

$$1 - x + x^2 - x^3 + \ldots = \frac{1}{1+x}, \qquad |x| < 1,$$

$$1 + 2x + 3x^2 + 4x^3 + \ldots = \frac{1}{(1-x)^2}, \qquad |x| < 1,$$

2. The result found in the above example involved the sampling time T. This shows that the z transform not only contains information about the original signal but also about the sampling time.

3. The result is only valid if |z| > 1. We shall say more about this type of restriction later.

z Transform Tables

From the exercises carried out with Fourier and Laplace transforms it is clear that when calculating transforms full use should be made of tables of transforms of commonly occurring functions. The same is true of z transforms. The more familiar you are with the use of the table provided the easier it will be to find z transforms and their inverses.

Example

Use the table provided to find the z transforms of the following signals:

(i) e^{-2t} sampled every second,
(ii) e^{-t} sampled every hundredth of a second,
(iii) $sin\left(\frac{\pi}{4}t\right)$, sampled every two seconds.

Key points

- Z transforms operate on sequences, not continuous signals.
- The z transform of a sequence $\{x_n\}$ is a continuous function of the variable z, X(z) and is defined by the summation $Z[\{x_n\}] = X(z) = \sum_{n=0}^{\infty} \frac{x_n}{z^k}$.
- The sequence $\{x_n\}$ often arises from sampling a continuous signal at equal time intervals.
- The z transform of a sequence $\{x_n\}$ is a continuous function of the variable z, X(z), and is defined by the summation .
- The z transform of a sequence can be evaluated directly from the definition above.
- In practice, full use is made of tables of z-transforms for finding and inverting z-transforms.

Table of Standard z Transforms

X(nT)-sampled	z Transform
a^n	$\frac{z}{z-a}$
$\delta(n)$	1
$H(n)$	$\frac{z}{z-1}$
nT	$\frac{Tz}{(z-1)^2}$
$(nT)^2$ ie when x(t)=t^2	$\frac{T^2z(z+1)}{(z-1)^3}$
e^{-anT}	$\frac{z}{z-e^{-aT}}$
$1-e^{-anT}$	$\frac{z(1-e^{-aT})}{(z-1)(z-e^{-aT})}$
nTe^{-anT}	$\frac{Tze^{-aT}}{(z-e^{-aT})^2}$
$(1-anT)e^{-anT}$	$\frac{z[z-e^{-aT}(1+aT)]}{(z-e^{-aT})^2}$
$\sin(n\omega T)$	$\frac{z\sin(\omega T)}{z^2-2z\cos(\omega T)+1}$
$\cos(n\omega T)$	$\frac{z(z-\cos(\omega T))}{z^2-2z\cos(\omega T)+1}$
$\sinh(n\omega T)$	$\frac{z\sinh(\omega T)}{z^2-2z\cosh(\omega T)+1}$
$\cosh(n\omega T)$	$\frac{z(z-\cosh(\omega T))}{z^2-2z\cosh(\omega T)+1}$
$e^{-anT}x(nT)$	$X(ze^{aT})$
$x(nT-n_0T)$	$\frac{X(z)}{z^{n_0}}$
$nx(n)$	$-z\frac{dX(z)}{dz}$
$x(nT+T)$	$zX(z)-zx(0)$
$x(nT+2T)$	$z^2X(z)-z^2x(0)-zx(T)$

Properties of z Transforms

Linearity

$$Z[a_1x_1(t) + a_2x_2(t)] = a_1Z[x_1(t)] + a_2Z[x_2(t)].$$

<u>Example</u>

Use the table provided to find the z transform of $t + e^{-3t}$ that has been sampled at intervals of length T.

$$Z[\{nT\} + \{e^{-3nT}\}] = \mathrm{Z}\big[\{\mathrm{nT}\}\big] + \mathrm{Z}[\{e^{-3nT}\}]$$

$$Z[\{nT\}] = \frac{TZ}{(z-1)^T}, \qquad Z[\{e^{-anT}\}] = \frac{z}{z-e^{-aT}}$$

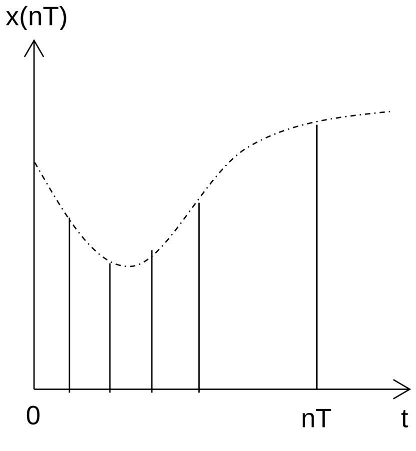
$$\mathrm{Z}\big[\{\mathrm{nT}\}\big] + \mathrm{Z}[\{e^{-3nT}\}] = \frac{Tz}{(z-1)^2} + \frac{z}{z-e^{-3T}}$$

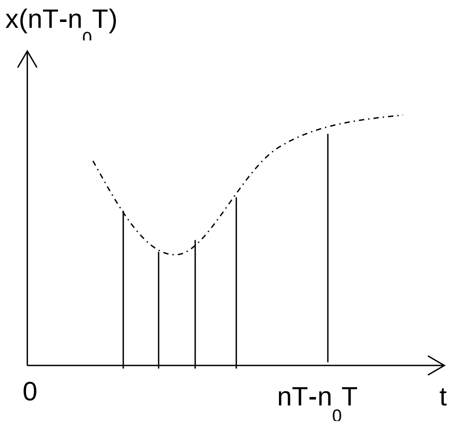
<u>Example</u>

Use the table provided to find the z transform of $e^{-3t} - 3te^{-3t}$ sampled every 2 seconds.

First Shift Theorem (time delay)

If $X(z) = Z[\{x(nT)\}]$ then $Z[\{x(nT - n_0T)\}] = \frac{1}{z^{n_0}}X(z),$ assuming $x(t) = 0 \; if \; t < 0.$





This result states that each delay of x(t) by time T multiplies the z transform by a factor $\frac{1}{z}$.

<u>Example</u>

Given x(nT) = 0, n = 0
= 1, n = 1, 2, 3, ... ,
find the z transform of x(nT – T), ie. the original signal delayed by T.

$$0 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \cdots \;\; = \frac{1}{z}\left(1 + \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \cdots\right) = \frac{1}{z} \times \frac{1}{\left(1 - \frac{1}{z}\right)}, \;\; \left|\frac{1}{z}\right| < 1$$

$$= \frac{1}{z-1}, \ |z| > 1$$

$$Z[x(nT - n_0T)] = \frac{1}{z^{n_0}} X(z)$$

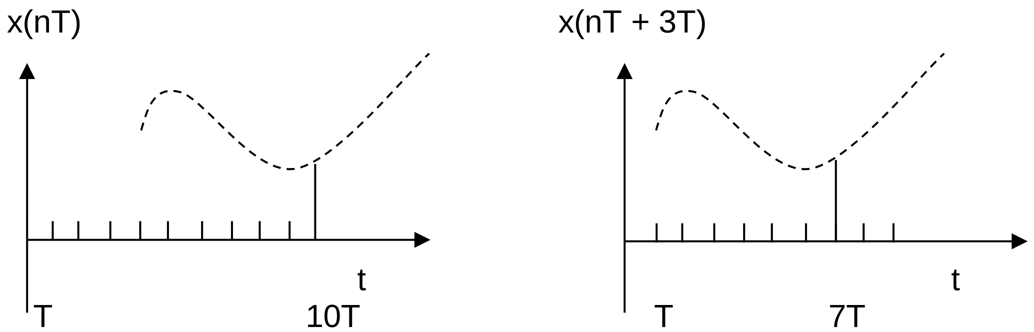
$$\frac{1}{z} \times \frac{1}{z-1} = \frac{1}{z(z-1)}$$

Example

Given $x(nT) = nT,$ find the z transform of $x(nT - 5T)$

The Second Shift Theorem (time advance)

If x(nT), n = 0, 1, 2, ... are samples of a signal x(t) then $x(nT + n_0T)$ are the samples of x(t) shifted n_0 sampling times to the left. This is equivalent to advancing the signal.



The theorem states that

$$Z[\{x(nT + n_0T)\}] = z^{n_0}X(z) - \sum_{n=0}^{n_0-1} x(nT)z^{n_0-n},$$

where

$$X(z) = Z[\{x(nT)\}].$$

We shall normally only be concerned with using this result when $n_0 = 1$ or $n_0 = 2$, for which cases we have:

$$Z[\{x(nT + T)\}] = zX(z) - zx(0),$$

and

$$Z[\{x(nT + 2T\}] = z^2X(z) - z^2x(0) - zx(T).$$

Example

What is the z transform of the sequence $\{(n+2)^2\}$?

$$x(n) = (nT + 2T)^2, \qquad Z[\{x(nT+2T)\}] = z^2X(z) - zx(T) - z^2x(0)$$

From the table of standard transforms, we have: $Z[n^2] = \frac{z(z+1)}{(z-1)^3}$
We can treat this sequence as $x(n) = (nT + 2T)^2, \; where \; T = 1$

Using the rule for advanced signal with n_0=2

$$Z[(n+2)^2] = \frac{z^3(z+1)}{(z-1)^3} - zx(1) - z^2x(0)$$

$$x(1) = 1 \text{ and } x(0) = 0$$

$$= \frac{z^3(z+1)}{(z-1)^3} - z$$

Complex Translation Theorem

$Z[\{e^{-anT}x(nT)\}] = X(e^{aT}z)$ where $X(z) = Z[\{v(nT)\}]$.

Example

Use the above theorem to find the z transform of the signal v(t) = te^{-at} sampled every T seconds.

$$Z[(nT)] = \frac{Tz}{(z-1)^2}$$

$$Z[x(nT)e^{-anT}] = \frac{Tze^{aT}}{(ze^{aT}-1)^2} = \frac{Tze^{-aT}}{(z-e^{-aT})^2}$$

Example

Use the complex translation theorem to find the z transform of the signal $x(t) = t^2e^{-3t}$ sampled every T seconds.

Derivative of a z Transform

$$Z[\{x(nT)\}] = x(0) + \frac{x(T)}{z^1} + \frac{x(2T)}{z^2} + \frac{x(3T)}{z^3} + \cdots$$

$$\frac{dZ[\{x(nT)\}]}{dz} = -\frac{x(T)}{z^2} - \frac{2x(2T)}{z^3} - \frac{3x(2T)}{z^4} - \cdots$$

$$= -\frac{1}{z}\sum_0^{\infty}\frac{nx(nT)}{z^n}$$

$$= -\frac{1}{z}Z[\{nx(nT)\}]$$

Example

Using $Z[a^n] = \frac{z}{z-a}$, demonstrate that $Z[na^n] = \frac{az}{(z-a)^2}$

From the derivative rule: $Z[\{nx(nT)\}] = -z\frac{dZ[\{x(nT)\}]}{dz}$

The derivative of $Z[a^n]$

$$\frac{d}{dz}\left(\frac{z}{z-a}\right) = \frac{(z-a)-z}{(z-a)^2} = \frac{-a}{(z-a)^2}$$

Multiply by –z to get the result

Example

Find the z transform of $n2^n$.

Key Points

- There are several useful properties of z transforms. The ones we have concentrated on are the first shift theorem (time delay), the second shift theorem (time advance), complex translation and the derivative.
- Linearity: $Z[a_1x_1(nT) + a_2x_2(nT)] = a_1Z[x_1(nT)] + a_2Z[x_2(nT)]$.
- First shift theorem (time delay): $[\{x(nT - n_0T\}] = \frac{1}{z^{n_0}}X(z)$.
- Second shift theorem (time advance): $Z[\{x(nT + n_0T)\}] = z^{n_0}Z[\{x(nT)\}] - \sum_0^{n_0-1} x(nT)z^{n_0-n}$.
- Complex translation theorem: $Z[\{e^{-anT}x(nT)] = X(e^{aT}\mathrm{z})$.
- Derivative of a z transform: $\frac{dZ[\{x(nT)\}]}{dz} = -\frac{1}{z}Z[\{nx(nT)\}]$

Inversion of z-Transforms

As with Fourier and Laplace transforms, we often want to move from the transformed description of a system back to the time domain, i.e. we need to recover x(nT) from X(z). This is the process of inverse transformation:

$$x(nT) = Z^{-1}[X(z)].$$

Example

Find $Z^{-1}\left[\frac{z}{z-2}\right]$.

Like the Laplace transform, we will be using the table to invert, but again, the signals will not always be presented in the correct form to simply read off the table. The method we will be employing to invert z transforms is partial fractions.
For the z transform there are three cases that we shall consider:

- Where one power of z is present in the numerator
- Where more than one power of z is present in the numerator
- Where there are no powers of z in the numerator

If you look at the table of transforms, except for the delta function, all transforms have a power of z in the numerator. We need to make sure that our method for inversion preserves this.

Example

Find $Z^{-1}\left[\frac{z}{(z-1)(z-2)}\right]$.

$$\frac{V(z)}{z} = \frac{1}{(z-1)(z-2)}$$

$$\frac{1}{(z-1)(z-2)} = \frac{A}{z-1} + \frac{B}{z-2}$$

$$1 = A(z-2) + B(z-1)$$

When z=1: $1 = A(-1),\ \ A = -1$
When z=2: $1 = B(1),\ \ \ B = 1$

$$\frac{V(z)}{z} = \frac{1}{(z-1)(z-2)} = \frac{-1}{z-1} + \frac{1}{z-2}$$

$$V(z) = \frac{-z}{z-1} + \frac{z}{z-2}$$

$$Z^{-1}[V(z)] = \{2^n - 1^n\} = 0,1,3,7 \ldots$$

Example

Find $Z^{-1}\left[\frac{z}{(z-5)(z-3)}\right]$

Example

Find $Z^{-1}\left[\frac{z^2}{(z+1)(z+2)}\right]$.

$$\frac{V(z)}{z}=\frac{z}{(z+1)(z+2)}$$

$$\frac{z}{(z+1)(z+2)}=\frac{A}{z+1}+\frac{B}{z+2}$$

$$z=A(z+2)+B(z+1)$$

When z=-1: $-1=A(1)$, $\quad A=-1$
When z=-2: $-2=B(-1)$, $\quad B=2$

$$\frac{V(z)}{z}=\frac{z}{(z+1)(z+2)}=\frac{-1}{z+1}+\frac{2}{z+2}$$

$$V(z)=\frac{2z}{z+2}-\frac{z}{z+1}$$

$$Z^{-1}[V(z)]=\{2(-2^n)-(-1^n)\}=1,-3,7,-15,31\ \ldots$$

Example

Find $Z^{-1}\left[\frac{z^2}{(z+5)(z+3)}\right]$

Example

Find $Z^{-1}\left[\frac{1}{(z-1)(z-2)}\right]$.

$$\frac{V(z)}{z} = \frac{1}{z(z-1)(z-2)}$$

$$\frac{1}{z(z-1)(z-2)} = \frac{A}{z-1} + \frac{B}{z-2} + \frac{C}{z}$$

$$1 = Az(z-2) + Bz(z-1) + C(z-1)(z-2)$$

When z=1: $1 = A(1)(-1)$, $A = -1$
When z=2: $1 = B(2)(1)$, $B = \frac{1}{2}$
When z=0: $1 = C(-1)(-2)$ $C = \frac{1}{2}$

$$\frac{V(z)}{z} = \frac{1}{z(z-1)(z-2)} = \frac{-1}{z-1} + \frac{1}{2(z-2)} + \frac{1}{2z}$$

$$V(z) = \frac{-1z}{z-1} + \frac{z}{2(z-2)} + \frac{z}{2z}$$

$$Z^{-1}[V(z)] = \left\{\frac{1}{2}(2^n) - (1^n) + \frac{1}{2}\delta(n)\right\}$$

Example

Find $Z^{-1}\left[\frac{1}{(z-5)(z-3)}\right]$

Solution of Difference Equations using z Transforms

We have seen that LTI systems varying continuously with time t can be described by ordinary differential equations. For discrete systems the governing equation is called a difference equation and relates samples at successive sampling times (rather than derivatives of different orders). A simple example of a difference equation is

$$x_{n+2} = x_{n+1} + x_n, \qquad n \geq 1,$$

which generates the well-known Fibonacci sequence if we take $x_0 = 1$, $x_1 = 1$. The z transform can be used to solve difference equations in much the same way that Laplace transforms can be used to solve ordinary differential equations.

Example

Solve the Fibonacci difference equation using z transforms.

$$x_{n+2} = x_{n+1} + x_n,\ \ n \geq 0,\ \ x_0 = x_1 = 1$$

Take the z transform of each term and use the second shift theorem

$$Z[\{x_n\}] = X(z), \qquad Z[\{x_{n+1}\}] = zX(z) - zx(0), \qquad Z[\{x_{n+2}\}] = z^2X(z) - z^2x(0) - zx(T)$$

The z transform of the difference equation is then

$$z^2X(z) - z^2 - z = zX(z) - z + X(z)$$

Rearrange this to make $X(z)$ the subject gives $X(z) = \frac{z^2}{z^2-z-1}$

$$\frac{X(z)}{z} = \frac{z}{z^2 - z - 1} = \frac{A}{z-\alpha} + \frac{B}{z-\beta} \text{ where } \alpha = \frac{1-\sqrt{5}}{2},\ \beta = \frac{1+\sqrt{5}}{2}$$

$$z = A(z-\beta)\ + B(z-\alpha)$$

Equate coefficients

$$z{:}\, 1 = A + B$$

$$1{:}\, 0 = -\beta A - \alpha B, \qquad \beta A = -\alpha B, \qquad A = \frac{-\alpha}{\beta}B$$

$$1 = \frac{-\alpha}{\beta}B + B, \qquad \beta = B(\beta - \alpha), \qquad B = \frac{\beta}{\beta - \alpha}$$

The result of X(z): $X(z) = -\frac{\alpha}{\sqrt{5}} \cdot \frac{z}{z-\alpha} + \frac{\beta}{\sqrt{5}} \cdot \frac{z}{z-\beta}$

$$x_n = -\frac{\alpha}{\sqrt{5}}\alpha^n + \frac{\beta}{\sqrt{5}}\beta^n = \frac{\alpha^{n+1}}{\sqrt{5}} + \frac{\beta^{n+1}}{\sqrt{5}}$$

$$= \frac{1}{\sqrt{5}}\left[\left(\frac{1-\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1+\sqrt{5}}{2}\right)^{n+1}\right]$$

Example

Solve the difference equation

$$x_{n+2} - 3x_{n+1} - 4x_n = 0,$$

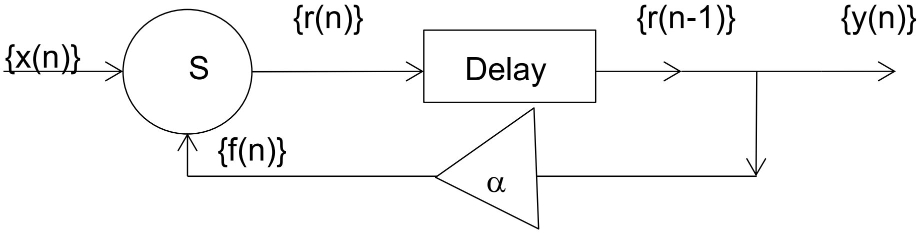
where $x_0 = 0$ and $x_1 = 1$.

Key Points

- We can find the inverse of a z transform using the standard table of transforms.
- We use partial fractions to manipulate our z transform into a form which we can read of our table.
- Digital systems, (strictly, linear, time invariant, digital systems) are described by difference equations and may be analysed by means of the z-transform.

Application to Digital Systems

Consider the discrete-time feedback system shown in the diagram below.



The components of the system are a summing unit S, a delay D and (negative) feedback loop with amplification α. The input sequence is x(0), x(1), x(2), ..., or {x(n)}, and the output sequence is y(0), y(1), y(2) ... or {y(n)}. The input signal and the feedback signal, f(k) say, combine to give an input r(n) = x(n) – f(n) to the delay. The output from the delay is r(n – 1). In summary:

y(n) = r(n – 1),
r(n) = x(n) – α r(n – 1).

If we take z transforms of these equations, assuming zero initial conditions, and use the first shift theorem then we have

$$Y(z) = \frac{1}{z}R(z),$$

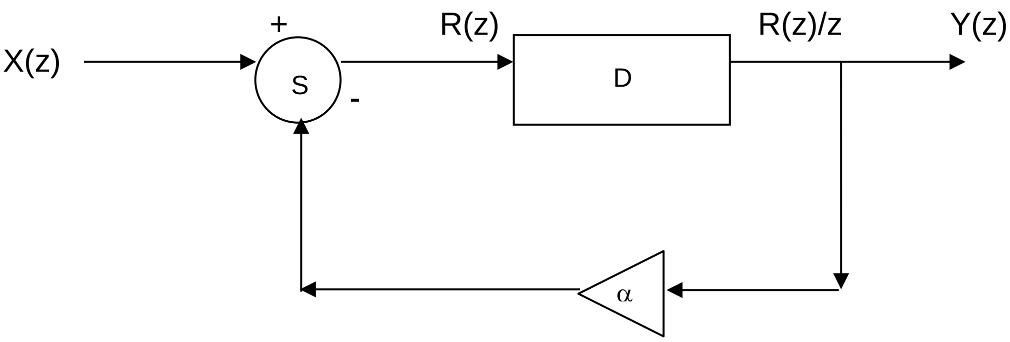
$$R(z) = X(z) - \frac{\alpha}{z}R(z) = X(z) - \alpha Y(z),$$

Where $Y(z) = Z[\{y(n)\}]$ etc. Elimination of R(z) gives

$$zY(z) = X(z) - \alpha Y(z),$$

$$Y(z) = X(z) \cdot \frac{1}{z+\alpha}$$

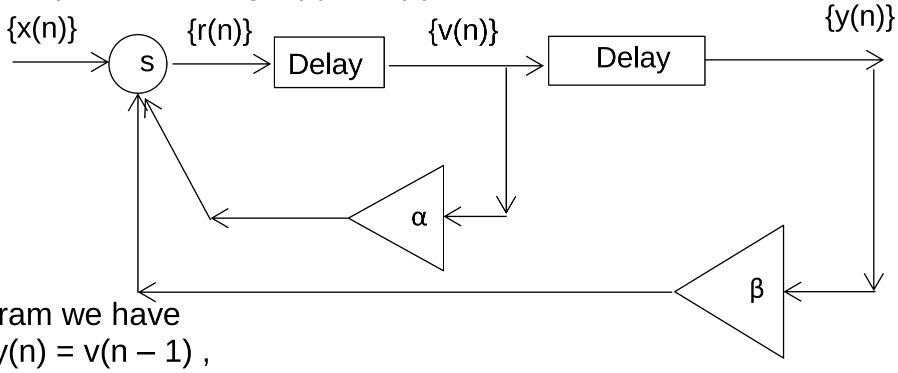
If x(n) were known then we could find X(z), then Y(z), and invert to find the output y(n). It is possible to miss out the stage of writing down the difference equations and to work directly with the z transforms, as shown in the diagram below:



Note that the introduction of an additional variable, R(z) in this case, which will later be eliminated, is usually necessary. The above example involves one delay block and gives rise to a first order system. The following example involves a second order system.

Example

Write down the difference equations relating the input sequence {x(n)}, the output sequence {y(n)} and the 'intermediate' sequences {r(n)} and {v(n)} for the system shown below. Find an equation relating Y(z) to X(z).



From the diagram we have

$$y(n) = v(n-1),$$
$$v(n) = r(n-1),$$
$$r(n) = x(n) - \alpha\, v(n) - \beta\, y(n).$$

Take z transforms:

$$Y(z) = \frac{1}{z}V(z), \qquad (1)$$

$$V(z) = \frac{1}{z}R(z), \qquad (2)$$

$$R(z) = X(z) - \alpha V(z) - \beta Y(z), \qquad (3)$$

(1) $\Rightarrow$ $V(z) = zY(z)$,
(2) $\Rightarrow$ $R(z) = zV(z) = z^2 Y(z)$,
(3) $\Rightarrow$ $z^2 Y(z) = X(z) - \alpha z Y(z) - \beta Y(z)$,

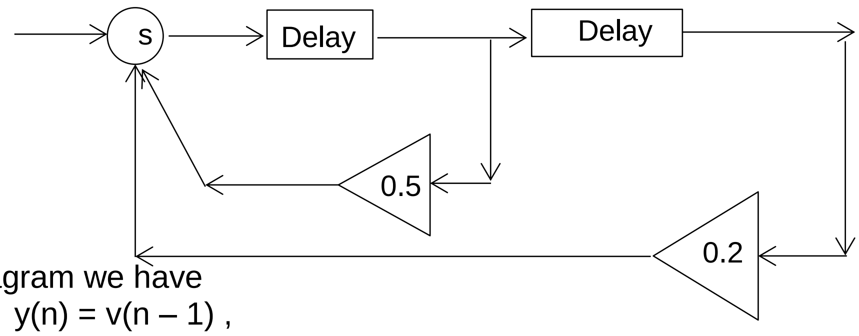
$$\therefore Y(z)(z^2 + \alpha z + \beta) = X(z),$$

$$Y(z) = \frac{X(z)}{z^2 + \alpha z + \beta}.$$

Equations (1), (2) and (3) could have been written down immediately if we had worked directly with z transforms.

Example

Write down the difference equation for the negative feedback system shown below.



From the diagram we have

$$y(n) = v(n-1),$$
$$v(n) = r(n-1),$$
$$r(n) = x(n) - 0.5\, v(n) - 0.2\, y(n).$$

Take z transforms:

$$Y(z) = \frac{1}{z}V(z), \qquad (1)$$

$$V(z) = \frac{1}{z}R(z), \qquad (2)$$

$$R(z) = X(z) - 0.5V(z) - 0.2Y(z), \qquad (3)$$

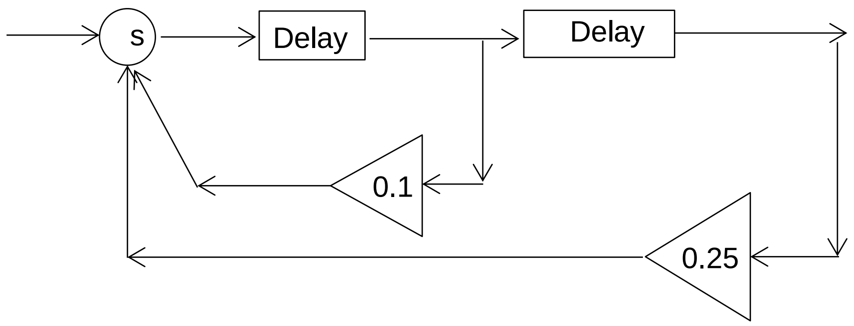
(1) $\Rightarrow$ $V(z) = zY(z)$,
(2) $\Rightarrow$ $R(z) = zV(z) = z^2Y(z)$,
(3) $\Rightarrow$ $z^2Y(z) = X(z) - 0.5zY(z) - 0.2Y(z)$,

$$\therefore Y(z)(z^2 + 0.5z + 0.2) = X(z),$$

$$Y(z) = \frac{X(z)}{z^2 + 0.5z + 0.2}.$$

<u>Example</u>

Write down the difference equation for the negative feedback system shown below.



The Digital Transfer Function

In the Laplace transform notes we defined the transfer function for an LTI continuous system by

$$G(s) = \frac{Y(s)}{X(s)}.$$

For a digital system the same relationship holds but with z transforms:

$$G(z) = \frac{Y(z)}{X(z)}.$$

G(z) is called the digital transfer function, or pulse transfer function. The function G(z) contains information about the system and does not depend on the input or output. We have already derived some examples of G(z) and in general this function takes the form

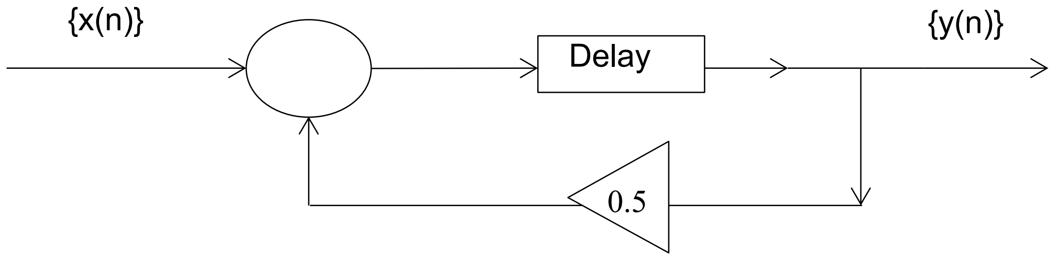
$$G(z) = \frac{Y(z)}{X(z)} = \frac{b_m z^m + b_{m-1} z^{m-1} + \cdots + b_1 z + b_0}{a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0},$$

$$= \frac{P(z)}{Q(z)},$$

where P(z) is a polynomial of degree m (the numerator) and Q(z) is a polynomial of degree n (the denominator).

__Example__

Find the transfer function of the system represent by the block diagram below



$$Y(z) = \frac{1}{z} R(z)$$

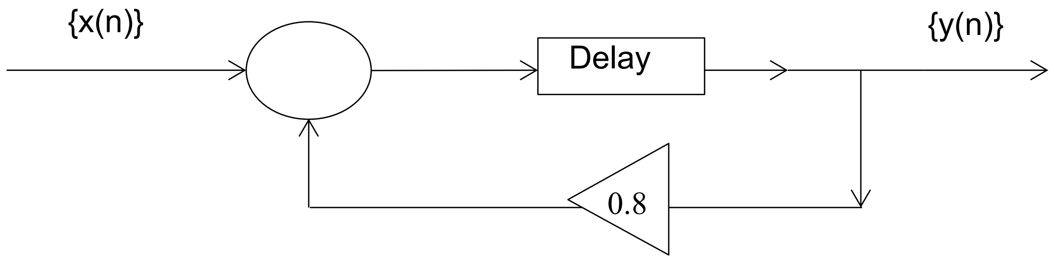
$$R(z) = X(z) - 0.5Y(z)$$

$$zY(z) = X(z) - 0.5Y(z)$$

$$G(z) = \frac{Y(z)}{X(z)} = \frac{1}{z + 0.5}$$

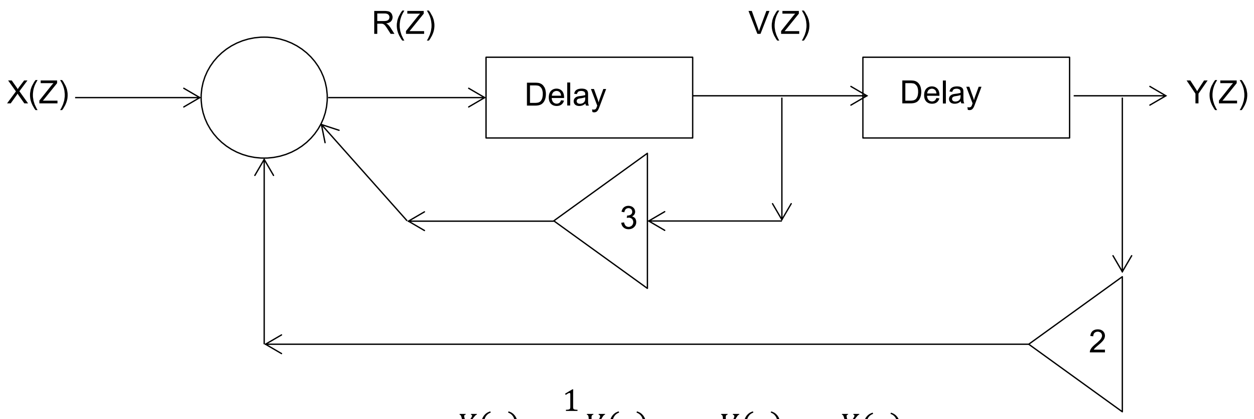
Example

Find the transfer function of the system represent by the block diagram below



Example

Find the transfer function of a two feedback loop



$$Y(z) = \frac{1}{z}V(z), \qquad V(z) = zY(z)$$

$$V(z) = \frac{1}{z}R(z), \qquad R(z) = zV(z) = z^2Y(z)$$

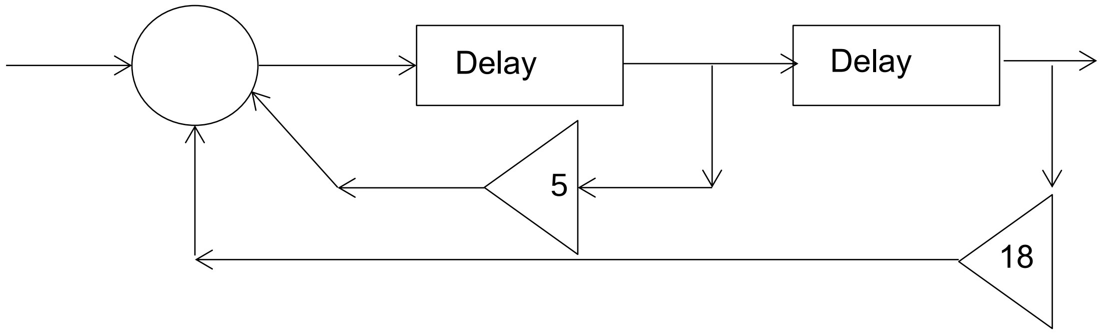
$$R(z) = X(z) - 3V(z) - 2Y(z), \qquad z^2Y(z) = X(z) - 3zY(z) - 2Y(z)$$

$$Y(z)(z^2 + 3z + 2) = X(z)$$

$$G(z) = \frac{Y(z)}{X(z)} = \frac{1}{z^2 + 3z + 2}$$

Example

Find the transfer function of a negative two feedback loop



Poles, Zeros and Stability

The roots of P(z) are called zeros and the roots of Q(z) are called poles. As with Laplace transforms, the position of the poles and zeros in the complex z plane determine the system stability. However, the stability condition is that each pole must lie inside the unit circle $|z| = 1$ in the complex z-plane.

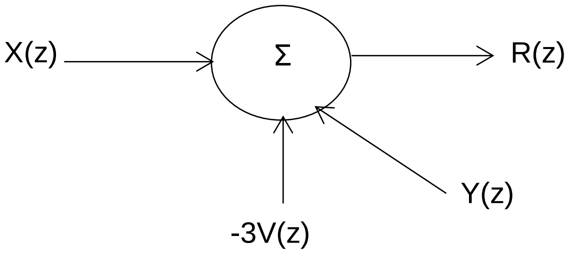
The stability of a digital system can be determined from the position of the poles and zeros of G(z) in the complex z plane. A system is stable if for every bounded input there is a corresponding bounded output. Consider a system which has the impulse response, $h(nT) = e^{anT} = e^{(\sigma+j\omega)nT}$. For stability, Re[a]=σ must be negative. The corresponding transfer function is $H(z) = Z[e^{anT}] = \frac{z}{z-e^{aT}}$, which has poles at
$z = e^{aT} = e^{(\sigma+j\omega)T} = e^{\sigma T}e^{j\omega T} = e^{\sigma T}\left(\cos(\omega T) + j\sin(\omega T)\right)$.

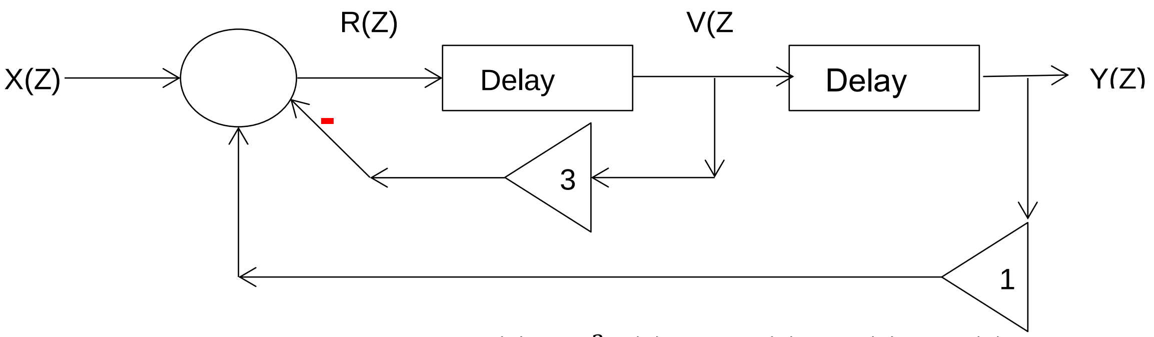
$z = \cos(\omega T) + j\sin(\omega T)$. This is the unit circle in the complex plane. The factor $e^{\sigma T}$ is a constant for a particular system that affects the magnitude of the circle in the complex plane. If σ=Re[a]<0, when the system is stable, 0<$e^{\sigma T}$<1. for Re[a]>0, $e^{\sigma T}$>1.

The stability condition is that each pole must lie inside the unit circle |z|=1 in the complex z plane. Remember that our series only converge for |z|<1.
At the zeros of the transfer function, Y(z) is zero i.e. there is no output signal so the system is unusable.

Example

Draw a block diagram to represent the system whose digital transfer function is given implicitly by z^2 Y(z) + 3zY(z) – Y(z) = X(z). Find the poles and zeros of the system and comment on its stability.





We are given the transfer function: $G(z) = z^2Y(z) + 3zY(z) - Y(z) = X(z)$.
Stability is found when the poles are zero:

$$z^2Y + 3z - 1 = 0$$

$$z = \frac{-3 \pm \sqrt{13}}{2}$$

$$z = 0.3, \qquad z = -3.3$$

Although 0.3 lies with the unit circle, -3.3 lies outside the unit circle, therefore the system is unstable.

Example

Draw a block diagram to represent the system whose digital transfer function is given implicitly by z^2 Y(z) -0.3zY(z) – 0.1Y(z) = X(z). Find the zeros of the system and comment on its stability.

The unit impulse function

For a continuous system the definition of the delta function was as follows:

$$\delta(t) = \begin{cases} \infty, & t = 0 \\ 0, & t \neq 0 \end{cases} \; and \; \int_{-\infty}^{\infty} \delta(t) = 1$$

For a digital system we have to alter this definition to:

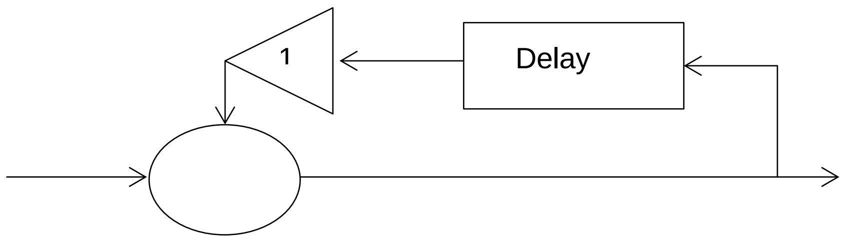
$$\sum_{n=0}^{\infty} \delta(n) = 1 \; and \; \delta(n) = 0 \quad \text{if } n \neq 0. \quad \text{Hence } \delta(0) = 1$$

Taking the z transform of this

$$Z(\delta(n)) = \sum_{n=0}^{\infty} \frac{\delta(n)}{z^n} = \frac{1}{1} + \frac{0}{z} + \frac{0}{z^2} + \frac{0}{z^3} + \cdots = 1$$

<u>Example</u>

Investigate the stability of the filter. Consider the impulse input. What is the output?



$$X(z) - \frac{1}{z} Y(z) = Y(z)$$

Transfer function: $G(z) = \frac{z}{z+1}$
Stability is found when the poles are zero:

$$z + 1 = 0$$
$$\text{z} = -1$$

This values lies on the unit circle, therefore the system is stable.

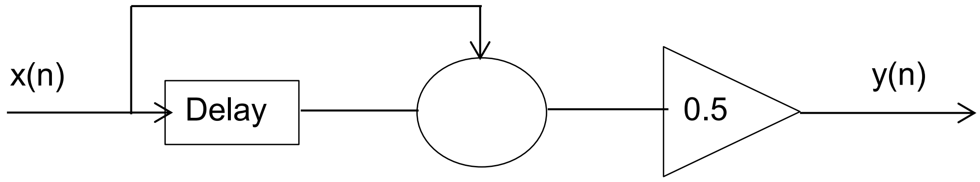
$$Y(z) = G(z)X(z), \qquad x(n) = \delta(n) \to X(z) = 1$$
$$Y(z) = G(z) \to Z^{-1}[Y(z)] = Z^{-1}\left[\frac{z}{z+1}\right]$$
$$\{y(n)\} = Z^{-1}\left[\frac{z}{z+1}\right] = (-1)^n = 1, -1, 1, -1 \ldots$$

<u>Example</u>

What is the impulse response of the system whose transfer function is $G(z) = \frac{6}{6z^2 - 5z + 1}$

Example

Find the transfer function for the low pass filter shown below and the response to the impulse input.



Frequency Response

To obtain the frequency response of a digital system from its transfer function by setting $z = e^{2j\pi f}$. We can then find its amplitude and phase spectra in the same way that we did for continuous signals.

Example

Given the transfer function $G(z) = \frac{z}{z-a}$, what is the frequency response?

$$G(z) = \frac{1}{1 - az^{-1}} = \frac{1}{1 - ae^{-2j\pi f}}$$

Amplitude: $|G(z)| = \sqrt{\mathcal{R}^2 + \mathcal{I}^2}$

Real part: $\frac{1-a\,cos(2\pi f)}{(1-acos(2\pi f))^2 + a^2 sin(2\pi f)^2}$

Imaginary part:– $\frac{-a\,sin(2\pi f)}{(1-acos(2\pi f))^2 + a^2 sin(2\pi f)^2}$

$$\sqrt{\left(\frac{1 - a\cos(2\pi f)}{(1 - \mathrm{a}\cos(2\pi f))^2 + a^2\sin(2\pi f)^2}\right)^2 + \left(\frac{-a\sin(2\pi f)}{(1 - \mathrm{a}\cos(2\pi f))^2 + a^2\sin(2\pi f)^2}\right)^2}$$

$$= \sqrt{\frac{(1 - a\cos(2\pi f))^2 + (a\sin(2\pi f))^2}{((1 - \mathrm{a}\cos(2\pi f))^2 + a^2\sin(2\pi f)^2)^2}} = \sqrt{\frac{(1 - acos(2\pi f))^2 + a^2 sin(2\pi f)^2}{((1 - acos(2\pi f))^2 + a^2 sin(2\pi f)^2)^2}}$$

Example

Given the transfer function $G(z) = \frac{z}{z-3}$, what is the frequency response?

Systems in Series

Consider the system represented by the diagram below in which the sampled input to system 1 produces a digital output which is then fed as input to system 2. The overall transfer function for the complete digital system is

$$\longrightarrow \boxed{G_1(z)} \longrightarrow \boxed{G_2(z)} \longrightarrow$$

$$G_A(z) = G_1(z) \times G_2(z).$$

But if we want the complete system to be the discrete version of two continuous systems with transfer functions $G_1(s)$ and $G_2(s)$ then the overall digital transfer function is not $G_1(z)$ x $G_2(z)$, as the following example explains.

<u>Example</u>

Find the overall pulse transfer function for a system consisting of two systems in series with transfer functions $\frac{1}{s+1}$ and $\frac{1}{s+2}$.

$$G(s) = \frac{1}{s+1} \times \frac{1}{s+2} = \frac{1}{(s+1)(s+2)}$$

$$1 = A(s+2) + B(s+1)$$
$$0 = A + B \; and \; 1 = 2A + B, \;\; 1 = 2A - A = A, \;\; so \; B = -1$$
$$= \frac{1}{s+1} - \frac{1}{s+2}$$

The impulse response function h(t) is therefore: $h(t) = e^{-t} - e^{-2t}$

And the sampled output sequence is $\{h_n\} = \{e^{-nT} - e^{-2nT}\}$

The z transform of this output sequence is

$$G(z) = Z[\{e^{-nT}\}] - Z[\{e^{-2nT}\}] = \frac{z}{z - e^{-T}} - \frac{z}{z - e^{-2T}} = \frac{z(e^{-T} - e^{-2T})}{(z - e^{-T})(z - e^{-2T})}$$

<u>Example</u>

Find the overall pulse transfer function for a system consisting of two systems in series with transfer functions $\frac{1}{s}$ and $\frac{1}{s-4}$.

Key Points

- Digital systems, (strictly, linear, time invariant, digital systems) are described by difference equations and may be analysed by means of the z-transform.

- A digital transfer function, $G(z)$, can be defined in the usual way, i.e. as the ratio of the z-transform of the output to the z-transform of the input.

- In general, $G(z)$ has poles and zeros (c.f. $G(s)$ with Laplace transforms) and, for stability, each pole must lie inside the unit circle, $|z| = 1$, in the complex z plane.

- Systems can be combined in series but care has to be taken when deriving the overall transfer function.

Convolution

As with the Laplace and Fourier transformations, the convolution theorem holds for the z transform:

Transform of Convolution = Product of Transforms

In the case of z transforms: $Z[x(t) * h(t)] = Z[x(t)] \cdot Z[h(t)]$

And $Z^{-1}[X(f) * H(f)] = Z^{-1}[X(f)] \cdot Z^{-1}[H(f)]$

For discrete-time or discrete-frequency functions, an equivalent to the convolutions integral exists; the convolution summation:

$$f(kT) * h(kT) = \sum_{m=-\infty}^{\infty} f(mT) \cdot h(kT - mT)$$

The process of convolution to obtain the system output sequence $y(k)$ for a given input sequence $x(k)$ then merely involves multiplication, delay and addition:

$$y(k) = \sum_{m=-\infty}^{\infty} x(m) \cdot h(k - m)$$

Where $h(m)$is the impulse response of the system.

<u>Example</u>

If we have the signals $x(n)$ and $h(n)$ which are defined as

$$x(n) = 2\delta(n) - 3\delta(n-2) + 4\delta(n-3)$$
$$h(n) = \delta(n) + 2\delta(n-1) + \delta(n-2)$$

Find $y(n)$.

$$y(n) = F^{-1}[X(z)H(z)]$$

$$X(z) = 2 - 3z^{-2} + 4z^{-3}$$
$$H(z) = 1 + 2z^{-1} + z^{-2}$$

$$X(z)H(z) = (2 - 3z^{-2} + 4z^{-3})(1 + 2z^{-1} + z^{-2})$$
$$= 2 + 4z^{-1} + 2z^{-2} - 3z^{-2} - 6z^{-3} - 3z^{-4} + 4z^{-3} + 8z^{-4} + 4z^{-5}$$
$$= 2 + 4z^{-1} - z^{-2} - 2z^{-3} + 5z^{-4} + 4z^{-5}$$

$$= 2\delta(n) + 4\delta(n-1) - \delta(n-2) - 2\delta(n-3) + 5\delta(n-4) + 4\delta(n-5)$$

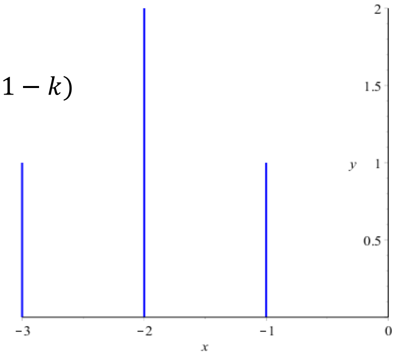
<u>Example</u>

Using the discrete convolution formula and the signals from the previous example, illustrate the relationship between convolution and multiplication.

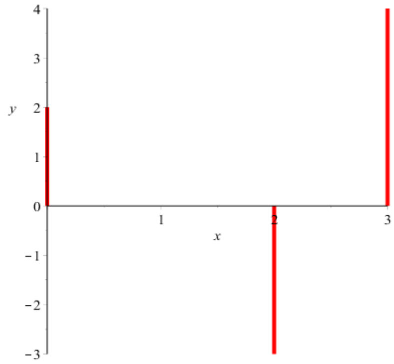
$$x(n) = 2\delta(n) - 3\delta(n-2) + 4\delta(n-3) = \{2,0,-3,4\}$$
$$h(n) = \delta(n) + 2\delta(n-1) + \delta(n-2) = \{1,2,1\}$$

we then choose one of these signals to ‘flip’ or reflect in the y axis.

$h(-1-k)$



$x(k)$



$$y(-1) = 1 \times 0 + 2 \times 0 + 1 \times 0 + 0 \times 0 + 0 \times -3 + 0 \times 4 = 0$$

$$y(0) = 1 \times 0 + 2 \times 0 + 1 \times 2 + 0 \times 0 + 0 \times -3 + 0 \times 4 = 2$$

$$y(1) = 1 \times 0 + 2 \times 2 + 1 \times 0 + 0 \times -3 + 0 \times 4 = 4$$

$$y(2) = 1 \times 2 + 2 \times 0 + 1 \times -3 + 0 \times 4 = 2 - 3 = -1$$

$$y(3) = 0 \times 2 + 1 \times 0 + 2 \times -3 + 1 \times 4 = -6 + 4 = -2$$

$$y(4) = 0 \times 2 + 0 \times 0 + 1 \times -3 + 2 \times 4 + 1 \times 0 = -3 + 8 = 5$$

$$y(5) = 0 \times 2 + 0 \times 0 + 0 \times -3 + 1 \times 4 + 2 \times 0 + 1 \times 0 = 4$$

$$y(6) = 0 \times 2 + 0 \times 0 + 0 \times -3 + 0 \times 4 + 1 \times 0 + 2 \times 0 + 1 \times 0 = 0$$

$$y(n) = \{2,4,-1,-2,5,4\}$$
$$= 2\delta(n) + 4\delta(n-1) - \delta(n-2) - 2\delta(n-3) + 5\delta(n-4) + 4\delta(n-5)$$

As before.

Example

Given: $h(n) = \{0.5,2,2.5,1\}$ and $x(n) = \{1,1,1\}$, use the convolution summation to find $y(n) = \sum x(n-k)h(k)$

Example

Given: $h(n) = \{0.5,2,2.5,1\}$ and $x(n) = \{1,1,1\}$, use the convolution summation to find $y(n) = \sum x(k)h(n-k)$, ie, flip $h(n)$ to illustrate that either function can be flipped.

When we take the convolution of two signals, the number of terms in the result is $y_i = y_1, y_2, y_3, \dots, y_{m+n-1}$. There are $m+n-1$ samples in the sequence.

Key Points

- We can find the convolution of two digital signals and the convolution theorem still applies.
- To find the convolution of two digital signals we use the convolution summation: $f(kT) * h(kT) = \sum_{m=-\infty}^{\infty} f(mT) \cdot h(kT - mT)$
- When using the summation, it does not matter which signal we choose to 'flip'.

Statistics

Descriptive Statistics

Introduction

Statistics is the branch of mathematics that involves the collection and organisation of data. Specifically, a statistic is a number derived from a set of data. The power of statistics is the ability to use a sample of data taken from a large population to make inferences about the entire population.

Definitions

Population: the entire set of all things that have the property we are measuring;
Sample: the group of things randomly selected from the population, which are measured;
Individuals: the members of our sample group;
Variables: the properties of the individuals which we count or measure;
Data: the values of the variables for each individual in our sample;
Frequency: the number of pieces of data that lie within a certain range of values;
Relative frequency: the fraction or proportion of a piece of data occurs. To find the relative frequencies, divide each frequency by the total number of pieces;
Cumulative relative frequency: the accumulation of the previous relative frequencies. To find the cumulative relative frequencies, add all the previous relative frequencies to the relative frequency for the current row.

When we examine variables, we are interested in their distribution. This distribution of a variable tells us what values it takes and how often it takes these values. This chapter presents methods for describing a single variable.

Data types

Data falls into two different categories:

- Qualitative
  - Binary (yes/no, true/false, success/failure, etc.)
  - Categorical (blood type, political party, word, etc.)
  - Ordinal (relative score, significant only for creating a ranking; Example: Is your general health: 1. Poor 2. Reasonable 3. Good 4. Excellent)
- Quantitative
  - Discrete: binomial (number of successes out of N possible), count/frequency (number of items in a given interval/area/volume)
  - Continuous: real-valued (temperature, relative distance, location parameter, etc.)

Graphs for discrete and categorical variables

A proper choice of graph depends on the nature of the variable.

- For categorical data, we use bar graphs. Bar charts should also be used for ungrouped discrete data.
- Histograms are only used for grouped data.
- Stem and leaf displays and frequency polygons are used for continuous data.

Bar charts

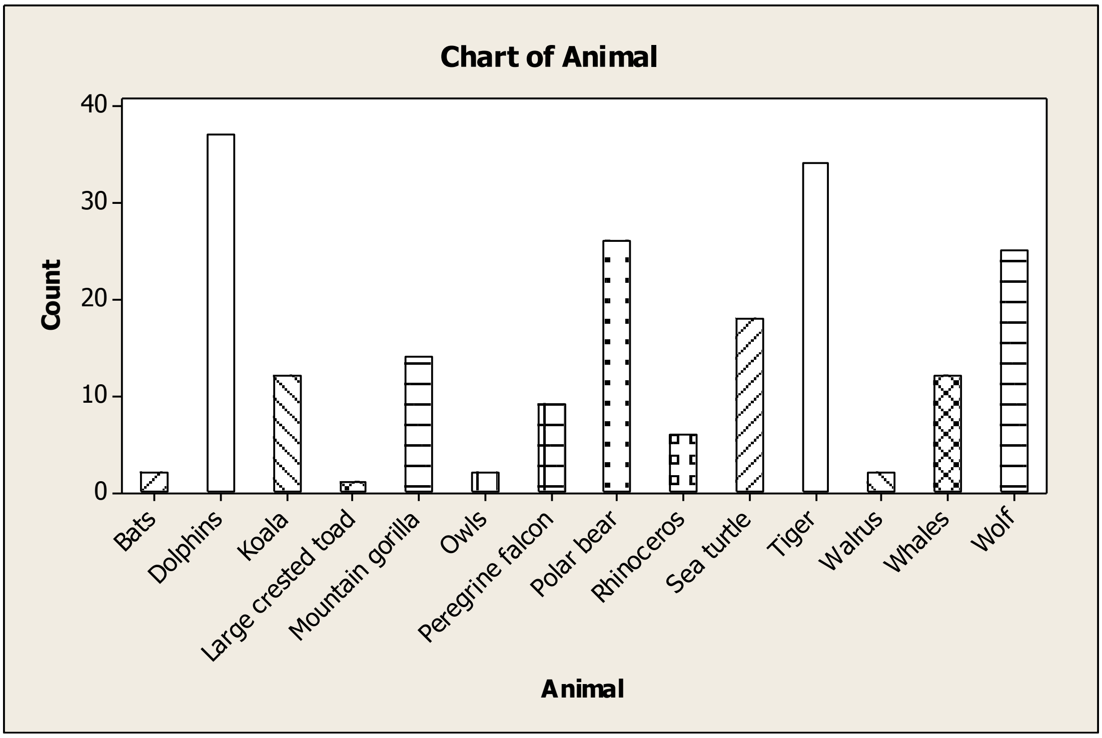
Bar charts are often used to present data in a pictorial form to illustrate the information collected and highlight important points:

- Bar charts a drawn with parallel bars placed vertically or horizontally.
- The width of each bar and the spacing between the bars are kept the same to avoid giving a misleading representation.
- The height of the bar is drawn to scale to represent the amount of the item.
- An individual bar representing a single category must not touch another.

Example

Here are the endangered animals secondary school children in the UK would most like to save. *(Source: CensusAtSchool).*

Animal	Frequency	%
Bats	2	1
Dolphins	37	18.5
Koala	12	6
Large crested toad	1	0.5
Mountain gorilla	14	7
Owls	2	1
Peregrine falcon	9	4.5
Polar bear	26	13
Rhinoceros	6	3
Sea turtle	18	9
Tiger	34	17
Walrus	2	1
Whales	12	6
Wolf	25	12.5



<u>Example</u>

Use the data above to create a bar chart showing the percentage of the animal secondary school children would most like to save.

Graphs for quantitative data

Histograms

A histogram is a graphical representation showing a visual impression of the distribution of data. It is an estimate of the probability distribution of a continuous variable. A histogram consists of tabular frequencies, shown as adjacent rectangles, created over discrete intervals which are called bins, with an area equal to the frequency or the frequency divided by the width of the interval. In the latter case, the total area of the histogram is equal to the number of data. A histogram may also be normalized displaying relative frequencies. It then shows the proportion of cases that fall into each of several categories, with the total area equalling 1.

Equal class widths

When the classes in the table are all the same width, then the height of the bars scale with the frequencies.
Step 1: find the actual range (minimum and maximum)
Step 2: divide the actual range (or a wide range) of the data into classes of equal width
Step 3: count the number of observations in each class, i.e. the frequency for each class
Step 4: draw the histogram

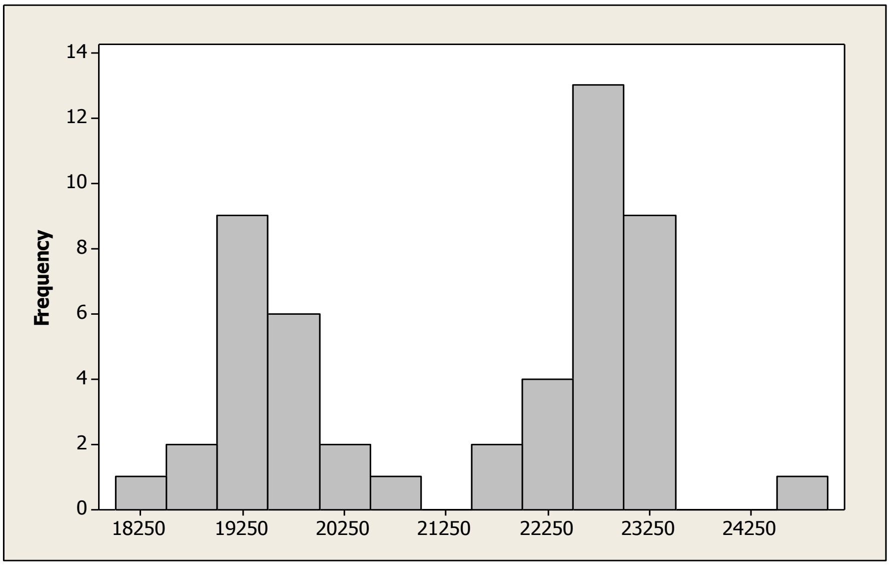
<u>Example</u>

The following table gives the velocities (km/s) of 50 galaxies from cluster A1775. It is thought (hypothesis) that is may be a double cluster, i.e. two galaxy clusters in close proximity.

22922	20210	21911	19225	18792	21993	23059	20785	22781	23303
22192	19462	19057	23017	20186	23292	19408	24909	19866	22891
23121	19673	23261	22796	22355	19807	23432	22625	22744	22426
19111	18933	22417	19595	23408	22809	19619	22738	18499	19130
23220	22647	22718	22779	19026	22513	19740	22682	19179	19404

The range of the data is from 18400 to 24909, so we choose as our classes [18000, 18500), [18500, 19000), … , [24500, 25000). Here, the chosen range is wider than that of the actual range. We use a tally to count the number of observations in each class.

Class	midpoint	tally	frequency
[18000, 18500)	18250	𝍷	1
[18500, 19000)	18750	𝍷𝍷	2
[19000, 19500)	19250	𝍸 𝍷𝍷𝍷𝍷	9
[19500, 20000)	19750	𝍸 𝍷	6
[20000, 20500)	20250	𝍷𝍷	2
[20500, 21000)	20750	𝍷	1
[21000, 21500)	21250		0
[21500, 22000)	21750	𝍷𝍷	2
[22000, 22500)	22250	𝍷𝍷𝍷𝍷	4
[22500, 23000)	22750	𝍸 𝍸 𝍷𝍷𝍷	13
[23000, 23500)	23250	𝍸 𝍷𝍷𝍷𝍷	9
[23500, 24000)	23750		0
[24000, 24500)	24250		0
[24500, 30000)	24750	𝍷	1



Note: there is no one right choice of the classes in a histogram.

- Too few gives a “skyscraper” graph
- Too many produces a “pancake” graph

You need to use your judgement in choosing classes to display the shape.

To interpret you histogram, you need to look for

- Overall pattern
- Deviation from the overall pattern.

Example

Using the data below construct an histogram to represent the distribution.

27	38	22	34	37	41	3	29	14	22
29	31	26	28	50	28	24	47	12	43
13	3	2	29	17	22	46	12	49	27
42	41	49	3	39	13	7	19	29	26
4	47	1	6	50	26	13	14	36	41
28	12	11	22	7	23	28	0	34	22

Class	**midpoint**	**tally**	**frequency**

Unequal class width

If the widths of the groups (class widths) into which the data has been divided are not equal, the heights of the bars must be adjusted so that the areas are in the correct proportions.
Situations where this may be necessary are as follows:

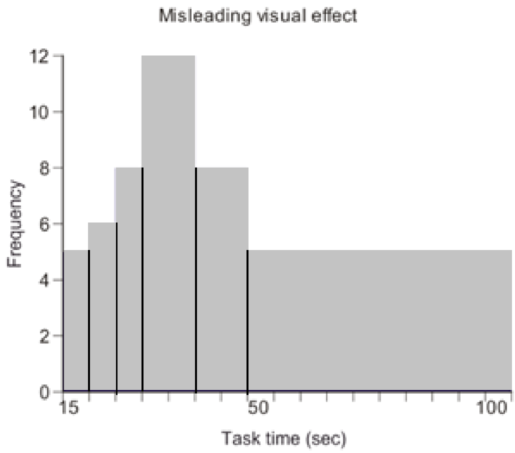
I. Data has already been grouped
II. Naturally occurring ranges (eg. Pre-school, primary school ages, etc.)
III. Extreme values make equal sized groups impractical e.g. noise, signal output.

Example

Consider the frequency distribution below which describes the time taken by 44 individuals to perform a particular task

Task time (s)	Frequency	Frequency density
15-20	5	5/5=1.0
20-25	6	6/5=1.2
25-30	8	8/5=1.6
30-40	12	12/10=1.2
40-50	8	8/10=0.8
50-100	5	5/50=0.1

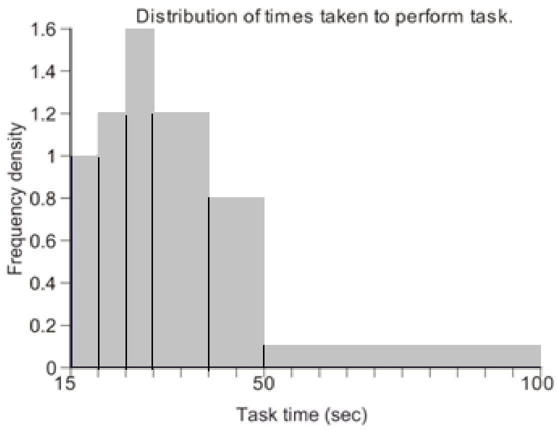
If we plot the frequencies, we get a misleading effect.



To give a fair comparison, we must scale the frequencies to get frequency densities.

$$frequency\ density = \frac{frequency}{class\ width}$$

In the correct histogram, we can clearly see that the majority of times taken lie in the range 20-40s.



Example

Below are the results from testing the lifetime of a component in hours. Construct a histogram using the data.

Life time (hours)	Frequency	Frequency density
0-10	2	
10-20	5	
20-30	10	
30-40	13	
40-50	60	
50-75	25	
75-100	9	
100-200	1	

Stem and leaf displays

For small data set (n<100) a stem and leaf display can also be used. For n>100 use a histogram.

Step 1: separate each observation into a stem (all the digits except for the last one) and a leaf (the last digit)
Step 2: write stems in a vertical column in increasing value order
Step 3: write each leaf in the row to the right of its stem, in increasing value order.
One purpose of a stem and leaf display is to clarify the shape of the distribution. Unlike the histogram, which reflects the real distribution, the stem and leaf graph does not have classes.

Example

The response time of 30 integrated circuits in microseconds has been measured:

4.6	4.0	3.7	4.1	4.1	5.6	4.5	6.0	7.2	3.4
3.4	4.6	3.7	4.2	4.6	4.7	4.1	3.7	3.4	3.3
3.7	4.1	4.5	4.6	4.4	4.8	4.3	4.4	5.1	3.9

In this table, the stems are the first digit, i.e. the integer part, the leaf is the last digit. If we order the data, we can quickly construct the stem and leaf display

3.3	3.4	3.4	3.4	3.7	3.7	3.7	3.9	4.0	4.1
4.1	4.1	4.1	4.2	4.3	4.4	4.4	4.5	4.5	4.6
4.6	4.6	4.6	4.7	4.8	4.8	5.1	5.6	6.0	7.2

3	3 4 4 4 7 7 7 7 9
4	0 1 1 1 1 2 3 4 4 5 5 6 6 6 6 7 8
5	1 6
6	0
7	2

<u>Example</u>

Below is a list of resistances from 20 resistors measured in ohms:

38, 37, 41, 26, 28, 50, 25, 43, 29, 46, 35, 39, 26, 47, 45, 30, 28, 40, 35, 40

Create a stem and leaf display to show the distribution.

Numerical Representation of Data

As well as displaying data graphically we will often wish to summarise it numerically particularly if we wish to compare two or more data set.

Measures of Location

The sample Mean, $\bar{x}$

If the n observations are $x_1, x_2, \ldots, x_n$, then their mean is

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

<u>Example</u>

Using the data for the 30 integrated circuits in the previous example

4.6	4.0	3.7	4.1	4.1	5.6	4.5	6.0	7.2	3.4
3.4	4.6	3.7	4.2	4.6	4.7	4.1	3.7	3.4	3.3
3.7	4.1	4.5	4.6	4.4	4.8	4.3	4.4	5.1	3.9

a. Find the mean of all data

$$\bar{x} = \frac{130.7}{30} = 4.357\ ms\ (3dp)$$

b. Find the mean of all data except for the two highest values

$$\bar{x} = \frac{130.7 - (6.0 + 7.2)}{30 - 2} = \frac{130.7 - (13.2)}{30 - 2} = \frac{117.5}{28} = 4.196\ ms\ (3dp)$$

So from this example, we see that the mean is influenced by the extreme observations. We say that the mean is not a resistant measure.

<u>Example</u>

Here are the resistance (ohm) for 15 resistors, what is the mean of the sample?
38, 41, 29, 31, 28, 50, 25, 43, 29, 46, 27, 42, 49, 29, 26

Median, M

The median is the midpoint of a distribution. To find the median of a set of data:

I. Arrange the n observations in an increasing order
II. If n is odd, M is the centre, the $\frac{n+1}{2}^{th}$ observation in the order list
III. If n is even, M is the mean of the two centre observation in the order list.

<u>Example</u>

Using the data for the 30 integrated circuits in the previous example in its order form

3.3	3.4	3.4	3.4	3.7	3.7	3.7	3.9	4.0	4.1
4.1	4.1	4.1	4.2	4.3	4.4	4.4	4.5	4.5	4.6
4.6	4.6	4.6	4.7	4.8	4.8	5.1	5.6	6.0	7.2

a. Find the median of all data

$$M = \frac{4.2 + 4.3}{2} = 4.25ms$$

b. Find the median of all data except the highest value

$$M = 4.3ms$$

We can see that the median is affected less by extreme values than the mean. We say that it is resistant to outliers or more robust than the mean.

<u>Example</u>

Find the median for the following data:

48 55 52 48 50 49 53 53 51 49 54 55 48

Find the median for the following data:

34 18 18 26 33 23 34 21 33 32 32 35 22 28

Mode

The mode of a set of data is that which occurs with the greatest frequency, i.e. the value that occurs the most often. However, it is rarely used in scientific work as measure of location unless the shape of the distribution is multimodal in which case no single measure of location gives a reasonable description.

<u>Example</u>

What is the mode of the integrated circuit data?

3.3	3.4	3.4	3.4	3.7	3.7	3.7	3.9	4.0	4.1
4.1	4.1	4.1	4.2	4.3	4.4	4.4	4.5	4.5	4.6
4.6	4.6	4.6	4.7	4.8	4.8	5.1	5.6	6.0	7.2

Both 4.1 and 4.6 occur with the greatest frequency (4).

Modal values are 4.1ms and 4.6ms.

<u>Example</u>

What is the mode for the resistor data?

38, 41, 29, 31, 28, 50, 25, 43, 29, 46, 27, 42, 49, 29, 26

Measuring spread

The mean and median provide two different measures of the centre of a distribution. But, a measure of location alone can be misleading.
We also need to measure the spread of a distribution. The spread of a distribution can be measured by the range, quartiles and standard deviation.

Range, $\boldsymbol{R}$

This is the simplest measure of spread. The range of a data set is given by

$$\boldsymbol{range = maximum\ value - minimum\ value}$$

This measure describes the extremes only, so it may not represent the main body of the data.

Example

Find the range for the following data set:
99.4 100 58.4 42.6 63.5 24.8 53.3 85.1 28.4 61.1

Maximum=100 Minimum=24.8

$$range = 100 - 24.8 = 75.2$$

Example

Find the range for the following data set:
9.89 5.02 2.46 5.40 0.59 3.80 6.62 2.62 7.46 4.73

The inter-quartile range (IQR)

The quartiles divide the ordered data into four groups, where each one contains the same number of observations. For this measure, we need to locate the quartiles. These are found in a similar way to the median. In fact the median is the second quartile.

To calculate the quartiles

Step 1: arrange the observations in increasing order and locate the median M in the ordered list of observations.
Step 2: Q_1 is the median of the observations whose position in the ordered list is to the left of the location of the overall median, i.e. the $\frac{n+1}{4}^{th}$ value.
Step 3: Q_3 is the median of the observations whose position in the ordered list is to the right of the location of the overall median, i.e. the $\frac{3(n+1)}{4}^{th}$ value.
Step 4: the inter-quartile range (IQR) is

$$IQR = Q_3 - Q_1$$

Example

Find the interquartile range of the following numbers:
5.45 7.87 8.97 1.34 0.35 6.22 6.55 2.84 2.07 9.89 2.46 0.59 6.62 7.46 5.02 5.40 3.80 2.62 4.73

There are 19 observations, the median is 5.02
First Quartile:
$Q_1 = \frac{n+1}{4} th$ observation, $\frac{19+1}{4} = \frac{20}{4} = 5$, 5th observation is 2.46.

$$Q_1 = 2.46$$

Last Quartile:
$Q_3 = \frac{3(n+1)}{4} th$ observation, $\frac{3(19+1)}{4} = \frac{3\times 20}{4} = 3 \times 5 = 15$, 15th observation is 6.62.

$$Q_3 = 6.62$$

$$IQR = Q_3 - Q_1 = 6.62 - 2.46 = 4.16$$

Example

Find the interquartile range of the following numbers:
34 18 18 26 33 23 34 21 33 32 35 22 28 18 31 28 30 28 26 20 32

Five number summary

We can use the five numbers:

$$Minimum, Q_1, M, Q_3, Maximum$$

to describe the distribution of a variable.

Example

Here is the number of runs a cricketer made in his last 19 innings:
0, 1, 0, 48, 105, 116, 49, 11, 1, 0, 4, 13, 41, 129, 65, 39, 24, 3, 10

Find the mean and the five number summary for the data.

Mean: $\frac{659}{19} \approx 34.68\, runs$

Order the data:
0, 0, 0, 1, 1, 3, 4, 10, 11, 13, 24, 39, 41, 48, 49, 65, 105, 116, 129

Median: 19 observations, $\frac{19+1}{2} = \frac{20}{2} = 10$, 10$^{\text{th}}$ observation is the median, median= 13 runs.

Q_1: $\frac{19+1}{4} = \frac{20}{4} = 5$, 5$^{\text{th}}$ observation is Q_1, $Q_1 = 1$ runs.

Q_3: $\frac{3(19+1)}{4} = \frac{3\times 20}{4} = 15$, 15$^{\text{th}}$ observation is Q_3, $Q_3 = 49$ runs.

Minimum = 0

Maximum = 129

Five number summary: 0, 1, 13, 49, 129

Example

Here are the runs totals for another cricketer in his last 19 innings:
27, 109, 125, 38, 52, 71, 11, 63, 0, 4, 106, 54, 6, 104, 29, 121, 1, 46, 5

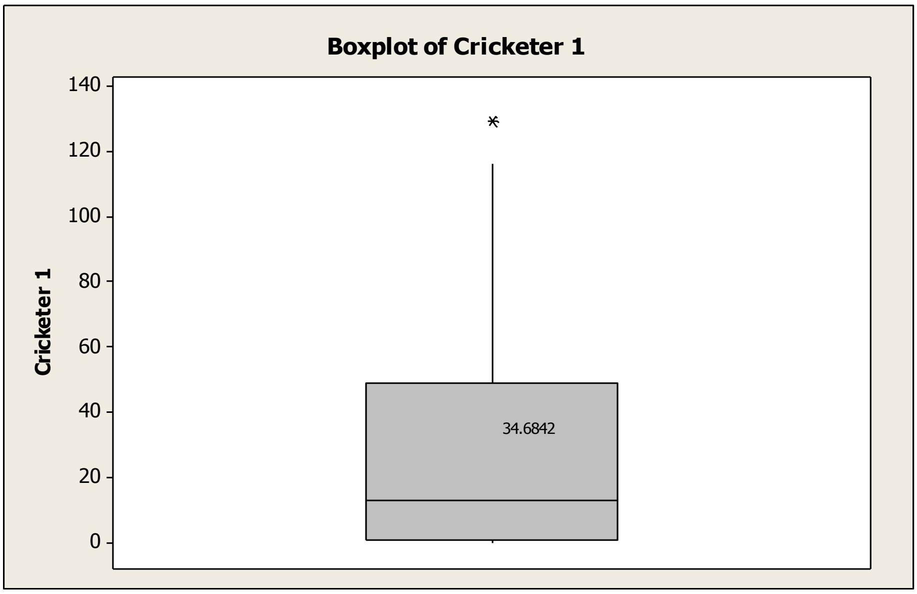
Find the mean and the five number summary for this data.

(source: http://www.espncricinfo.com/)

These five numbers lead to a new graph, the **Boxplot**.

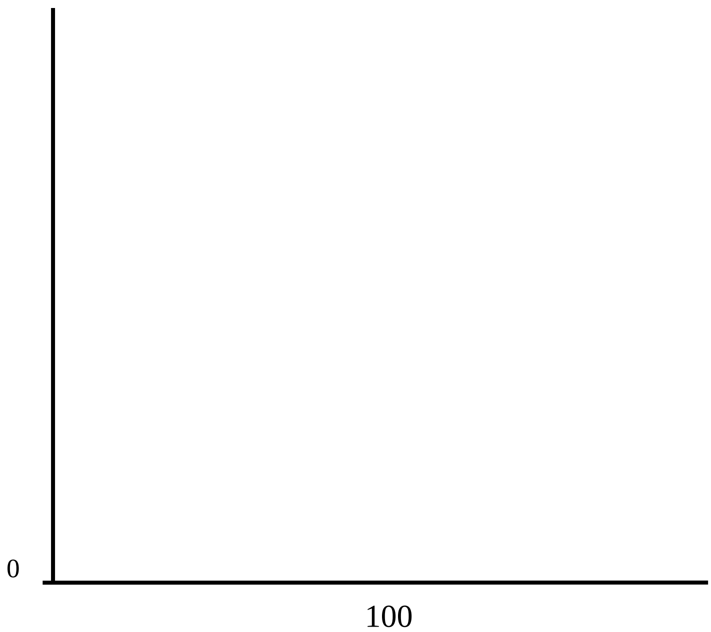
<u>Example</u>

Using the five number summary from the previous example, produce a boxplot for the data.



<u>Example</u>

Using the five summary from the previous example to produce a boxplot for the data.



The sample standard deviation, s, and the variance s^2

The standard deviation measures spread by calculating bow far each observation is from the mean. The sample standard deviation is given by:

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}^2)} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$$

where the number $n - 1$ is called the degrees of freedom of the standard deviation. The variance of the sample is s^2.

- $s = 0$ only when there is no spread, i.e. all of the observations are the same
- s is not resistant, because it is sensitive to outliers in the same way as the mean.

<u>Example</u>

Find the mean, standard deviation and variance of the following data set:

383.4 385.7 386.1 383.9 384.1 384.6 385.2

Note: Keep full figure accuracy during the evaluation of s, rounding values before the end of the calculation can cause wildly different results.

	x_i	x_i^2
	383.4	146995.56
	385.7	148764.49
	386.1	149073.21
	383.9	147379.21
	384.1	147532.81
	384.6	147917.16
	385.2	148379.04
Totals	2693	1036041.48

$$\bar{x} = \frac{2693}{7} = 384.7$$

$$s^2 = \frac{1036041.48 - 7 \times \left(\frac{2693}{7}\right)^2}{7-1} = \frac{1036041.48 - 7 \times \left(\frac{2693}{7}\right)^2}{6}$$

$$s^2 = 0.9848\ (4pd)$$

$$s = \sqrt{s^2} = 0.9924\ (4dp)$$

Example

Find the mean, standard deviation and variance of the following data set:
5.54 6.22 7.87 6.55 5.97 9.88 1.34 2.84 0.35 2.07

	x_i	x_i^2
	5.54	
	6.22	
	7.87	
	6.55	
	5.97	
	9.88	
	1.34	
	2.84	
	0.35	
	2.07	
Totals		

Key points

- There are two types of data, qualitative and quantitative.
- We use bar charts for qualitative data.
- Histograms are used for quantitative data and can have equal and unequal class widths.
- If a histogram has unequal class widths, the frequency density is plotted rather than the frequency. $frequency\ density = \frac{frequency}{class\ width}$.
- We use stem and leaf displays for small data sets.
- The mean of a data set is given by $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.
- The median of the data is the midpoint of the distribution. If n is odd, M is the $\frac{n+1}{2}th$ observation. If n is even it is the mean of the middle two observation. The list has to be ordered.
- Mode is the value that occurs with the greatest frequency in the data set.
- The range of the data is maximum value – minimum value.
- The inter-quartile range (IQR) is given by the last quartile – first quartile.
- Five number summery is given by the minimum, first quartile, median, last quartile, maximum. These can be used to create a boxplot.
- The standard deviation of a sample is given by $s = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)}$.
- The variance is given by s^2.

Probability

Definitions

A random phenomenon has an outcome that cannot be predicted but that nonetheless has a regular distribution in very many repetitions. For example, if we flip a coin, we cannot predict whether we will get a head or tall. But, if we flip the coin thousands of times, we will find that the number of heads is almost the same as the number of tails.

The sample space, *S*, is the set of all possible outcomes of the random phenomenon. Sets of outcomes are called events.
Probability is the frequency with which a particular random event occurs, provided that the population is sufficiently large. We can use long-run frequencies to predict the probability of an event occurring (population $\rightarrow$sample).
The probability that the event $A \in S$ will happen is denoted by $P(A)$. This is a measure of how likely it is that *A* will occur.
An event is an outcome from a test, e.g. rolling an even number on a dice, drawing an ace from a pack of cards.
Conditional probability measures the probability an event given that (by assumption, presumption, assertion or evidence) another event has already occurred.

Probability Rules

Value

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- If $P(A) = 0$, *A* cannot happen.
- If $P(A) = 1$, *A* must happen.

One way of calculating probabilities is to take what has happened in the past as an indication of what will happen in the future. Thus, we use the relative frequency (or proportion of times an event happened in the past) as an estimate of the required probability (or how likely an event is to happen in the future).
Another way of calculating a probability is

$$probability\ of\ an\ event\ happing = \frac{Number\ of\ ways\ it\ can\ happen}{Total\ number\ of\ outcomes}$$

<u>Example</u>

In the past, 100,000 electrical components have been produced by a particular company. Of these, 3,240 have been defective. Estimate the probability that a component taken at random from the company's production line will be defective.

$$\frac{number\ of\ defective}{total\ number\ manufactured} = \frac{3240}{100000} = 0.0324$$

<u>Example</u>

A company manufactures electrical components. In the past, 125,000 of these components have been manufactured and 104,000 of these components faultless. Estimate the probability that a component taken at random from the company's production line will be faultless.

Mutual exclusivity

Two or more events are said to be mutually exclusive if they cannot occur at the same time.

$$P(A\ and\ B) = 0\ or\ P(A \cap B) = 0$$

What is the probability of drawing a card from a standard pack of playing cards which an ace and a king.

__**Example**__

$$P(Ace\ and\ King) = 0$$

A playing card cannot be both a king and an ace.

Not Mutually Exclusive

Events that are not mutually exclusive CAN happen at the same time. Examples of these events are:

- Rolling a 2 and an even number with a single dice
- Drawing a diamond and an 8 from a deck of cards

Independence

We say that *A* and *B* are independent if knowing that event *A* has occurred has no effect on the probability that event *B* will occur and vice versa. So for independent events, the conditional probability that event *B* occurs given that event *A* has already occurred is

$$P(B|A) = P(B).$$

__**Example**__

A card is drawn from a standard pack of playing cards and then replaced. What is the probability of now drawing an ace?
As the card we previously drawn have been replaced, it has no effect on the probability of the next card to be drawn.

$$P(Ace) = \frac{4}{52} = \frac{1}{13}$$

Dependence

We say that *A* and *B* are dependent if knowing that event *A* has occurred it effects the probability that event *B* will occur and vice versa.

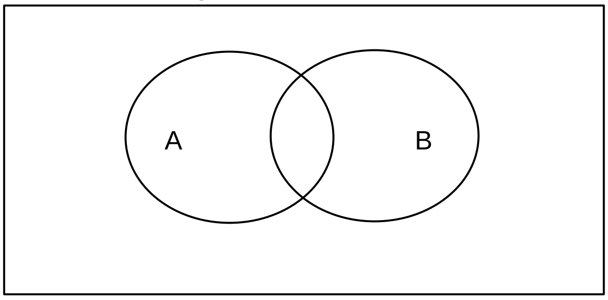
$$P(A|B) = \frac{P(A \cap B)}{P(B)},\ \ \mathrm{P(B|A)} = \frac{P(A \cap B)}{P(A)}$$

__**Example**__

A six-sided die is thrown. What is the probability that the number is prime, given that it is odd?
$P(odd) = \frac{3}{6} = \frac{1}{2}$ and of these odd numbers only two are prime.

$$P(prime|odd) = \frac{P(prime\ and\ odd)}{P(odd)} = \frac{\left({}^{1}\!/_{2} \times {}^{2}\!/_{3}\right)}{{}^{1}\!/_{2}} = \frac{{}^{1}\!/_{3}}{{}^{1}\!/_{2}} = \frac{2}{3}$$

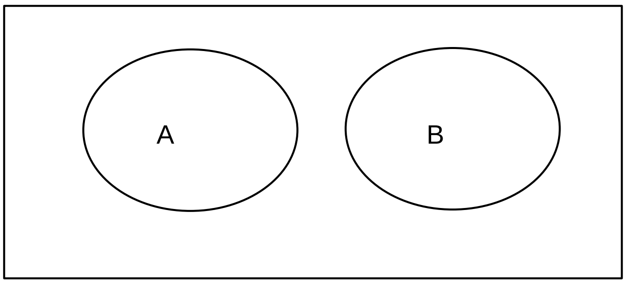
The addition rule and Venn diagrams



Consider the event *A* and *B*. the areas enclosed in the Venn diagram above represent the probabilities that *A, B,* both or neither occur. If *A* and *B* are not mutually exclusive than an event can occur which satisfies both criteria.
The complement is the area represented by the areas that we are not interested in.
The probability that **either** *A* **or** *B* occurs is

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

If events *A, B* are mutually exclusive, then our Venn diagram would look as follows:



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

The addition rule simplifies to

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

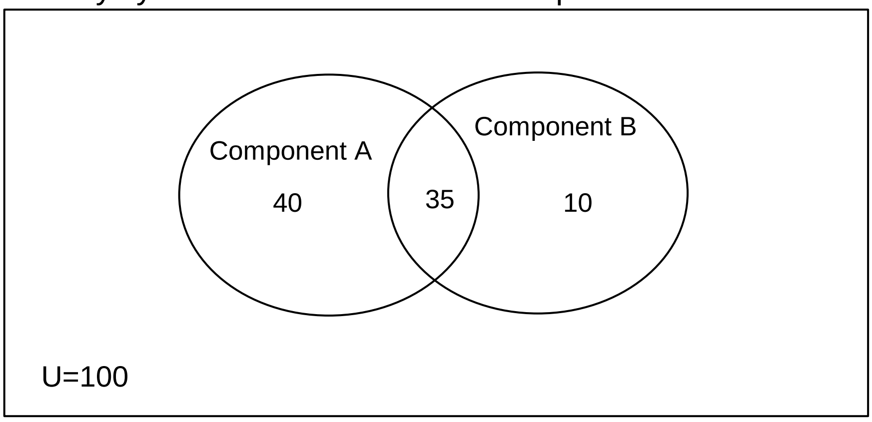
The addition rule can be applied to any number of criteria not just two. For any individual event *A*

$$P(A\ does\ NOT\ occur) = 1 - P(A\ does\ occur)$$

<u>Example</u>

A subsystem is made up of two components, A and B. 100 of these systems are tested and it is found that in 75 of the systems component A fails, 45 of the systems component B fails and in 35 systems both component A and B fail.

a. What is the number of systems in which either component A or component B fails?
b. In how many systems will neither of the components fail?

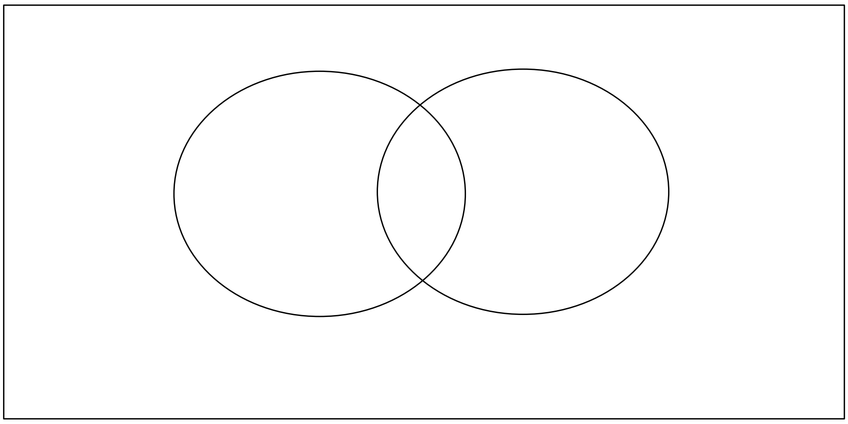


We know that in 35 systems both components failed, so this information is written in the area where the two circles intersect.
Now complete the remaining numbers for the other two sets. In 75 systems component A failed, 35 of those component B also failed, so we have 75-35=40. There are 40 systems where component A failed only. In 45 systems component B failed, 35 of those component A also failed, so we have 45-35=10. There are 10 systems where component B failed only.
Now answer the questions.
The number of systems in which either component A or component B fails is the union of the two sets. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 75 + 45 - 35 = 120 - 35 = 85$. There are 85 systems in which component A or component B failed.
The number of systems in which none of the components failed is the complement of $A \cup B$. We know that there are 100 systems so the complement is found by subtracting the number of systems where either component A or component B failed from the total number of systems. So we have 100-85=15. There are 15 systems in which neither component A nor component B failed.

Example

A system is made up of two components, A and B. 150 of these systems are tested, in 85 systems component A failed, in 70 systems component B failed and in 50 systems both components failed.

a. In how many systems did ONLY component A fail?
b. In how many systems did ONLY component B fail?
c. In how many systems did either component A or component B fail?
d. In how many systems did neither component A or B fail?



Example

In the example previously, we estimated that the probability that a randomly chosen electrical component would be defective is 0.0324. What is the probability that such a component is satisfactory?

$$P(component\ is\ satisfactory) = 1 - P(component\ is\ defective) = 1 - 0.0324 = 0.9676$$

Example

The probability that a randomly chosen component is satisfactory is 0.974. What is the probability that a component is defective?

The multiplication rule

The probability that events *A* and *B* both occur is given by

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

where $P(A|B)$ is the probability of *A* occurring given that *B* has occurred. Hence for independent event for which $P(A|B) = P(A)$ this rule simplifies to

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

<u>Example</u>

The following table gives the lifespans (L) of 1500 components

Lifespan (hours)	No of components
$L \geq 1000$	210
$900 \leq L < 1000$	820
$800 \leq L < 900$	240
$700 \leq L < 800$	200
$L < 700$	30

a. What is the probability that a component which is still working after 800 hours will last for at least 900 hours?

$P(L > 800) = \frac{1270}{1500}$ and $P(L > 900) = \frac{1030}{1500}$

$$P = \frac{{}^{1030}\!/_{1500}}{{}^{1270}\!/_{1500}} = 0.8110$$

b. What is the probability that a component which is still working after 900 hours will continue to last for at least 1000 hours?

$P(L > 900) = \frac{1030}{1500}$ and $P(L > 1000) = \frac{210}{1500}$

$$P = \frac{{}^{210}\!/_{1500}}{{}^{1030}\!/_{1500}} = 0.2039$$

<u>Example</u>

A system is known to suffer from two faults: fault *A* and fault *B*. The probability that the system develops fault *A* at some stage during its lifetime is 0.2. The corresponding probability for fault *B* is 0.15. The probability the system develops both faults during its lifetime is 0.03.

a. Are the events ‘the system develops fault *A’* and ‘the system develops fault *B*’ mutually exclusive or independent?
b. What is the probability that the system develops
   i. Fault *A* only during its lifetime?
   ii. Fault *B* only during its lifetime?
   iii. Fault *A* or fault *B* or both during its lifetime?

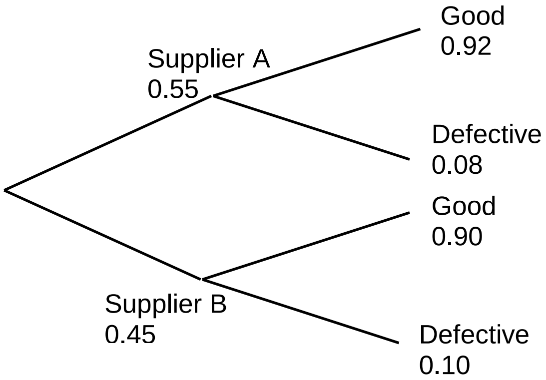
Tree diagrams

A tree diagram is a pictorial aid to the evaluation of conditional probabilities.

Example

A company which manufactures computers receives 55% of its parts from supplier A and 45% of its parts from supplier B. it is known that 92% of parts from supplier A are good, while 90% of parts from supplier B are good. What is the probability that a randomly selected part is

a. From supplier B and good?
b. From supplier A and defective?
c. Good?
d. Given that a part selected is good, what is the probability that it came from B?



$$P(Supplier\ B\ and\ good) = 0.45 \times 0.90 = 0.405$$

$$P(Supplier\ A\ and\ defective) = 0.55 \times 0.08 = 0.044$$

$$P(component\ is\ good) = P(supplier\ A\ and\ good) + P(supplier\ B\ amd\ good)$$

$$= 0.55 \times 0.92 + 0.45 \times 0.90 = 0.506 + 0.405 = 0.911$$

$$P(a\ component\ is\ good\ and\ has\ come\ from\ supplier\ B)$$

$$P(supplier\ B|good) = \frac{P(supplierB \cap good)}{P(good)} = \frac{0.405}{0.911} = 0.445\ (3dp)$$

Example

A system is made up of two components, A and B. The probability that component A fails is 0.15, the probability that component B survives is 0.92. The probability of each component failing is independent of the other. What is the probability that

a. The system survives?
b. If the system fails that component A has failed.

Key Points

- The probability of an event is always between 0 and 1. $0 \leq P(A) \leq 1$.
- The probability of all possible outcomes in a sample space *S* is $P(s) = 1$.
- If $P(A) = 0, A$ cannot happen.
- If $P(A) = 1, A$ must happen.
- If events are mutually exclusive then they cannot occur at the same time, $P(A \cap B) = 0$.
- If events are independent, then if one event has already occurred this has no effect on the probability that the other event will occur. $P(B|A) = P(B)$.
- The addition rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- If events *A* and *B* are mutually exclusive, the rule simplifies to $P(A \cup B) = P(A) + P(B)$.
- The addition rule can be applied to any number of criteria: $P(A\ does\ NOT\ occur) = 1 - P(A\ does\ occur)$.
- The multiplication rule: $P(A \cap B) = P(A) \times P(B|A) = P(B) \times P(A|B)$. For independent events the rule simplifies to $P(A \cap B) = P(A) \times P(B)$.
- Tree diagrams can aid the evaluation of conditional probabilities.

Probability distributions

A quantitative random variable (r.v) is a variable taking numerical values determined by the outcome of a random phenomenon. Random variables can be either discrete or continuous.

Definitions

Discrete random variable is a random variable which can only take a countable number of values.

Continuous random variable is a random variable where the data can take infinitely many values.

Example

X is the number of time a price of equipment will break down each month.
Y is the lifetime of a battery.

Are X and Y discrete or continuous random variables?

X is a discrete random variable as the variable can only take a countable number of values.

Y is a continuous random variable as the variable can take infinitely many values.

Example

X is the resistance of an electrical component.
Y is the number of times a queen is drawn from a standard deck of playing cards.
Are X and Y discrete or continuous random variables?

Discrete distributions

For any discrete probability distribution with probabilities p_1, p_2, p_3, … p_n

$$\sum_{i=1}^{n} p_i = 1$$

with each probability p_i satisfying $0 \leq p_i \leq 1$.

This distribution can be shown graphically by a bar chart. The height of each bar then represents the probability of that particular value of the random variable occurring.

Example

Let X be the number of faults on a manufactured item at random. The probability distribution for X is given in the following table.

No. of faults	0	1	2	3	4
Probability	0.05	0.26	0.19	0.3	0.2

Find the following probabilities,

a) $P(X \leq 2)$
b) $P(X \geq 3)$
c) $P(X \leq 4)$

a) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X + 2) = 0.05 + 0.26 + 0.19 = 0.5$

b)$P(X \geq 3) = P(X = 3) + P(X = 4) = 0.3 + 0.2 = 0.5$

c)$P(X < 4) = 1 - P(X = 4) = 1 - 0.2 = 0.8$

Example

X has probability distribution $f(x) = \frac{x+1}{15}$, where $x = 0, 1, 2, 3\ ,4$.

X	0	1	2	3	4
Probability	1/15	2/15	3/15	4/15	5/15

Find the following probabilities,

a) $P(X < 2)$
b) $P(1 \leq X \leq 3)$
c) $P(X = 1 \cup X = 4)$
d) $P(X > 2)$

Mean of a discrete probability distribution

The mean or Expected value of a discrete probability distribution is given by

$$E(X) = \mu = \sum_{i=1}^{n} x_i p_i.$$

Example

Let X be the number of faults on a manufactured item at random. The probability distribution for X is given in the following table.

No. of faults	0	1	2	3	4
Probability	0.05	0.26	0.19	0.3	0.2

Find the expected value.

$$\mu = 0 \times 0.05 + 1 \times 0.26 + 2 \times 0.19 + 3 \times 0.3 + 4 \times 0.2$$

$$= 0 + 0.26 + 0.38 + 0.9 + 0.8$$

$$= 2.34 \, faults$$

Example

X has probability distribution $f(x) = \frac{x+1}{15}$, where $x = 0,\ 1,\ 2,\ 3\ ,4$

X	0	1	2	3	4
Probability	1/15	2/15	3/15	4/15	5/15

Find the expected value for the table above.

Standard deviation and Variance of a discrete probability distribution

The variance of a discrete probability distribution is given by

$$s^2 = Var(X) = \sigma^2 = \sum_{i=0}^{n} x_i^2 p_i - \mu^2$$

The variance is the value in the probability distribution squared minus the Expected value or mean squared.

The standard deviation of a discrete probability distribution is given by $s = \sqrt{s^2}$.

Example

Let X be the number of faults on a manufactured item at random. The probability distribution for X is given in the following table.

No. of faults	0	1	2	3	4
Probability	0.05	0.26	0.19	0.3	0.2

Find the standard deviation and the variance for the table above.

$$s^2 = 0^2 \times 0.05 + 1^2 \times 0.26 + 2^2 \times 0.19 + 3^2 \times 0.3 + 4^2 \times 0.2 - 2.34^2$$

$$= 0 + 0.26 + 0.76 + 2.7 + 3.2 - 5.4756 = 6.92 - 5.4756 = 1.4444$$

$$s = \sqrt{1.4444} = 1.202$$

Example

X has probability distribution $f(x) = \frac{x+1}{15}$, where $x = 0,\ 1,\ 2,\ 3\ ,4$

X	0	1	2	3	4
Probability	1/15	2/15	3/15	4/15	5/15

Find the standard deviation and the variance for the table above.

In the next sections we shall consider a number of probability distributions which are important in electrical and communication engineering. These are:

- Discrete probability distributions
  - The binomial distribution
  - The Poisson distribution
- Continuous probability distributions
  - The exponential distribution
  - The normal (Gaussian) distribution
  - The lognormal distribution
  - The Weibull distribution

Key points

- A quantitative random variable is a variable taking numerical values determined by the outcome of a random phenomenon.
- Random variables can be either discrete or continuous.
- Discrete random variable is a random variable which can only take a countable number of values.
- Continuous random variable is a random variable where the data can take infinitely many values.
- For any discrete probability distribution with probabilities p_1, p_2, p_3,…p_n, $\sum_{i=1}^{n} p_i = 1$.
- The Mean or Expected value for a discrete probability distribution is $\sum_{i=0}^{n} x_i p_i$.
- The variance of a discrete probability distribution is given by $\sum_{i=0}^{n} x_i^2 p_i - \mu^2$
- A continuous r.v. can take all values in a certain interval.

Combinations, Permutations, the Binomial distribution and the Poisson distribution

Factorial Notation

$$n! = n \times (n-1) \times (n-2) \times \ldots \times 1, \qquad \text{where n is a positive integer.}$$

By definition $0! = 1$.

Permutations

The number of ways of choosing *r* items from a total of *n* items $(r \leq n)$ when the order is important is known, mathematically, as the number of permutations of *r* items from *n* and is defined by

$$n^r$$

Example

In a combination lock, there are 10 numbers to choose from and you need to choose three of them, what is the number of permutations?

$$10^3 = 1000 \text{ permutations}$$

Example

A pin number is a random selection of 4 numbers from a possible 10. What is the number of permutations?

Permutations without repetition

In this case we are not allowed to have the same item repeated. This means that the number of available choices is reduced each time. This is defined by

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

Example

What is the number of permutations for drawing 5 cards from a standard deck of playing cards without replacement?

$$\frac{52!}{(52-5)!} = \frac{52!}{47!} = 52 \times 51 \times 50 \times 49 \times 48 = 311875200$$

Example

A password consists of 3 different letters of the alphabet, where each letter is only used once. How many different possible passwords are there?

Your calculator should have this function built in.

Combinations

The number of ways of choosing *r* items from a total of *n* items $(r \leq n)$ when the order is unimportant is known, mathematically, as the number of combinations of *r* items from *n* and is defined by

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

This is often read as 'n choose r'. Your calculator should have this function built in.

Example

In the National Lottery six balls are chosen from a total of 49. How many ways are there of choosing six balls from 49?

$$^{49}C_6 = \frac{49!}{6!\,(49-6)!} = \frac{49!}{6!\,43!} = \frac{49 \times 48 \times 47 \times 46 \times 45 \times 44}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 13983816$$

Example

How many different committees of 5 people can be chosen from 10 people?

The Binomial distribution

The Binomial distribution is an example of a discrete probability distribution. It is the probability distribution of the number of successes in a sequence of *n* independent yes/no experiments, each of which yields success with probability *p*.

When a trail only has two possible outcomes it is termed a Bernoulli trail. An example of this would be flipping a coin, assuming that the coin cannot land on its edge; there are only two possible outcomes.

The number of trails *n* and the probability of success *p* are termed the parameters of the binomial distribution. If the r.v. *X* has the binomial distribution with parameter *n* and *p* we write

$$X \sim B(n, p).$$

The binomial distribution gives the probability of obtaining *x* successes from *n* Bernoulli trails.

If we let *X* be the random variable giving the number of success in *n* Bernoulli trails,

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

where

p is the probability of a success on any individual trail;

$^{n}C_x$ is the number of ways of getting *x* successes in *n* trails

$p^x(1-p)^{n-x}$ is the probability associated with each of the $^{n}C_x$ ways.

The mean and the variance of the Binomial random variable are,

$$\mu = \sum_{x=0}^{n} xP(x) = np, \qquad \sigma^2 = \sum_{x=0}^{n} P(x)(x-\mu)^2 = np(1-p)$$

As usual the standard deviation is the square root of the variance.

There are a number of assumptions associated with the binomial distribution

1. The total number of Bernoulli trails, *n*, is fixed in advance of the experiment.
2. The probability of success in any individual trial, *p*, is fixed across all trails.
3. The individual trails are independent of one another.

Example

A fair coin is flipped 10 times. What is the probability of obtaining 7 heads? What is the mean of the number of heads obtained in 10 flips? What is the variance?

$$B(7,0.5) = \frac{10!}{7!\,(10-7)!}0.5^7(1-0.5)^{10-7}$$

$$= \frac{10!}{7!\,3!}0.5^7(0.5)^3 = \frac{10\times 9\times 8}{3\times 2\times 1}0.5^7(0.5)^3 = 120(0.5)^7(0.5)^3 = 0.117$$

$$\mu = 10\times 0.5 = 5$$

$$\sigma^2 = 10\times 0.5(1-0.5) = 5\times 0.5 = 2.5$$

Example

Past data tells us that 10% of the components produced by a particular company are faulty. A sample of size 20 is drawn from a large batch of components for testing. The entire batch will be rejected if more than two of the components are found to be faulty. What is the probability that two components will be faulty? What is mean number of faulty components? What is the variance?

The Poisson distribution

The Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time and/or space if these events occur with a known average rate and independently of the time since the last event.

Definition

A discrete random variable *X* is said to have a Poisson distribution with the parameter λ>0 if, for, *k* = 0, 1, 2,…

$$P(X=x) = \frac{\lambda^k e^{-\lambda}}{k!}$$

We write this as

$$X \sim Po(\lambda)$$

X follows a Poisson distribution with parameter λ
For the Poisson distribution, the mean and variance is

$$\mu = \sigma^2 = \lambda$$

Example

Flaws occur on wire randomly and independently at a rate of six every 10 meters. What is the probability that a randomly selected piece of wire length 10m will have

a) no flaws?
b) no more than two flaws?

The average number of flaws per 10 meters is $6 = \lambda$

$$P(X = 0) = \frac{6^0 e^{-6}}{0!} = e^{-6} = 0.00248$$

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

$$= \frac{6^0 e^{-6}}{0!} + \frac{6^1 e^{-6}}{1!} + \frac{6^2 e^{-6}}{2!} = e^{-6} + 6e^{-6} + 18e^{-6}$$

$$= 0.0620$$

Example

Breakdowns in a communication system occur at random and independently with an average of 19 per year. What is the probability that in a particular month there are:

a) no breakdowns?
b) Less than three breakdowns?
c) At least four breakdowns?

Key points

- Permutation is the number of ways of choosing items when the order is important.
- This is defined mathematically as n^r.
- Permutations without repetition is defined by $\frac{n!}{(n-r)!}$.
- Combinations is the number of ways of choosing items when the order is not important.
- This is defined mathematically as ${}^nC_r = \frac{n!}{r!(n-r)!}$
- The binomial distributions is used when there are only two outcomes of an event.
- If a random variable follows a binomial distribution, we write this as $X \sim B(n, p)$.
- To evaluate the probability of a binomial distribution we use $P(X = x) = {}^nC_x p^x (1-p)^{n-x}$.
- The mean of a binomial distribution is $\mu = np$.
- The variance of a binomial distribution is $\sigma^2 = np(1-p)$.
- The Poisson distribution expresses the probability of a given number of events in a fixed time and/or space interval.
- If a random variable follows a Poisson distribution, we write this as $X \sim Po(\lambda)$.
- To evaluate the probability of a Poisson distribution we use $P(X = x) = \frac{\lambda^k e^{-\lambda}}{k!}$.
- The mean of a Poisson distribution is $\mu = \lambda$.
- The variance of a Poisson distribution is $\sigma^2 = \lambda$

Continuous distributions

A continuous r.v. can take all values in a certain interval. A density curve describes the probability distribution for a continuous r.v.

Probability distributions for continuous variables are scaled so that the total area under the density curve is one unit. This is equivalent to the probabilities for a discrete distribution adding up to 1.

With discrete variables we often interested in the probability that our variable equals a particular value. For example, the probability that a manufactured item has exactly two faults. Such probabilities are known as point probabilities. With continuous variables, point probabilities are of no interest. This is because they are always effectively zero. Instead, we turn our attentions to interval probabilities.

PDF and CDF of a continuous random variable

Definitions

We use the probability density function (PDF) or the cumulative distribution function (CDF) to describe the distribution of a continuous random variable.

Generally, any function $f(x)$can serve as a pdf provided that $\int_{-\infty}^{\infty} f(x)dx = 1$ and $f(x) \geq 0$ for $-\infty < x < \infty$.

The cumulative distribution function is defined by

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

Where $F(x) = \int_{-\infty}^{\infty} f(x)dx$.

Sometimes it is more convenient to work with $F(x)$ then with $f(x)$.

In general, if the random variable *X* has pdf $f(x)$, then the probability of *X* is

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

<u>**Example**</u>

Suppose a random variable Y has a density curve with height 1 over the interval from 0 to 1. Find the following probabilities.

a) $P(Y \leq 0.5)$
b) $P(0.3 \leq Y \leq 0.7)$
c) $P(Y \geq 0.8)$
d) $P(Y \leq 0.5 \; or \; Y \geq 0.8)$

$$f(y) = 1, \qquad 0 \leq y \leq 1$$

$$P(Y \leq 0.5) = \int_0^{0.5} 1dy = [y]_0^{0.5} = 0.5 - 0 = 0.5$$

$$P(0.3 \leq Y \leq 0.7) = \int_{0.3}^{0.7} 1 \, dy = [y]_{0.3}^{0.7} = 0.7 - 0.3 = 0.4$$

$$P(Y \geq 0.8) = \int_{0.8}^{1} 1 \, dy = [y]_{0.8}^{1} = 1 - 0.8 = 0.2$$

$$P(Y \leq 0.5 \; or \; Y \geq 0.8) = P(Y \leq 0.5) + P(Y \geq 0.8) = 0.5 + 0.2 = 0.7$$

Example

Suppose a random variable has the pdf $f(x) = 3x^2, 0 \leq x \leq 1$. Find the following probabilities

a) the probability that X is no more then 0.5
b) the probability that X is between 0.25 and 0.5
c) the probability that X is greater than or equal to 0.75

Expected value and variance of a continuous random variable

The expected value of a random variable is given by

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

The variance of a random variable is given by

$$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Example

Suppose a random variable Y has a density curve with height 1 over the interval from 0 to 1. Find the expected value and variance for this distribution.

$$E(Y) = \int_0^1 y \times 1\, dy = \int_0^1 ydy = \left[\frac{y^2}{2}\right]_0^1 = \frac{1^2}{2} - \frac{0^2}{2} = 0.5$$

$$V(Y) = \int_0^1 y^2 \times 1 \times dy - 0.5^2 = \int_0^1 y^2 dy - 0.25 = \left[\frac{y^3}{3}\right]_0^1 - 0.25$$

$$= \frac{1}{3} - 0.25 = 0.0833$$

Example

Suppose a random variable has the pdf $f(x) = 3x^2$ over the interval from 0 to 1. Find the expected value and variance for this distribution.

Given some data on the continuous r.v., the fundamental questions are:

1. Which distribution generally fits the data best?
2. What are the implications?

These implications include evaluating probabilities and interpreting the estimates of the parameters and associated reliability characteristics.

Exponential distribution

The exponential distribution is the probability distribution that describes the times between events in a Poisson process. That is a process in which events occur continuously and independently at a constant average rate.
We write this as

$$X \sim \mathrm{Exp}(\lambda)$$

The exponential probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

It has only one parameter, λ.

The cumulative distribution function gives the probability that the gap between 2 successive events, X, is less than some given value, c,

$$F(c) = P(X \leq c) = 1 - e^{-\lambda c}$$

In general, if an event is occurring randomly in time or space, and it occurs at a constant rate, λ,

- the number of events in some time interval follows a Poisson distribution;
- the distance/time between the events follows an exponential distribution.

<u>Example</u>

Let T be the random variable of the lifetime of an electronic component. $T \sim \mathrm{Exp}\left(\frac{1}{10}\right)$.
What is the probability the component will fail after 15 years? What is the probability that the component will fail before 2 years?

$$P(T > 15) = 1 - P(T \leq 15) = 1 - \left(1 - e^{-\frac{1}{10} \times 15}\right) = e^{-\frac{3}{2}} = 0.223$$

$$P(T \leq 2) = 1 - e^{-\frac{1}{10} \times 2} = 1 - e^{-0.2} = 1 - 0.818730753 \approx 0.181$$

<u>Example</u>

On a safari, the number of animals seen, per hour, follows an exponential distribution, where $\lambda = \frac{1}{4}$. What is the probability of seeing an animal before 45 minutes? What is the probability of seeing an animal after 30 minutes?

Expected value and variance of an exponential distribution

The expected length of the gap between successive events is

$$E(X) = \frac{1}{\lambda}$$

The variance for the exponential distribution is

$$Var(X) = \frac{1}{\lambda^2}$$

Example

If breakdowns occur at a rate of $\lambda = 2$ per year, what is the expected time between breakdowns? What is the variance?

$$E(\text{time between breakdowns}) = \frac{1}{2}\text{ year}$$

$$Var(\text{time between breakdowns}) = \frac{1}{2^2}\text{year} = \frac{1}{4}\text{year}$$

Example

If flaws in a wire occur at a rate of 0.015 per meter, what is the expected distance between flaws? What is the variance?

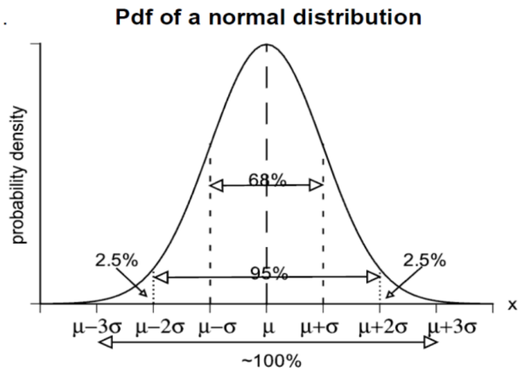
The normal distribution

The normal distribution is a symmetric bell-shape, i.e. Gaussian. A random variable that is the sum of other (not necessarily normally distributed) random components, under fairly lax conditions, will be normally distributed. Hence the mean of a large sample is normally distributed.
The pdf of a random variable with a normal distribution is

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

We say that X is normally distributed with mean, μ, and variance σ^2, which we denote by

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



Expected value and standard deviation of the normal distribution

The expected value for the normal distribution is

$$E(X) = \mu$$

This parameter is at the centre of the distribution.

The standard deviation $\sigma = \sqrt{\sigma^2}$, is the shape parameter which tells us how spread out the distribution is.

For a normal distribution, 95% of the population lies within 1.96 standards deviations of the mean, i.e. $(\mu - 1.96\sigma, \mu + 1.96\sigma)$.

99.74% of the population lies within 3 standard deviations of the mean, i.e. $(\mu - 3\sigma, \mu + 3\sigma)$.

The standard normal distribution

The standard normal distribution has a mean of zero and a variance of unity. Any normal distribution can be shifted and scaled to that of the standard one.
If

$$X \sim N(\mu, \sigma^2), \qquad Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

Note that the denominator of *Z* is the standard deviation, not the variance. This is important.

By setting $\mu = 0 \text{ and } \sigma = 1$, this reduces the complexity of the pdf to

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

but this is still a function that requires numerical integration. A table of standard normal probabilities is provided in these lecture notes.

<u>Example</u>

Use the normal distribution table to find

a) $P(Z > 1.34)$
b) $P(Z < 2.1)$
c) $P(-1.2 < Z < 2)$

a) $P(Z > 1.34) = 0.0901$

b) $P(Z < 2.1) = 1 - P(Z > 2.1) = 1 - 0.0179 = 0.9821$

c) $P(-1.2 < Z < 2) = 1 - P(Z < -1.2) - P(Z > 2) = 1 - 0.1151 - 0.0228 = 0.8621$

<u>Example</u>

Use the normal distribution table to find

a) $P(Z > 0.67)$
b) $P(Z < 1.94)$
c) $P(0.15 < Z < 2.81)$

Example

Noise voltages on a sliding contact are normally distributed with mean 12 micro volts and standard deviation of 1.5 micro volts. What proportion of these noise voltages are

a) greater than 15 micro volts?
b) less than 14.2 micro volts?
c) Between 11 and 14 microvolts?

a) $P(X > 15) = P\left(Z > \frac{15-12}{1.5}\right) = P(Z > 2) = 0.228$

b) $P(X < 14.2) = 1 - P(X < 14.2) = 1 - P\left(Z < \frac{14.2-12}{1.5}\right) = 1 - P(Z < 1.467) = 1 - 0.0708 = 0.9292$

c) $P(11 < X < 14) = 1 - P(X < 11) - P(X > 14)$

$$= 1 - P\left(Z < \frac{11-12}{1.5}\right) - P\left(Z > \frac{14-12}{1.5}\right)$$

$$= 1 - P(Z < -0.667) - P(Z > 1.333) = 1 - 0.2514 - 0.0918 = 0.6568$$

Example

A manufacturer claims that certain components have resistances, which are normally distributed, with a mean and standard deviation of 0.5 ohm and 0.045 ohm, respectively.

a) What is the probability that a randomly chosen component has a resistance in excess of 0.6 ohm?
b) What proportion of components is expected to have resistances in excess of 0.45 ohm?

The lognormal distribution

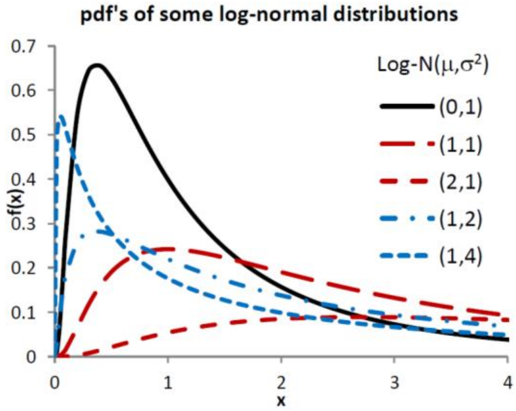
A log-normal distribution is a continuous probability distribution of a random variable whose logarithm is normally distributed. If X is a random variable with normal distribution, then $Y = e^x$ has a log-normal distribution; likewise if Y is log-normal distributed, then $X = \log(Y)$ has a normal distribution. The log-normal distribution is the distribution of a random variable that takes only positive real values.

A variable might be modelled as log-normal if it can be thought of as the multiplicative product of many independent random variables each of which is positive.

A random variable has a lognormal distribution in $\ln(x)$ or $\log_a(x)$ has a normal distribution. Its pdf is

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right]} \; or \; f(x) = \frac{1}{\ln(a)\sigma x\sqrt{2\pi}} \exp\left[-\frac{(\log_a(x)-\mu)^2}{2\sigma^2}\right]$$

where the parameters μ and σ^2 are the mean and variance of the underlying normal distribution. σ^2 relates mostly to the skewness of the distribution of X. the larger it is, the more skewed the lognormal distribution is. Both parameters affect the position of the peak.



Calculating probabilities and percentage points for the lognormal distribution involves working with the normal distribution and transforming to the original lognormal scale as appropriate.

Example

A probability plot suggest that the lifetime X (in hours) of a component has a lognormal distribution with the mean and standard deviation of $\ln(X)$ being 10 and 0.5 respectively.

a) What proportion of components have lifetimes over 50000 hours?
b) What lifetime is exceeded by 10% of components?

a) $P(X > \ln(50000)) = P\left(Z > \frac{\ln(50000)-10}{0.5}\right) = P(Z > 1.64) = 0.0505$

b) $P(Z > 1.28) = 0.1003$

$$1.28 = \frac{\ln(x) - 10}{0.5}$$

$$0.64 = \ln(x) - 10$$
$$10.64 = \ln(x)\,, x = 41772.77$$

The lifetime exceeded by 10% of components is 41772 hours.

Example

The annual rainfall in Plymouth may be assumed to have a log-normal distribution. The logarithm of the rainfall figures (in mm) has been calculated and the results give a mean of 5.9 and a standard deviation of 1.3. Assuming that current conditions prevail,

a) What is the probability of more than 1200mm of rain in a particular year?
b) What is the probability of less than 150 mm of rain in a particular year?
c) What value of annual rainfall is the least that is expected in 95% of all years?

The Weibull distribution

The parameters of the Weibull distribution have large impact on its shape. It is widely used in reliability modelling because of the variety of shapes it can take. Thus it can be used to model many different types of life time data.
The pdf for the Weibull distribution is

$$f(x) = \alpha\beta^{-\alpha}x^{\alpha-1}\exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right], \qquad x > 0$$

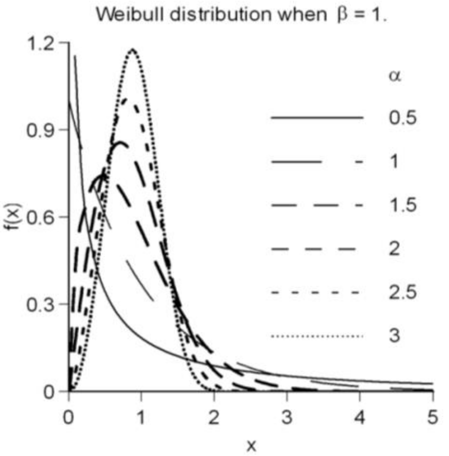
It is often much easier to work with its cdf

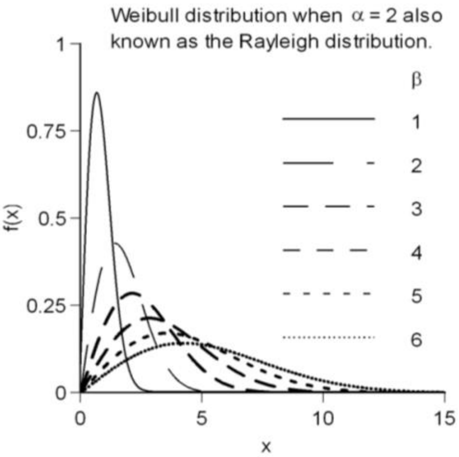
$$F(x) = P(X < x) = 1 - \exp\left[-\left(\frac{x}{\beta}\right)^{\alpha}\right].$$

This distribution has two positive parameters: α- shape parameter and β- scale parameter, so named as α has the greatest effect on the shape while increasing β spreads the basic distribution shape out.

Special cases

- When α=1, the Weibull pdf reduces to $f(x) = \frac{1}{\beta}e^{-x/\beta}$, which is the pdf of an exponential distribution with rate $\frac{1}{\beta}$ or mean *β.*
- When α=2, the Weibull pdf reduces to a form known as the Rayleigh distribution which is sometimes met in reliability literature.





<u>Example</u>

The lifetime, in 1000s of hours, of a certain type of valve has a Weibull distribution with $\alpha = 1.5$ and $\beta = 4.7$.

a) What is the probability that a particular valve will fail in less than 2000 hours?
b) What is the probability that a particular valve will survive for more than 3500 hours?

$$P(X < 2000) = 1 - \exp\left(-\left(\frac{2}{4.7}\right)^{1.5}\right) = 1 - e^{-0.277586414} = 1 - 0.757610093 \approx 0.242$$

$$P(X > 3500) = 1 - P(X < 3500) = 1 - \left(1 - \exp\left(-\left(\frac{3.5}{4.7}\right)^{1.5}\right)\right) = \exp\left(-\left(\frac{3.5}{4.7}\right)^{1.5}\right)$$

$$= e^{-0.642621541} \approx 0.526$$

Example

The lifetime of certain drill bits is well approximated by a Weibull distribution with $\alpha = 2$ and $\beta = 3$, in drilling hours.

a) What is the probability that a drill bit survives for less than 1 drilling hour?
b) What is the probability that a drill bit survives for more than 5 drilling hours?

Key points

- A continuous r.v. can take all values in a certain interval.

- A density curve describes the probability distribution of a continuous r.v.

- The probability distribution is scaled such the area under the density curves is equal to 1.

- The expected value and variance of a continuous distribution is found by

$$E(X) = \int_{-\infty}^{\infty} x \cdot f(x)dx\,, \qquad V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$$

Parameters

The exact distribution to use depends on the values of parameters which may be estimated from the data. These parameters and their effect on the distribution are

Distribution		Parameters	Interpretation
Discrete	Binomial	n, p	p is the probability of success, n is the number of trials
	Poisson	λ	Mean rate of occurrence
Continuous	Normal	μ, σ^2	Location (mean) and spread (variance)
	Lognormal	μ, σ^2	Location and spread of log(X)
	Exponential	λ	Reciprocal of the mean
	Weibull	α, β	Shape and scale

Mechanisms

These distributions may arise if the variable measured has the type shown below

Distribution	Process
Binomial	A fixed number of experiments each with only 2 possible outcomes
Poisson	The number of events that occur at a constant but random rate
Normal	The result of adding many independent effects
Lognormal	The result of multiplying many independent effects
Exponential	The gaps between Poisson events that occur at a constant but random rate
Weibull	The gaps between event occurring with a variety of failure patterns

Finding Probabilities

In general, probabilities for discrete distributions are found by summing point probabilities while continuous ones are calculated by integrating the pdf.

Distribution	Method
Binomial	$P(X = x) = {}^{n}C_{x} p^{x}(1-p)^{n-x}$
Poisson	$P(X = x) = \frac{\lambda^{x} e^{-\lambda}}{x!}$
Normal	Use tables
Lognormal	As for Normal but the log(X)
Exponential	Use cdf: $P(X \leq x) = 1 - \exp(-\lambda x)$
Weibull	Use cdf: : $P(X \leq x) = 1 - \exp(-(x/\beta)^{\alpha})$

Reliability

Introduction

When manufactures claim that their products are very reliable they essentially mean that the products can function as required for a long period of time, when used as specified. In order to assess and improve the reliability of an item we need to be able to measure it. Thus a more formal definition is required.
When an item stops functioning satisfactorily it is said to have failed. The time-to-failure or life-time of an item is intimately linked to this reliability and this is a characteristic that will vary from item to item even if they are supposedly identical. For example, say we have 100 identical brand new light bulbs that we plug into a test circuit, turn on simultaneously and observed how long they last. We would not, of course, expect them all to fail at the same time. Their times to failure will differ. Furthermore, there is some random element to their failure times so their lifetime is a random variable whose behaviour can be modelled by a probability distribution. This is the basis of reliability theory.

Reliability analysis enables us to answer such questions as:

- What is the probability that a unit will fail before a given time?
- What percentage of items will last longer than a certain time?
- What is the expected lifetime of a component?

Fundamental concepts associated with reliability

Some important functions

Suppose the failure density function (pdf) is given by $f(x)$, we can define some other functions as follows:

- The cdf is $F(t) = \int_0^t f(u)du$
- The reliability function (the probability that a unit survives to time t) is $R(t) = 1 - F(t)$. The reliability function is always a decreasing function of time.
- The hazard function which is the instantaneous failure rate is $h(t) = \frac{f(t)}{R(t)}$.

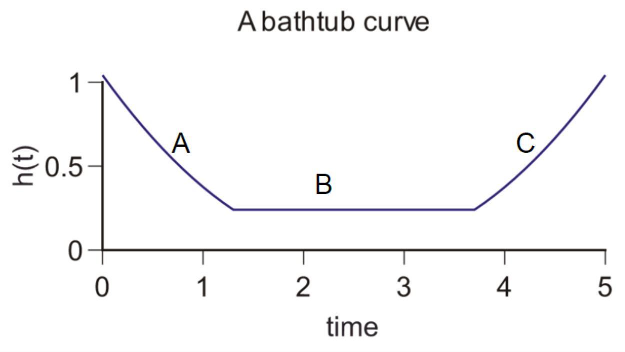
The hazard function indicates the risk of a unit failing after time t has elapsed. However, it is not a probability (eg. it can take values greater than 1). If we want to work in probabilities we must use the reliability function, whose relationship to $h(t)$ is

$$R(t) = \exp\left(-\int_0^t h(u)du\right)$$

- The cumulative hazard function is $H(t) = \int_0^t h(u)hu = -\ln(R(t)$. An upwardly-curving cumulative hazard function (chf) indicates an increasing failure rate as the component ages, whilst a downwardly-curving chf indicates a decreasing failure rate.
- The mean time between failures (MTBF) is a concept which is frequently used in reliability work. It is defined to be the average or expected lifetime of an item.
$$\text{MTBF} = \int_0^\infty tf(t)dt$$
- $f(t), F(t), R(t), h(t)$ and $H(t)$ give mathematically equivalent descriptions of the lifetime, $T,$ in the sense that, given any one of these functions, the other four functions may be deduced.

Bathtub curve

From the graph of $h(t)$ we obtain a useful picture of the likely risk of failure of a unit throughout its lifetime. In practice, the hazard function of a unit or system is often assumed to change in the following manner, the so called bathtub curve.



For such a curve, the different periods of the items life are modelled by different standard distributions for $f(t)$.

For example:

- Region A: decreasing failure rate: often due to so called 'infant mortality' or 'burn in'. Use Weibull distribution.
- Region B: constant failure rate for majority of lifetime. Use exponential distribution.
- Region C: increasing failure rate: often due to items 'wearing out' or 'ageing'. Use the Weibull distribution.

It should be noted that, in reality, an items hazard function (instantaneous failure rate) is unlikely to remain constant throughout the majority of its lifetime. Most items age gradually throughout this period but constant is an acceptable simplification. In practice, $h(t)$, other than the Bathtub curve may be obtained, in which case other combinations of distributions can be used to model the items lifetime at the different periods throughout its life, as appropriate.

Lifetime following an exponential distribution

We have previously seen that the exponential distribution is a useful model of time (or length) between failures, when the failures are happening at random and at a constant rate, λ. The pdf is

$$f(t) = \lambda e^{-\lambda t}$$

The associated functions are:

- The cdf: $F(t) = \int_0^t \lambda e^{-\lambda u} du = \left[\frac{\lambda e^{-\lambda u}}{-\lambda}\right] = -e^{-\lambda t} + 1$
- The reliability function: $R(t) = 1 - F(t) = 1 - \left(-e^{-\lambda t} + 1\right) = e^{-\lambda t}$
- The hazard function: $h(t) = \frac{f(t)}{R(t)} = \frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$

Referring back to the so called bathtub curve for the hazard function of a system or item, we can see that within region B the hazard function is constant. That is, within region B the lifetime of the system can be modelled by an exponential distribution.

- The mean time before failure: $\text{MTBF} = \frac{1}{\lambda}$

Example

A system has a failure rate of 2×10^{-3} failures/hour. What is the mean time between failures?

$$\lambda = 2 \times 10^{-3}$$
$$\text{So, MTBF} = \frac{1}{2 \times 10^{3}} = 500 \text{ hours.}$$

Example

The number of failures of a circuit, per year, follows a Poisson distribution with mean 0.23. Therefore, the time between failures follows an exponential distribution. What is the mean time between failures?

Example

A component has a failure rate of 5 failures/ 10^6 hours. What is the probability that the component will still be working after 10,000 hours of operation?

$$\lambda = 5 \times 10^{-6}$$
$$P(X > 10000) = R(10000) = e^{-0.000005 \times 1000} = 0.951229424$$

Example

Assume that a guided missile has a true MTBF of 2 hours and a constant failure rate. What is the probability of the missile failing to complete a two-hour mission?

Lifetime following a Weibull distribution

Associated functions

If lifetimes have a Weibull distribution

- The pdf: $f(t) = \alpha\beta^{-\alpha}t^{\alpha-1}\exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right]$
- The cdf: $F(t) = 1 - \exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right]$
- The reliability function: $R(t) = \exp\left[-\left(\frac{t}{\beta}\right)^{\alpha}\right]$
- The hazard function: $h(t) = \alpha\beta^{-\alpha}t^{\alpha-1}$. This will give various forms for the hazard function depending on the value of α.
- It can also be shown that $\text{MTBF} = \beta\Gamma\left(1 + \frac{1}{\alpha}\right)$, where Γ denotes the gamma function. Tabulated values of the gamma function are available to assist in calculating the mean time between failures.

Example

The failure function of a component (in thousands of hours) has the following Weibull parameters, $\alpha = 2$ and $\beta = 0.5$.

a) Find the probability that the component will fail before 2000 hours of operation.
b) What is the expected mean time before failure for these components?

$$P(X > 2000) = R(2) = \exp\left[-\left(\frac{2}{0.5}\right)^2\right] = e^{-16} = 0.0000001125351747$$
$$\approx 0.000000113$$

Required probability= $1 - 0.0000001125531747 = 0.999999887 \approx 1$

$$\text{MTBF} = 0.5 \times \Gamma\left(1 + \frac{1}{2}\right) = 0.5 \times \Gamma(1.5) = 0.5 \times 0.8862 = 0.4431$$

$$\text{MTBF} = 443.1 \text{ hours}$$

Example

A relay is critical in the operation of a piece of communications equipment. If the lifetime of the relay, in 10^6 actuations, follows a Weibull distribution where $\alpha = 2.04$ and $\beta = 21.19$:

$$f(t) = 0.004t^{1.04}\exp\left[-\left(\frac{t}{21.19}\right)^{2.04}\right]$$

a) Estimate the expected lifetime of this relay.
b) Write down the reliability function and calculate the probability that one of these relays will still be operating after a million actuations.

Lifetimes following normal or lognormal distributions.

Although the normal distribution is the most commonly used distribution in statistics, it is rarely used as a failure distribution function. The lognormal distribution is used more frequently in which case log(lifetime) is taken to have a normal distribution. The associated functions of these distributions are more complicated than those for the exponential and Weibull distributions although statistical software can deal with them just as easily.

Key points

- Reliability analysis allows us to find the probability that a unit will fail before a given time.

- Reliability analysis allows us to find the percentage of items which will last longer than a certain time.

- Reliability analysis allows us to find the expected lifetime of a component.

- The cdf is $F(t) = \int_0^t f(u)du$

- The reliability function (the probability that a unit survives to time t) is $R(t) = 1 - F(t)$. The reliability function is always a decreasing function of time.

- The hazard function which is the instantaneous failure rate is $h(t) = \frac{f(t)}{R(t)}$

- If we want to work with probabilities we need to use the following reliability function
$$R(t) = \exp\left(-\int_0^t h(u)du\right)$$

- The cumulative hazard function is $H(t) = \int_0^t h(u)hu = -\ln(R(t)$.

- The mean time between failures (MTBF) is a concept which is frequently used in reliability work. It is defined to be the average or expected lifetime of an item.
$$\text{MTBF} = \int_0^\infty tf(t)dt$$

- We often assume the hazard function changes in a manner that produces the bathtub curve.

	Exponential	Weibull
$f(t)$	$\lambda e^{-\lambda t}$	$\alpha\beta^{-\alpha}t^{\alpha-1}\exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right]$
$F(t)$	$1 - \lambda e^{-\lambda t}$	$1 - \exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right]$
$R(t)$	$e^{-\lambda t}$	$\exp\left[-\left(\frac{t}{\beta}\right)^\alpha\right]$
$h(t)$	λ	$\alpha\beta^{-\alpha}t^{\alpha-1}$
MTBF	$\frac{1}{\lambda}$	$\beta\Gamma\left(1 + \frac{1}{\alpha}\right)$

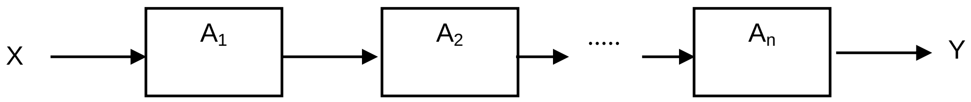
System Reliability

Suppose that we have to calculate the reliability of a system made up of several components. The total reliability can be calculated by calculating the reliability of each individual component and combining these individual reliabilities. The way in which they are combined depends on the way in which the components are connected. That is, whether they are connected

- in series;
- in parallel;
- in a combination of both.

Components in series

Consider a system of *n* components connected in series so that the system will only work (i.e. a signal will pass from X to Y) if all the components work.



If a components failure independent of that of the others, it is easy to show that if R_1, R_1,…, R_n the reliabilities of the individual components, than the reliability of the system is given by

$$R_{S-system} = R_1 \times R_2 \times \ldots \times R_n = \prod_{i=1}^{n} R_i \, .$$

__Example__

Consider a system of three components connected in series, each component having a constant failure rate. The rate for components A, B, and C are 0.2, 0.4, 0.5 per 10000 hours, respectively.

a) Find the overall reliability of the system.
b) Find the probability that the system is still working after 20000 hours

As the question informs us that the failure rate is constant, we use the exponential distribution.

$$R(t) = e^{-\lambda t}$$
$$R_s = e^{-0.2t} \times e^{-0.4t} \times e^{-0.5t} = e^{-1.1t}$$
$$R(2) = e^{-1.1\times 2} = e^{-2.2} = 0.110803158 \approx 0.1108$$

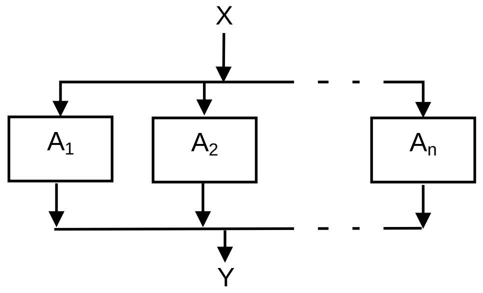
__Example__

Consider a system of four components connected in series, each component having a constant failure rate. The rate for components *A*, *B*, *C*, and *D* are 0.05, 0.1, 0.25, and 0.15 per 1000 hours, respectively.

a) Find the overall reliability of the system.
b) Find the probability that the system is still working after 5000 hours.

Components in parallel

If *n* components are connected in parallel, so that the system works (signal from X to Y) as long as at least one of the components works,



The reliability of the system (again assuming independent failures) is

$$R_{p-\,system} = 1 - (1 - R_1)(1 - R_2)\ldots(1 - R_n) = 1 - \prod_{i=1}^{n}(1 - R_i) = 1 - \prod_{i=1}^{n} F_i$$

where F_i is the cdf.

<u>Example</u>

Component A has a constant failure rate of 1.5 per thousand hours and component B has a constant failure rate of two per 1000 hours are connected in parallel.

a) Find the overall reliability of the system, working in thousands of hours
b) what is a probability that the system is still working after 1000 hours?

$$R_A(t) = e^{-1.5t} \text{ and } R_B(t) = e^{-2t}$$

$$R_{system}(t) = 1 - \left[\left(1 - R_A(t)\right)\left(1 - R_B(t)\right)\right] = 1 - [(1 - e^{-1.5t})(1 - e^{-2t})]$$

$$= 1 - [1 - e^{-1.5t} - e^{-2t} + e^{-1.5t}e^{-2t}] = e^{-1.5t} + e^{-2t} - e^{-1.5t-2t}$$

$$= e^{-1.5t} + e^{-2t} - e^{-3.5t}$$

$$R(1) = e^{-1.5\times1} + e^{-2\times1} - e^{-3.5\times1} = 0.32826806 \approx 0.3283$$

<u>Example</u>

Three components, *A, B* and *C* are connected in parallel, with constant failure rates 0.4, 0.21 and 0.34, per hour, respectively.

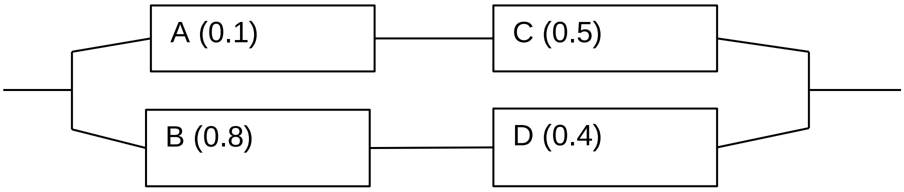
a) Find the overall reliability of the system.
b) what is a probability that the system is still working after 35 hours?

Mixed systems

To evaluate reliability of a system comprising of both parallel and series sections, divide the system into series only and parallel only subsystems. Find the reliability of each subsystem as above than combining a suitable manner.

Example

Consider the system depicted below, where the numbers in brackets indicate the constant failure rates per year for each of the components.



a) What is the reliability of the system?
b) Find the probability that system will still be operational after six months.

The first step is to split the system into two branches where both are a 'subsystem' in series.

Reliability of top branch = $R_{top}(t) = e^{-0.1t} \times e^{-0.5t} = e^{-0.6t}$
Reliability of bottom branch = $R_{bottom}(t) = e^{-0.8t} \times e^{-0.4t} = e^{-1.2t}$
Now we can combine these two subsystems into a larger parallel system.

$$R_{system}(t) = 1 - [(1 - e^{-0.6t})(1 - e^{-1.2t})] = 1 - [1 - e^{-0.6t} - e^{-1.2t} + e^{-1.8t}]$$
$$= e^{-0.6t} + e^{-1.2t} - e^{-1.8t}$$
$$R(0.5) = e^{-0.6\times0.5} + e^{-1.2\times0.5} - e^{-1.8\times0.5} = e^{-0.3} + e^{-0.6} - e^{-0.9}$$
$$= 0.883060197 \approx 0.883$$

Example

Components of type A have a constant failure rate of 1.5 per thousand hours. Components of type B have lifetimes which follow a Weibull distribution with $\alpha = 3$ and $\beta = 2$ in thousands of hours.

a) State the reliability function for each component.
b) If two subsystems, consisting of an A component and a B component connected in series, are connected in parallel to form a component, what is the reliability of the component?

Key points

- Components in series: $R_{S-system} = R_1 \times R_2 \times \dots \times R_n = \prod_{i=1}^{n} R_i$.
- Components in parallel: $R_{p-\,system} = 1 - (1 - R_1)(1 - R_2) \dots (1 - R_n) = 1 - \prod_{i=1}^{n}(1 - R_i) = 1 - \prod_{i=1}^{n} F_i$.
- The systems including some components connected in series and some in parallel we need to break the system into suitable subsystems and combined appropriately.

Estimation

Point estimation

A point estimate is a rule or formula that tells us how to use a sample data to calculate a single number estimate of a population parameter.

In general, we wish to estimate population parameters using suitable sample statistics. Here, given a sample $x_1, x_2, \dots, x_n$, we will consider how we can use

- the sample mean, $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$, to estimate that of the population, μ;
- the sample standard deviation, $s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}$ to estimate, σ;

The standard error

As the size of the sample increases, the mean of the sample approaches the meaning of the population, to the error in approximating the population mean by the sample mean decreases. It would be zero if we tested the entire population. The error in our estimate, known as the standard error, is

$$SE = \frac{\sigma}{\sqrt{n}}, \qquad \text{where } n \text{ is the sample size.}$$

Example

A random sample of size 50 of lightbulbs is taken and tested for the length of life. The sample has a standard deviation of 1.75, what is the standard error for this sample?

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{1.75}{\sqrt{50}} = 0.247487373 \approx 0.247$$

Confidence intervals for population means

If we want to estimate the mean for a population, μ, but we only have the mean of a sample, $\bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k$, we can only do so to within some margin of error.

The large numbers theorem and the central limit theorem

The large numbers theorem I:
Let $x_1, x_2, \dots, x_n$ be identically-distributed variables whose true mean (expected value) is μ (the same for all variables). Then,

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) = \mu.$$

Large numbers theorem II:

Let $x_1, x_2, \dots, x_n$ be identically-distributed variables whose true mean are μ_k (different). Then,

$$\lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} \mu_k\right).$$

Central limit theorem:

Let $x_1, x_2, \ldots, x_n$ be identically-distributed variables whose true mean and standard deviation are μ and $\sigma \neq 0$. As the sample mean is the scaled sum of the *x*'s, it has a normal distribution with $n \to \infty$:

$$\lim_{n\to\infty} \bar{x} = \lim_{n\to\infty}\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

$$\lim_{n\to\infty} Z = \lim_{n\to\infty}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right) \sim N(0,1)$$

Note that in the central limit theorem we do not specify the distribution of the variables-it may be any.
There are different formulations of the central limit theorem. For example, Lyapunov's theorem:
Let $x_1, x_2, \ldots, x_n$ be identically-distributed variables, each with finite true mean (expected value) $\mu_i = E(x_i)$ and variance $\sigma_i = E(\mu_i - x_i)^2 \neq 0$. If for some $\delta > 0$, the Lyapunov's condition

$$\lim_{n\to\infty}\left(\frac{\sum_{k=1}^{n}(x_k-\mu_k)^{2+\delta}}{(\sum_{k=1}^{n}\sigma_k^2)^{\delta}}\right) = 0$$

is satisfied, then a sum of $\frac{\sum_{k=1}^{n}(x_k-\mu_k)}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}}$ converges in distribution to a standard normal random variable, as n goes to infinity:

$$\frac{\sum_{k=1}^{n}(x_k-\mu_k)}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}} \sim N(0,1)$$

In practice it is usually easiest to check the Lyapunov's condition for $\delta = 1$.

σ known

Based on the central limit theorem, we can estimate the possible margin of error in our sample mean compared to the true mean and hence obtain a confidence interval for the true mean:

$$\mu \in \left((\bar{x}) - z\frac{\sigma}{\sqrt{n}}, (\bar{x}) + z\frac{\sigma}{\sqrt{n}}\right)$$

$P\left(-z \leq \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \leq z\right) = 1 - 2\alpha$ is the probability for $\mu \in \left((\bar{x}) - z\frac{\sigma}{\sqrt{n}}, (\bar{x}) + z\frac{\sigma}{\sqrt{n}}\right)$
where z is the critical value of the standard normal distribution that corresponds the confidence level α.

- The confidence interval is symmetric about the sample mean;
- the critical value has to be found from the equation $F(z) = \int_{-\infty}^{z} \exp\left(-\frac{x^2}{2}\right) dx = 1 - \alpha$;
- $z\frac{\sigma}{\sqrt{n}}$ is the margin of error.

Suppose that we can accept a 5% risk of being wrong. Then, we could be 95% sure that the mean of the population lives within a confidence interval. We can never be a hundred percent sure without measuring the entire population.

Example

Wet chemical etching is used to remove silicon from the backs of wafers before metallisation during the manufacture of semiconductors. The standard deviation of the process is known to be 0.39. The average etch rate of a random sample of 10 batches has been found to be 9.94 mm/min. Find the 95% confidence interval for the true etching rate. Find the 99% confidence interval for the true etching rate.

$$\mu \in \left(9.94 \pm 1.96 \times \frac{0.39}{\sqrt{10}}\right) = (9.94 \pm 1.96 \times 0.123328828) = (9.94 \pm 0.241724504)$$

$$\mu \in (9.698275496,10.1817245)$$

$$\mu \in (9.70, 10.18)$$

$$\mu \in \left(9.94 \pm 2.58 \times \frac{0.39}{\sqrt{10}}\right) = (9.94 \pm 2.58 \times 0.123328828) = (9.94 \pm 0.318188378)$$

$$\mu \in (9.621811622,10.25818838)$$

$$\mu \in (9.62, 10.26)$$

Example

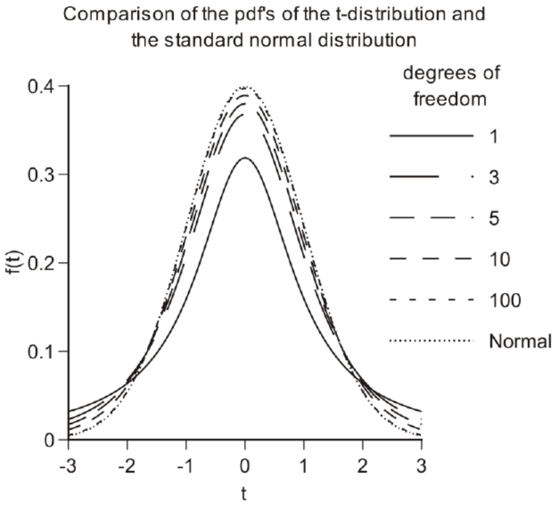
The weight, w, of a component is normally distributed and the standard deviation of the weights is known to be 0.139g. The mean weight of a batch of 30 components is found to be 2.04g. Find the 95% confidence interval for the true mean weight. Find the 99% confidence interval for the true mean weight.

σ unknown

In practice, we usually do not know σ. In this case, our best available estimate is $s = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}$ so we replace the population standard deviation by that of the sample and allow for the extra level of uncertainty in our figures by taking our critical values from a t distribution:

$$\mu \in \left(\bar{x} - t\frac{s}{\sqrt{n}}, \bar{x} + t\frac{s}{\sqrt{n}}\right).$$

This can be applied to any large sample but only to those from normal distributions if the sample size is small.



The t-distribution

A t-distribution depends on the sample size and n:

$$t_n(x) = \frac{\Gamma\left(\frac{n}{2}\right)}{\sqrt{2\pi}\Gamma\left(\frac{n-1}{2}\right)} \times \frac{1}{\sqrt{n-1}} \times \left(1 + \frac{x^2}{n-1}\right)^{-n/2}$$

Notice that it is a bell-shaped curve like the normal distribution but for small sample sizes, it is wider.

$$\lim_{n\to\infty} t_n(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

The spread is given by the number of degrees of freedom where $DF = n - 1$.

The area below the curve between $-t$ and t is the probability required. A table of these critical values is provided in these notes.

Example

A different, less expensive, etching chemical is trialled. The etching rate in mm/min for a random sample size 6 are

10.2 9.4 10.0 9.8 10.1 9.9

a) Constructing 95% confidence interval for the mean etching rate using this new chemical.
b) Is there any evidence of the 95% confidence level that the etching rate is less than 10 mm/min?

$$\bar{x} = \frac{10.2 + 9.4 + 10.0 + 9.8 + 10.1 + 9.9}{6} = \frac{59.4}{6} = 9.9$$

$$s^2 = \frac{1}{6-1} \times [(10.2 - 9.9)^2 + (9.4 - 9.9)^2 + (10.0 - 9.9)^2 + (9.8 - 9.9)^2 + (10.1 - 9.9)^2 + (9.9 - 9.9)^2]$$

$$= \frac{1}{5} \times [0.3^2 + (-0.5)^2 + 0.1^2 + (-0.1)^2 + 0.2^2 + 0^2] = \frac{1}{5} \times 0.4 = 0.08$$

$$s = \sqrt{0.08} = 0.282842712$$

Our t value: $n - 1 = 6 - 1 = 5$, we need the t value for 5 degrees of freedom at $\alpha = 0.025$

$$t_{5,0.025} = 2.571$$

$$\mu \in \left(9.9 \pm 2.571 \times \frac{\sqrt{0.08}}{\sqrt{6}}\right) = (9.9 \pm 2.571 \times 0.115470053) = (9.9 \pm 0.296873508)$$

$$\mu \in (9.603126492, 10.19687351)$$

$$\mu \in (9.60, 10.20)$$

As 10mm/min is within the 95% confidence interval, there is no evidence in the 95% confidence level that the etching rate is less than 10 mm/min.

Example

A manufacturer claims that certain components have resistances, which are normally distributed, with a mean and standard deviation of 0.5 ohm and 0.045ohm, respectively. You test six randomly selected components and get the following resistances, measured in ohms:

0.44 0.48 0.53 0.45 0.57 0.48

Use this sample to evaluate the 95% confidence interval for the true mean resistance of these components. Does this sample provide any evidence at the 95% confidence level that the mean resistance is not 0.5 ohm?

For the above examples:

What assumption does this calculation require?

We must be confident that the data are a simple random sample and that the distribution of etching rate is approximately normal because you’re such a small sample. We must learn how the sample chose and see whether it can be regarded as a simple random sample.

How do you verify the assumptions?

We would need to check that there were no outliers and the data was not strongly skewed. It is difficult to assess normality with only six observations. In practice, we would probably rely on the fact that past measurements of this type have been roughly normal.

<u>Example</u>

35 readings of the power output from the electronic device are found have a mean of 54.62 J and a standard deviation of 0.34 J.

a) Obtain a 95% confidence interval for the true power output for this device.
b) What is the minimum size sample that we needed to obtain 99% confidence interval the same margin of error as found in part a?

$$\mu \in \left(54.62 \pm 1.96 \times \frac{0.34}{\sqrt{35}}\right) = (54.62 \pm 1.96 \times 0.057470489) = (54.62 \pm 0.112642159)$$
$$\mu \in (54.50735784, 54.73264216)$$
$$\mu \in (54.51, 54.73)$$

The margin of error for the 95% confidence interval is $E = 1.96 \times \frac{0.34}{\sqrt{35}} = 0.112642159$

The margin of error for the 99% confidence interval is $E = 2.56 \times \frac{0.34}{\sqrt{n}} = 0.112642159$, what value of n do we require?

$$n = \left(\frac{2.56 \times 0.34}{E}\right)^2 = \left(\frac{2.56 \times 0.34}{0.112642159}\right)^2 = 60.64504373 \approx 61$$

The minimum sample size to obtain a 99% confidence interval with the same margin of error as the 95% confidence interval is 61.

Hypothesis testing

The confidence interval for the population proportion can also be used to test for the true proportion for the population.
Using the formula below:

$$\left|\sqrt{\frac{n}{p(1-p)}}\left(\hat{P}(A) - p\right)\right| \leq z$$

If this statement is true, then we accept the hypothesis that has been tested. If this statement does not hold, then we reject the hypothesis that has been tested.

<u>Example</u>

In a Bernoulli scheme, after $n = 3000$ trials, the frequency of the event A was found to be $n_A = 1450$. Using the central limit theorem and the table for the cumulative standard normal distribution, check the hypothesis that the probability of A is 0.5 within the 98% confidence interval.

$$p = 0.01 \ \rightarrow z = 2.33$$
$$\left|\sqrt{\frac{3000}{0.5(1-0.5)}} \times \left(\frac{29}{60} - \frac{30}{60}\right)\right| \leq 2.33$$
$$\left|\sqrt{\frac{3000}{0.5^2}} \times \left(-\frac{1}{60}\right)\right| \leq 2.33$$

$$\left|\sqrt{\frac{3000}{0.25}} \times \left(-\frac{1}{60}\right)\right| \leq 2.33$$

$$\left|\sqrt{12000} \times \left(-\frac{1}{60}\right)\right| \leq 2.33$$

$$|-1.825741858| \leq 2.33$$

This statement is true, therefore we have no evidence at the 98% confidence level to reject the manufacturers claim.

Example

A manufacturer of resistors claims that less than 0.4% of their products are defective. 12 defective resistors are found in a random sample of 1500. Is there any evidence that the claim is false at the 95% confidence level?

Which interval to use?

The conference interval to use remain depends on

- whether the original distribution of X is (at least approximately) normal
- whether σ is known
- sample size of n

The following table is one way of summarising the choice:

σ	n	X is normal	X not normal
Known	≤ 30	σ and z	σ and z
	> 30	σ and z	-
Unknown	≤ 30	s and z	s and z
	> 30	s and t	-

So the confidence intervals described here cannot be used if X is not normally distributed in the sample is small.

Correlation and regression

Correlation

Scatter plots

To study the relationships between variables, we must measure each of the variables in the same group of individuals.

If we think that a variable x may cause changes in another variable y, i.e. y depends on x, then

- x is an explanatory or independent variable;
- y is a response or dependent variable.

If you suspect that there is a relationship between two or more variables, you should **ALWAYS** plot y against (each) x as simple, unconnected points i.e. produce a scatter plot (s), with the

- x on the horizontal axis;
- y on the vertical axis.

Interpreting the scatter plot

Look for the overall pattern:

- form
- directions
- positive association
- negative association
- strength

Look for deviation from the overall pattern.

Correlation

Correlation, r, measures the strength and direction of the ***linear*** association between two quantitative variables x and y.
Given a dataset that contains n pairs of values, x_i and y_i, $i = 1, \ldots n$, the correlation between the variables x and y is given by

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left[\left(\frac{x_i - \bar{x}}{s_x}\right)\left(\frac{y_i - \bar{y}}{s_y}\right)\right]$$

Where s_x and s_y are the sample standard deviations of x and$y$, respectively.

Your calculator is likely to have the correlation function built in. You may need to set it to regression mode.

- r only measures the linear relationship
- $-1 \leq r \leq 1$
- $r > 0$-positive association
- $r < 0$-negative association
- $|r| = 1$-perfect correlation (points lie exactly on a straight line)

Example

Find the correlation between the following two datasets where

$$\bar{x} = 3, \quad \bar{y} = 4, \quad s_x = 1, \quad s_y = 0.763763$$

x	y	$x - \bar{x}$	$y - \bar{y}$	$\frac{x - \bar{x}}{s_x}$	$\frac{y - \bar{y}}{s_y}$	$\left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$
3	4	0	0	0	0	0
2	3.5	-1	-0.5	-1	-0.65465	0.65465
4	5	1	1	1	1.3093	1.3093
					Σ	1.96395

$$r = \frac{1}{3-1} \times 1.96395 = 0.981975 \approx 0.982$$

For this example we have strong positive association.
Use this simple dataset to learn how to use your calculator to evaluate r.

Example

Find the correlation between the following two datasets

$$\bar{x} = 5 \qquad \bar{y} = 4 \qquad s_x = \sqrt{19.2} \qquad s_y = \sqrt{5.2}$$

x	y	$x - \bar{x}$	$y - \bar{y}$	$\frac{x - \bar{x}}{s_x}$	$\frac{y - \bar{y}}{s_y}$	$\left(\frac{x - \bar{x}}{s_x}\right)\left(\frac{y - \bar{y}}{s_y}\right)$
12	1					
8	7					
5	4					
3	6					
2	4					
0	2					
					Σ	

Regression

If we have created a scatter plot of some variables that we suspect to be related, and that plot suggests that there might be a relationship between them, we like to drive a formula that links them. You recognise that any experimental results will incur error we do not want an equation that simply joins up all the points on the scatter. Rather, what an equation of a straight line that smooths out the variation. Warming to do this is to fit a function that minimises the square of the difference between the measured y's and those predicted by the equation. This is called least-squares fitting.

Linear regression

The simplest relationship that can exist between two variables is a linear one.
Given $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, would like to draw the best line through the points on the scatter plot.

The least-squares straight line is $\hat{y} = a + bx$
where $a = \bar{y} - b\bar{x}$, $b = \frac{S_{xy}}{S_{xx}}$ and $S_{xx} = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$, $S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$

The correlation between x and y, r, can also be expressed in terms of S_ _:

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

Note:

- $\hat{y}$ is a response predicted by the equation while y is the corresponding measured value.
- b is the gradient of the linear relationship
- a is the intercept
- The regression line passes through the point $(\bar{x}, \bar{y})$.
- $r^2$is the fraction of the variation in the values of Y explain by the linear regression line.

Example

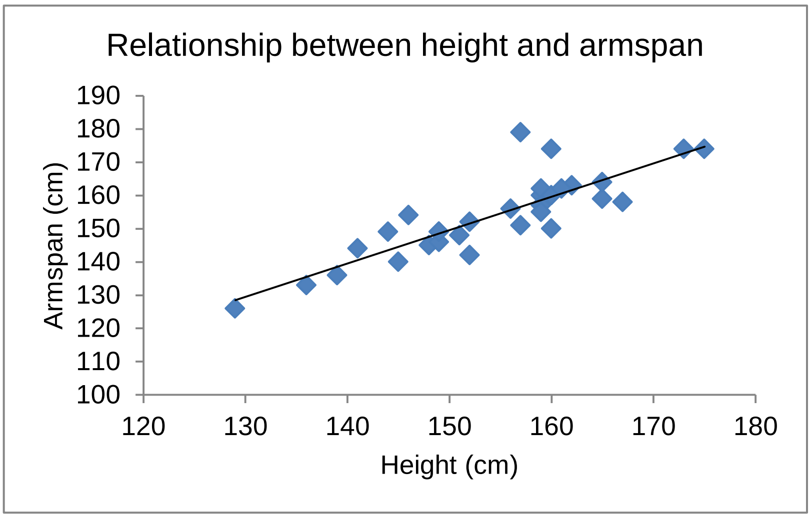
The height and arm span of 32 year 7 males has been collected.

a) Identify the explanatory and the response variables.
b) The scatter plot of the raw data is shown below and suggests the linear fit is appropriate. This should always be done before proceeding to fit a line.
c) Fit the linear least squares regression line and determine the correlation between height and armspan, given that

$$\bar{x} = 154.3226\text{cm}, \qquad \bar{y} = 153.9032\text{cm}$$
$$S_{HH} = 3268.774, \qquad S_{AA} = 4534.71, \qquad S_{HA} = 3281.968$$

The explanatory variable is height.

The response variable is armspan.



$$\hat{y} = a + bx$$

$$b = \frac{S_{HA}}{S_{HH}} = \frac{3281.968}{3268.774} = 1.004036$$

$$a = \bar{y} - b\bar{x}$$

$$= 153.9032 - 1.004036 \times 154.3226$$
$$= 153.9032 - 154.945504$$
$$= -1.042304$$

$$\hat{y} = -1.042304 + 1.004036x$$

$$r = \frac{S_{HA}}{\sqrt{S_{HH}S_{AA}}} = \frac{3281.968}{\sqrt{3268.774 \times 4534.71}} = \frac{3281.968}{\sqrt{14822942.15}}$$

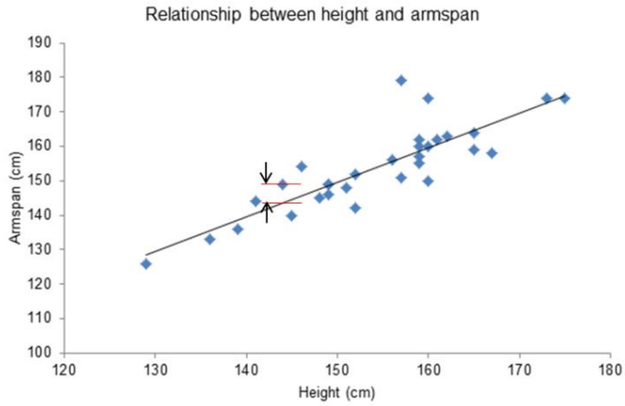
$$= \frac{3281.968}{3850.057421} = 0.852446506 \approx 0.852$$

Example

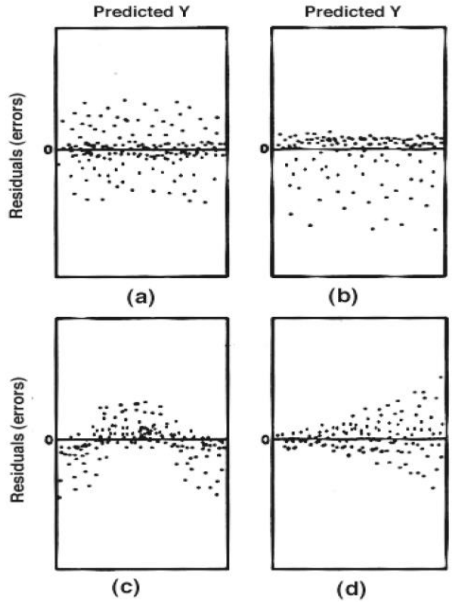
For the data below, find the linear least squares regression equation and the correlation between the variable.

x	y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

Examining the fit



We should examine the fit of the regression line to see whether the assumed form, here linear, is suitable and if so, we should check for influential observations and assess how well the equation fits the data.
The residuals are the difference between the observed response variable values and those predicted by the regression equation. The horizontal lines on this plot show how the residual for an arm span of 149cm is defined.
A scatter plot of the residuals against either the explanatory variable or against the response variable must be produced. A residual plot allows us to check that none of the assumptions made in deriving the equation have been violated.



If the scatter plot of the residuals

a) is a random scatter of uniform density, a linear relationship is likely;
b) has non-uniform density, a periodic factor may be needed.
c) Is curved or has another pattern, a non-linear relationship may exist. Try transferring the data.
d) Increases or decreases in spread about the centre line as x increases, the variance is not consistent. Try transferring the data.

In cases, (b), (c) and (d) the equation would not be valid and should not be used.
An influential observation is a point in the original dataset that particularly affects the gradient of the line. It need not have a large residual. Exclusion of such a point in computing the regression line with lead to very different mind to that obtained when it is included. Often, these points are on the extreme of the explanatory variable. The trustworthiness of an influential observation should be investigated to determine whether or not it should be included.
Once you have what seems to be a valid linear regression model, compute r^2. This tells us the proportion of the variation explained by the model. It provides a measure of accuracy that can be expected when the equation is used for prediction.

Example

Using the previous example for the height and arm span data.

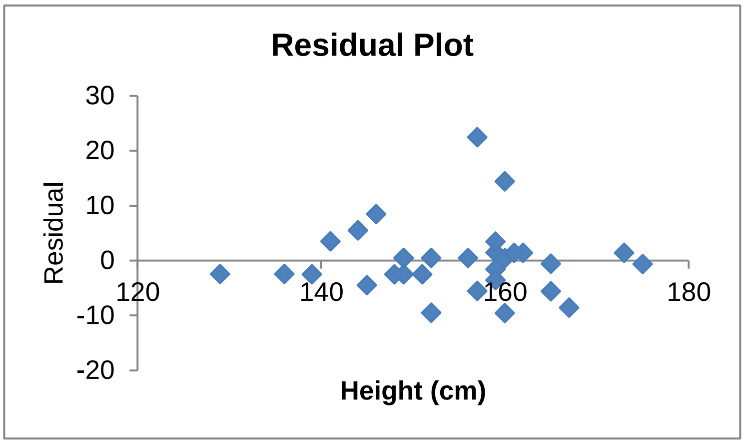
a) Calculate residual for someone who has a height of 165cm and an arm span of 159cm.
b) Comment on the residual plot.
c) What percentage of variation is explained by the regression line?
d) Using the regression line to predict the arm span of a year 11 male who has a height of 178cm.

$$\hat{y} = a + bx = -1.042304 + 1.004036x$$

$x = 165$, $\quad \hat{y} = -1.042304 + 1.004036 \times 165 = -1.042304 + 165.66594 = 164.623636$

Residual: observed –predicted$=159 - 164.623636 = -5.623636$

The residual plot shown a random scattering of residual which are of uniform density and so a linear relationship is likely.



$$r = \frac{3281.968}{3850.057421},$$

$$r^2 = \left(\frac{3281.968}{3850.057421}\right)^2 = 0.726665047$$

$$r^2 \approx 0.73$$

$$\hat{y} = -1.042304 + 1.004036x, \qquad x = 178$$

$$\hat{y} = -1.042304 + 1.004036 \times 178 = -1.042304 + 178.718408 = 177.676104$$

$$177.7\text{cm}$$

This is a reliable estimate because the equation explains 73% of the variation of the data and residual plot suggest that the variance of the data is fairly consistent. Clearly there are other highly influential factors which have not been included in the sample model. Inclusion of these other factors would increase the R squared value and the improved residual plot.

Example

For the data below:

a) Calculate the residual for x=3
b) What percentage of variation is explained by the regression line?
c) Using the regression line, predict the response to x=4.25.

x	y
1.00	1.00
2.00	2.00
3.00	1.30
4.00	3.75
5.00	2.25

Cautions about correlation and regression

We have seen that correlation and regression are powerful tools for describing the relationship between two variables. In practice we need to know the limitations.

- They describe only linear relationship.
- They are not resistant to outliers.
- Using extrapolation to predict for outside the range of x values is usually not accurate.
- If the data has been averaged prior to regression, it may be from more than one population.
- You should always be aware that there may be some 'lurking variable' on which both of your variables depend. Association does not imply causation. The best way to get good evidence that x causes why is to do an experiment in which we change x and hence keep under lurking variables under control.

Linearising techniques

So far we have considered only linear relationships between variables. Before calculating the correlation coefficient regression line, this must be checked with the scatter plot. If there is a curved relationship it can often be made linear simply by transforming one or both of the variables.
Some common transformations are:

- $\ln(y)$ and x $\quad \ln(y) = a + bx$ so $y = e^{a+bx} = e^a e^{bx} = Ce^{bx}$
- y and $\ln(x)$ $\quad y = a + b\ln(x)$
- $\ln(y)$and $\ln(x)$ $\quad \ln(y) = a + b\ln(x)$ so
  $y = \exp(a + b\ln(x)) = e^a \exp(b\ln(x)) = C\exp[\ln(x^b)] = Cx^b$
- y and $\frac{1}{x}$ $\quad y = a + \frac{b}{x}$
- y and $\sqrt{x}$ $\quad y = a + b\sqrt{x}$
- y and x^2 $\quad y = a + bx^2$

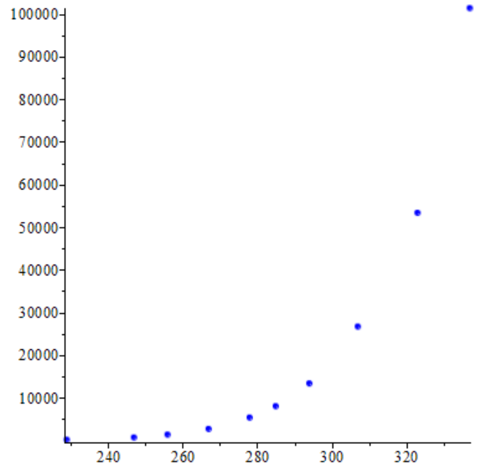
In practice, try if you transformations and see which has the best linearising effect using scatter diagrams. Check the r^2 value-the closer it is to one, the better the linear fit is likely to be.
Having linearising data, proceed with the regression analysis as before, and using the transformed variables. If the y variable has been transformed, transforming equation obtained so that its subject is simply y.

<u>Example</u>

Below is some data taken from an experiment

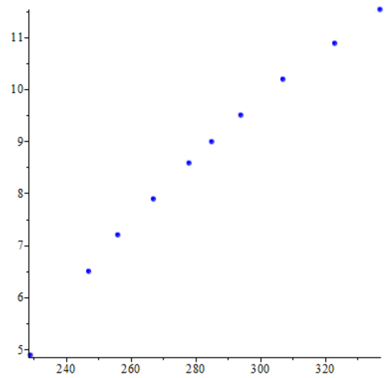
x	229	247	256	267	278	285	294	307	323	337
y	133	666	1333	2666	5332	7999	13332	26664	53328	101325

Find the least squares regression equation and the correlation coefficient.



The data does not look to be linear; therefore the data needs to be transformed.

x	229	247	256	267	278	285	294	307	323	337
y	133	666	1333	2666	5332	7999	13332	26664	53328	101325
ln(y)	4.89	6.50	7.20	7.89	8.58	8.99	9.50	10.19	10.88	11.53



$$\bar{x} = \frac{229 + 247 + 256 + 267 + 278 + 285 + 294 + 307 + 323 + 337}{10} = \frac{2823}{10} = 282.3$$

We need to use the ln(y) values, the values which linearised the data when finding the mean.

$$\bar{y} = \frac{4.89 + 6.5 + 7.2 + 7.89 + 8.58 + 8.99 + 9.5 + 10.19 + 10.88 + 11.53}{10} = \frac{86.15}{10} = 8.615$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$$

$$S_{xx} = (229^2 + 247^2 + 256^2 + 267^2 + 278^2 + 285^2 + 294^2 + 307^2 + 323^2 + 337^2) - 10 \times 282.3^2$$
$$= 807367 - 10 \times 79693.29 = 807367 - 796932.9 = 10434.1$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2$$

$$S_{yy} = (4.89^2 + 6.5^2 + 7.2^2 + 7.89^2 + 8.58^2 + 8.99^2 + 9.5^2 + 10.19^2 + 10.88^2 + 11.53^2) - 10 \times 8.615^2$$
$$= 780.0921 - 10 \times 74.218225 = 780.0921 - 742.18225 = 37.90985$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}$$

$$S_{xy} = (229 \times 4.89 + 247 \times 6.5 + 256 \times 7.2 + 267 \times 7.89 + 278 \times 8.58 + 285 \times 8.99 + 294 \times 9.5 + 307 \times 10.19 + 323 \times 10.88 + 337 \times 11.53) - 10 \times 282.3 \times 8.615$$
$$= 1119.81 + 1605.5 + 1843.2 + 2106.63 + 2385.24 + 2562.15 + 2793 + 3128.33 + 3514.24 + 3885.61 - 24320.145$$
$$= 24943.71 - 24320.145 = 623.565$$
$$\hat{y} = a + bx, \qquad a = \bar{y} - b\bar{x}, \qquad b = \frac{S_{xy}}{S_{xx}}$$
$$b = \frac{623.565}{10434.1} = 0.059762221$$

$$a = 8.615 - 0.059762221 \times 282.3 = 8.615 - 16.87087526 = -8.255875255$$
$$\hat{y} = -8.256 + 0.0598x$$

We had to transform the data first. This equation relates ln(y) to x. To make y the subject:

$$\hat{y} = e^{-8.256+0.0589x} = e^{-8.256}e^{0.0598x} = 0.0002596956992e^{0.0598x}$$
$$\hat{y} = 0.000260e^{0.0598x}$$

x	229	247	256	267	278	285	294	307	323	337
y	133	666	1333	2666	5332	7999	13332	26664	53328	101325
ln(y)	4.89	6.50	7.20	7.89	8.58	8.99	9.50	10.19	10.88	11.53
Predicted response	5.4382	6.5146	7.0528	7.7106	8.3684	8.787	9.3252	10.1026	11.0594	11.9866
residual	-0.5482	-0.0146	0.1472	0.1794	0.2116	0.203	0.1748	0.0694	-0.1794	-0.4566

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{623.565}{\sqrt{10434.1 \times 37.90985}} = \frac{623.565}{\sqrt{395555.1659}} = \frac{623.565}{628.9317657} = 0.991466855$$
$$r^2 = 0.991466855^2 = 0.983006525$$

The equation is a very reliable estimate as 98% of the variation of the data is explained by the equation.

Example

It is known that the variable x and y are related. An experiment is carried out to determine the relationship between them and the results are shown below.

x	0.6667	0.3333	0.5	0.1667	0.2	0.1429	0.1429	0.1	0.0769	0.0714
y	0	2	5	3	5	4	7	10	8	9
x	0.0909	0.0769	0.0714	0.0769	0.0667	0.0556	0.0667	0.0625	0.0526	0.0588
y	11	14	12	16	11	15	23	20	22	18
x	0.0476	0.0417	0.0385	0.0417	0.037	0.04	0.037	0.0333	0.0345	0.0345
y	21	25	24	27	23	28	32	29	27	30

Find the correlation coefficient for the transformed data, the relationship between the variable x and y and plot the residuals.

$$\bar{x} = \frac{3.4656}{30} = 0.11552, \qquad \bar{y} = \frac{481}{30} = 16.03333$$

$$S_{xx} = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2 = 0.99758 - 30 \times 0.11552^2 = 0.597234$$

$$S_{yy} = \sum_{i=1}^{n} y_i^2 - n(\bar{y})^2 = 10351 - 30 \times 16.03333^2 = 2638.967$$

$$S_{xy} = \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} = 29.4067 - 30 \times 0.11552 \times 16.03333 = -26.1584$$

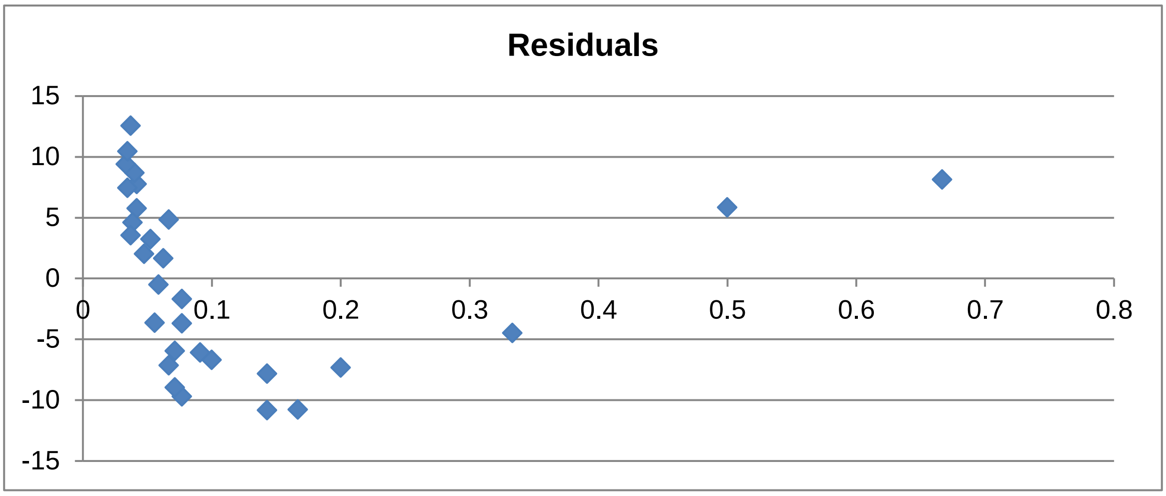
$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{-26.1584}{\sqrt{0.597234 \times 2638.967}} = -0.6589$$

$$\hat{y} = a + bx, \qquad a = \bar{y} - b\bar{x}, \qquad b = \frac{S_{xy}}{S_{xx}}$$

$$b = \frac{-26.1584}{0.597234} = -43.7993$$

$$a = 16.03333 - -43.7993 \times 0.11552 = 21.09303$$

$$\hat{y} = 21.09303 - 43.7993x$$



Example

The number of components used by company per month is recorded in the table below:

Month	0	1	2	3	4	5	6	7	8
No. of components	1520	2806	3210	18630	19800	32560	65495	150457	245904

Find the least squares regression equation and the correlation coefficient.

Key points

- Correlation, $-1 \leq r \geq 1$, only measures the linear relationship between two variables.
- Regression analysis steps
  - Plot the data as a scatter diagram
  - Decide whether a transformation is needed-plots of the transformed data may help
  - Compute the linear regression equation using whichever variable seem linearly related
  - Compute the residuals and plot them as a scatter diagram
  - Are the residuals a random scatter? If not, consider a transformation and go back to step three.
  - Compute R squared.
  - Report or use the equation.

Process control using Mean and Range charts

Some random variation occurs in any seemingly identical but repeated process. If a particular property of the process or its product is monitored regularly, some variation of the results may be due to the measurement and some may result from the process itself. The main reason for checking the process regularly is so that any problems can be identified quickly.

To control the process, it is necessary to keep a check on the current state of the accuracy (mean) and precision (spread) of the distribution of at least one critical factor. This may be achieved with the aid of control charts which have been designed so that a non-specialist can complete them and recognise when there is a problem.

Control charts

Control charts are concerned with two sources of manufacturing or operational variations:
- inherent causes (or common causes) a variation
- assignable courses (or special causes) a variation

The purpose of control charts is to try and identify any assignable causes of variation in the system, so that their cause can be corrected or eliminated.

Most frequently used charts that are used are the Means and Ranges charts which are used together.

Samples should be taken to set up control charts, when it is believed that the process is in statistical control. For the control charts of put into use, it is important to confirm that when the samples taken the process was indeed in statistical control, i.e. the distribution of individual items was reasonably stable.

A control chart consists of:
- points representing a statistic (e.g. a mean, range, proportion) measurements of a quality characteristic in samples taken from the process different times (data)
- the mean of the statistic using all the samples is calculated (e.g. the mean of the means, the mean of the ranges, mean of the proportions)
- a centre line is drawn at the value of the mean of the statistic
- the standard error (e.g. standard deviation/$\sqrt{n}$ for the mean) of the statistic is also calculated using all the samples
- upper and lower control limits (sometimes called “natural process limits”) that indicate the threshold at which the process output is considered statistically ‘unlikely’ and are drawn typically at 3 standard errors from the centre line

The chart may have other optional features, including:
- upper and lower warning limits, drawn as separate lines, typically two standard errors above and below the centre line
- division into zones, with the addition of rules governing frequencies of observations in each zone
- annotation with events of interest, as determined by the Quality Engineer in charge of the process quality.

Suppose we have the data

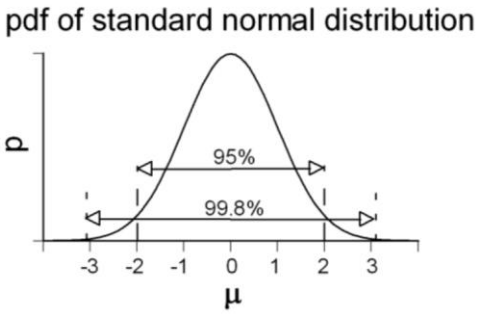
Sample	Measurements	$\bar{x}$	R
1	x_{11} ... x_{1n}	$\bar{x}_1$	R_1
2	x_{21} ... x_{2n}	$\bar{x}_2$	R_2
…	… … …	…	…
k	x_{k1} ... x_{kn}	$\bar{x}_k$	R_k

where $\bar{x}_i = \frac{1}{n}\sum_{j=i}^{n} x_{ij}$ $\quad R_i = \max_j(x_{ij}) - \min_j(x_{ij})$ $\quad 2 \le n \le 12$ and $nk \ge 50$.
When the process is in control, the x_{ij} are assumed independent and $N(\mu, \sigma^2)$. We want to use the data to construct a means chart and arrange chart with

Action limits
set such that the probability of a sample mean (range) being above the upper (below the lower) action limits is ≤ 0.001;

Warning limits
such that the probability of a sample mean (range) being above the upper (though the lower) warning limits is no more than 0.025.

Means chart



***μ* and *σ* known**

Since $x_{ij} \sim N(\mu, \sigma^2)$, $\bar{x}_i = \frac{1}{n}\sum_{j=i}^{n} x_{ij} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$,
$Z = \left(\frac{\bar{x}_i - \mu}{\sigma/_n}\right) \sim N(0,1)$.

Using the standard normal distribution, the action limits are chosen so that $P(Z > A) = P(Z < -A) = 0.001.$ So $A = 3.09.$

For the warning limits, we require that $P(Z > W) = P(Z < -W) = 0.025.$ So $W = 1.96$
The simpler values of $W = 2$ and $A = 3$ are used in many applications.

Now we can evaluate the action and warning limits for the means of our data:

For the actual limits (UAL and LAL) $\quad \frac{\bar{x}_A - \mu}{\sigma/_n} = \pm 3.09$
which we can arrange to give $\text{LAL} = \bar{x}_{LA} = \mu - 3.09\frac{\sigma}{\sqrt{n}}$ and $\text{UAL} = \bar{x}_{UA} = \mu + 3.09\frac{\sigma}{\sqrt{n}}$.

Similarly for the warning limits (UWL and LWL) $\frac{\bar{x}_W - \mu}{\sigma/_n} = \pm 1.96$
and hence $\quad \text{LWL} = \bar{x}_{LW} = \mu - 1.96\frac{\sigma}{\sqrt{n}}$ and $\text{UWL} = \bar{x}_{UW} = \mu + 1.96\frac{\sigma}{\sqrt{n}}$.

So if μ and σ are known, these limits can be used to construct the means chart.

***μ* and *σ* not known**

If we cannot use past experience, so that μ and σ are effectively known, we use the Process or Grand mean as an estimate:

$$\hat{\mu} = \bar{\bar{x}} = \frac{1}{k}\sum_{i=1}^{k} \bar{x}_i\,, \qquad \text{where } k \text{ is the number of samples of size } n.$$

The same idea cannot be applied to estimating σ. Instead we rely on work that has related to the population standard deviation to the range of a sample. This approximation is easy to compute without a scientific calculator:

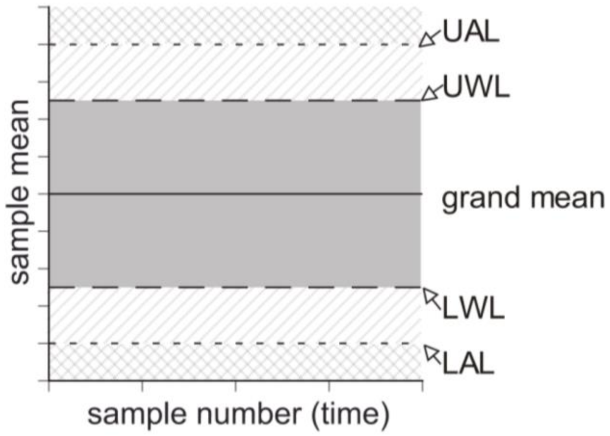
$$\hat{\sigma} = \frac{\bar{R}}{d_n}, \qquad \text{where } \bar{R} = \frac{1}{k}\sum_{i=1}^{k} R_i$$

and d_n is the Hartley's conversion constant, which is tabulated. The range method for estimating $\hat{\sigma}$ is applicable for $2 \le n \le 12$. For $n = 4$ or 5, often used, it is entirely satisfactory. So when μ and σ are unknown the action and warning lines are placed at

$$\bar{\bar{x}} \pm 3.09\frac{\bar{R}}{d_n\sqrt{n}} \qquad \text{and} \qquad \bar{\bar{x}} \pm 1.96\frac{\bar{R}}{d_n\sqrt{n}} \qquad \text{respectively.}$$

Drawing the chart

To plot these on the means chart we turn our usual normal stand distribution around thus



Our chart has 3 zones

- zone 1-stable-solid grey
- zone 2-warning-hashed
- zone 3-action-diamonds

Range chart

The control limits on the range chart are asymmetrical about the mean range since the distribution of sample ranges is positively skewed.
The formulae the setting the action, warning and mean lines on the range chart are:

- Upper action line at: $D_{0.001}\bar{R}$
- Upper warning lines at: $D_{0.025}\bar{R}$
- Process or grand mean at: $\bar{R}$
- Lower warning line at: $D_{0.975}\bar{R}$
- Lower action line at: $D_{0.999}\bar{R}$

where the $D_{-\,-}$ are given in the table of Hartley's constants.

<u>Example</u>

The mean lengths of 25 samples of size 4, of steel rods, taken in chronological order, are shown below. Plot the means and calculate and plot the action and warning limits for the means chart for this process in the following two cases. Compute and plot the mean and limits for the range chart.

a) $x_{ij} \sim N(150{,}16)$
b) μ and σ are unknown

Sample number	Sample mean (mm)	Sample range (mm)
1	147.50	10
2	147.00	19
3	144.75	13
4	150.00	8
5	155.50	4
6	149.75	12
7	146.25	18
8	148.00	14
9	153.25	9
10	149.75	9
11	152.00	4
12	150.00	10
13	148.00	10
14	150.75	16
15	150.25	11
16	147.00	5
17	151.50	7
18	151.75	9
19	152.00	6
20	149.25	13
21	152.50	14
22	152.75	11
23	150.25	13
24	152.25	8
25	150.50	17

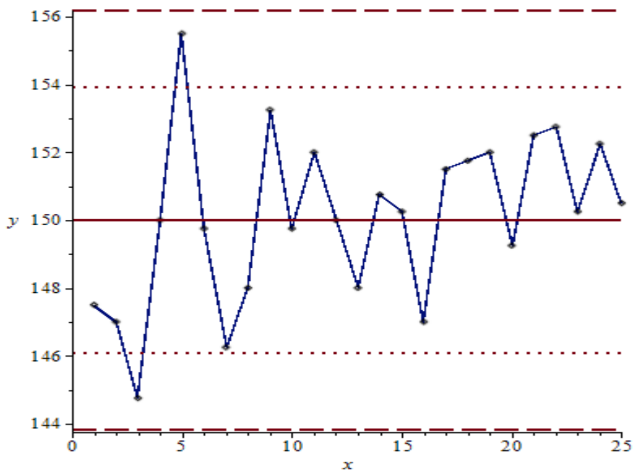
Means chart:

UAL: $150 + 3.09 \times \frac{4}{\sqrt{4}} = 150 + 3.09 \times \frac{4}{2} = 150 + 3.09 \times 2 = 150 + 6.18 = 156.18$

UWL: $150 + 1.96 \times \frac{4}{\sqrt{4}} = 150 + 1.96 \times \frac{4}{2} = 150 + 1.96 \times 2 = 150 + 3.92 = 153.92$

LWL: $150 - 1.96 \times \frac{4}{\sqrt{4}} = 150 - 1.96 \times \frac{4}{2} = 150 - 1.96 \times 2 = 150 - 3.92 = 146.08$

LAL: $150 - 3.09 \times \frac{4}{\sqrt{4}} = 150 - 3.09 \times \frac{4}{2} = 150 - 3.09 \times 2 = 150 - 6.18 = 143.82$



Process mean:

$$\bar{\bar{x}} = \frac{1}{25} \times \sum_{i=1}^{25} \bar{x}_i\,, \qquad \sum_{i=1}^{25} \bar{x}_i = 3752.5, \qquad \bar{\bar{x}} = \frac{1}{25} \times 3752.5 = 150.1$$

Mean range:

$$\bar{R} = \frac{1}{25} \times \sum_{i=1}^{25} R_i\,, \qquad \sum_{i=1}^{25} R_i = 270, \qquad \bar{R} = \frac{1}{25} \times 270 = 10.8$$

UAL: $150.1 + 3.09 \times \frac{10.8}{2.059\sqrt{4}} = 150.1 + 3.09 \times \frac{10.8}{4.118} = 105.1 + 3.09 \times 2.622632346$

$= 150.1 + 8.103933949 = 158.2039339 \approx 158.20$

UWL:$150.1 + 1.96 \times \frac{10.8}{2.059\sqrt{4}} = 150.1 + 1.96 \times \frac{10.8}{4.118} = 105.1 + 1.96 \times 2.622632346$

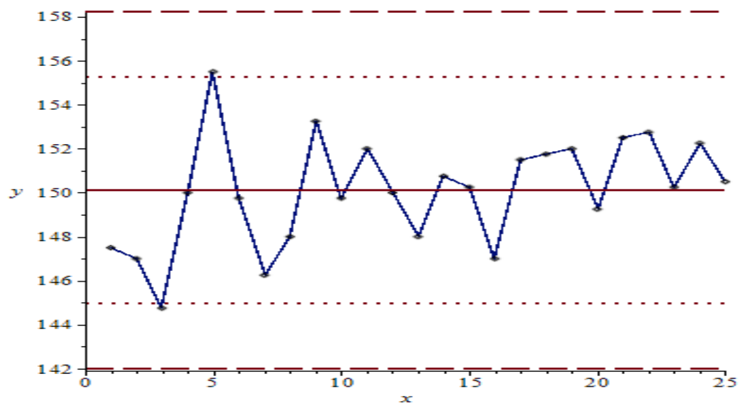
$= 150.1 + 5.140359398 = 155.2403594 \approx 155.24$

LWL:$150.1 - 1.96 \times \frac{10.8}{2.059\sqrt{4}} = 150.1 - 1.96 \times \frac{10.8}{4.118} = 105.1 - 1.96 \times 2.622632346$

$= 150.1 - 5.140359398 = 144.9596406 \approx 144.96$

LAL:$150.1 - 3.09 \times \frac{10.8}{2.059\sqrt{4}} = 150.1 - 3.09 \times \frac{10.8}{4.118} = 105.1 - 3.09 \times 2.622632346$

$= 150.1 - 8.103933949 = 141.9960661 \approx 142.00$



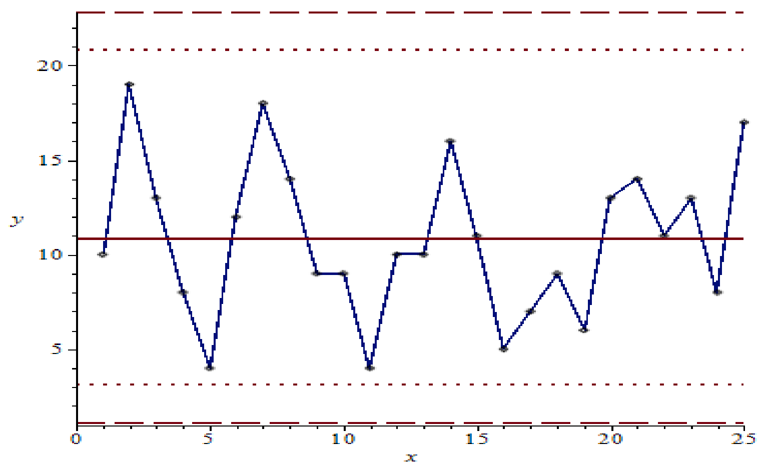
Range chart:

UAL: $2.57 \times 10.8 = 22.756 \approx 22.8$

LWL: $0.29 \times 10.8 = 3.132 \approx 3.1$

UWL: $1.93 \times 10.8 = 20.844 \approx 20.8$

LAL: $0.1 \times 10.8 = 1.08 \approx 1.1$



Example

In the production of transistor amplifiers the output voltage due to a fixed input voltage of 200mV is used for quality control purposes. Twenty samples of size 5 have been taken at regular intervals in order to create Shewhart control charts. The means and range of these sample voltages are

Sample no.	1	2	3	4	5	6	7	8	9	10
$\bar{x}$	0.89	0.69	0.74	0.85	0.77	0.78	0.87	0.71	0.81	0.69
R	0.19	0.28	0.25	0.38	0.36	0.43	0.76	0.16	0.37	0.29

Sample no.	11	12	13	14	15	16	17	18	19	20
$\bar{x}$	0.75	0.89	0.90	0.84	0.79	0.69	0.78	0.81	0.87	0.88
R	0.31	0.24	0.16	0.46	0.29	0.43	0.17	0.34	0.29	0.24

Use these results to evaluate grand means and control limits for the mean and range charts and plot these limits.

Process mean:

$$\bar{\bar{x}} = \frac{1}{20} \times \sum_{i=1}^{20} \bar{x}_i\,, \qquad \sum_{i=1}^{20} \bar{x}_i = 16, \qquad \bar{\bar{x}} = \frac{1}{20} \times 16 = 0.8$$

Mean range:

$$\bar{R} = \frac{1}{20} \times \sum_{i=1}^{20} R_i\,, \qquad \sum_{i=1}^{20} R_i = 6.4, \qquad \bar{R} = \frac{1}{20} \times 6.4 = 0.32$$

Are we in control?

We have stated that

- sample should be taken to set up control charts, when it is believed that the process is in statistical control;
- before the control charts are put into use for the process capability is assessed, it is important to confirm that when the samples were taken the process was indeed in statistical control.

Assessing the state of control

A process is said to be in statistical control when all the variations have been shown to have random or common causes.

The randomness of the variations is best illustrated by

- collecting at least 50 observations of data and grouping them into samples (or sets) of at least 4 observations;
- presenting the results in the form of both mean and range control charts-the limits of which are either worked out from the data may be set by some predetermined parameters.

If the process in which the data was collected is in statistical control there will not be

- mean or range values which lie outside the action limits (zone 3);
- more than about 1 in 40 values between the warning and action limits (zone 2);
- two consecutive mean or range values outside the same warning limit (zone 2);
- a run or trend of 5 or more values which also infringes a warning or action limit (zone 2/3);
- a run or more of 6 means which lie either above or below the ground mean (zone 1);
- trends of more than 6 sample means which are either raising or falling (zone 1).

(These are relatively conservative rules. Others use more points in a run or trend and some control the number of consecutive points slating about the grand mean.)

If the process is out of control, special courses must be identified and eliminated. The process can then be re-examined to see if it is in statistical control.

Once any special courses of variation in identified and eliminated,

- either another set of samples from the process is taken and the control chart limits recalculated,
- or approximate control chart limits are recalculated by simply excluding the out of control results which special cases have been found and corrected. By extending the points affected by known causes of variation, we have a better estimate of variation due to common causes only.

If the process is shown to be in statistical control, the next task is to compare the limits of this control with the tolerance sought.

<u>**Example**</u>

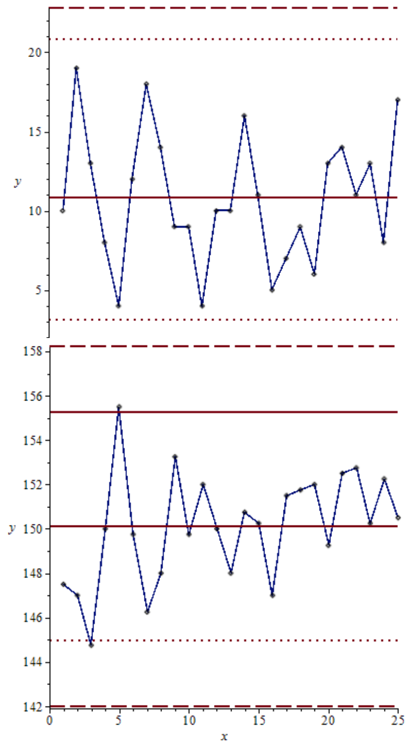
Using the data for the steel rod example, assess the state of control.

We examine the range chart first, because the range mean determines the positions of the limits on this chart and it also affects those on the mean chart.
The range is in control, all the points lie inside the warning limits, which means that the spread of the distribution remained constant. The process is in control with respect to the range or spread.

Examining the mean chart.

There are two points which fall in the warning zone:

- they are not consecutive nor in the same one,
- of the total points plotted on the charts, we are expecting 1 in 40 to be in each warning zone when the process is stable,
- there are not 40 results available, and we have to make a decision.



In this example, it is reasonable to assume that the two points in the warning zones have arisen from the random variation of the process and do not indicate an out of control situation.
There are no runs or trends of six or more points on the charts, so the process is judged to be statistical control.
Therefore, the mean and range charts may now be used to control the process.

<u>Example</u>

Using the control charts produced for the voltage of the transistor, assess the state of control.

Tolerance limits

A clear distinction must be made between the tolerance limits set down in the product specification and limits on control charts.

- Tolerance limits: based on the functional requirements of the products.
- Control chart limits: based on the stability and capability of the process.

The process may be unable to meet specification requirements, but still in a state of statistical control.

A comparison of processes capability and tolerance can only take place when it has been established that the process is in statistical control.

A quantitative assessment of capability with respect to the specified requirements can be found in Oakland (1996): Statistical process control, chapter 10.

Do we continue to stay in control?

When the process has been shown to be in control, the mean and range chart may be used to make decisions about the state of the process during further operation.
As each sample is taken, its mean and range is calculated in each of these is plotted on the appropriate chart. When adding new samples to the chart the original limits are used, we **do not** recalculate them.

If both points in zone 1

- the process appears to have remained stable
    - check that this mean is not the sixth in a run or trend-if okay that process continue.

If either the points plot in zone 3

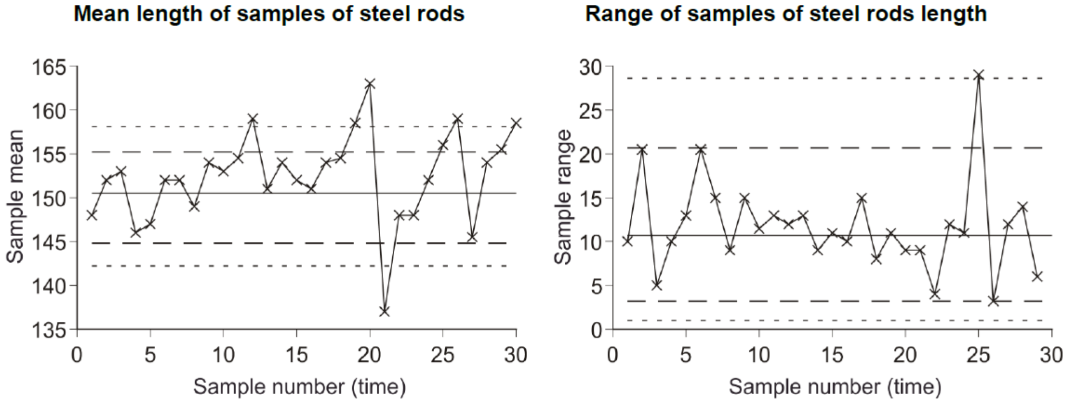
- the process should be investigated and action taken. The latest estimate of the mean and its difference from the original process mean or target value should be used to assess the size of any correction.

A point in zone 2

- suggest that there may have been an assignable change and that another sample must be taken in order to check whether the new samples mean and range
    - lie in zone 1-no further action is needed. Let the process continue;
    - lie in zone 2/3-take action now to find and rectify the problem.

<u>Example</u>

Continuing the steel rod example to show how the charts are used. The plots below show the mean and range charts for the next 30 samples taken from the steel cutting process.



In practice, the analysis of each chart would be as each point is added.

The process is well under control, i.e. all points within the action lines with no long runs or trends in the means, up to and including sample 11.

Sample 12 shows that corrective action must be taken.

Sample 18 is the sixth sample in a run above the mean, since the corrective action was taken at sample 12. Another sample should be taken immediately, rather than wait for the next sampling period.

The mean of sample 19 is in the action zone, as well as being the 7^{th} point in a run above the mean. Corrective action must be taken and another sample to be taken to check it.

After taking action, sample 20 gives a mean well above the action line, indicating that the corrective action has caused the process to move in the wrong direction. A correction of opposite sense to that just made must be applied and check a new sample.

The action following sample 20 results in over-correction and sample mean 21 is below the lower action line. Take another sample immediately.

The sample mean drifts upwards between samples 22 and 26 while the sample range becomes more variable and that for sample 25 exceeds the upper action limit. Make a correction and take another sample immediately.

The range of sample 26 suggests that the correction just made has been effective but now, the mean is in the action zone again. Make another correction and sample immediately.

This brings the sample mean back within zone 1 but the means of the next 3 samples rise until the upper action limit is infringed again at sample 30. There appears to be a problem. Production should stop until its cause has been found and rectified.

The process equipment was investigated as a result of this- a worn adjustment screw was slowly and continually vibrating open, allowing an increase speed of rod through the cutting machine.

This situation would not been identified as quickly in the absence of the process control charts. This simple example illustrates the power of control charts in both process control and in early warning of equipment trouble.

Example

The resistance (ohms) of twenty samples of size 4 are given in the table below which were taken after the test circuit had not been used for 48 hours.

$\bar{x}$	7.6	6.5	8.5	9.8	6.3	8.5	6.7	7.1	8.1	10.7
R	22	9	23	9	9	6	13	8	4	6

$\bar{x}$	9.7	7.8	9.5	8.7	9.5	8.9	11.5	7.7	9.1	11.0
R	7	10	7	14	5	13	8	11	4	7

Construct and interpret the mean and range charts of this data.
It is known that the results for the first few samples may be unreliable as the temperature of the circuit may not have reached equilibrium. Remove samples up to and including any suspect samples and refit the control limits.
Following the removal of the initial unreliable samples, is the process in control?

Key Points

In this short introduction to control charts, we have learnt how to construct and apply simple means and ranges to charts. The limits we are using are

		Means	
Limit	Ranges	μ and σ known	μ and σ unknown
UAL	$D_{0.001}\bar{R}$	$\mu + 3.09\frac{\sigma}{\sqrt{n}}$	$\bar{\bar{x}} + 3.09\frac{\bar{R}}{d_n\sqrt{n}}$
UWL	$D_{0.025}\bar{R}$	$\mu + 1.96\frac{\sigma}{\sqrt{n}}$	$\bar{\bar{x}} + 1.96\frac{\bar{R}}{d_n\sqrt{n}}$
Grand mean	$\bar{R}$	μ	$\bar{\bar{x}}$
LWL	$D_{0.975}\bar{R}$	$\mu - 1.96\frac{\sigma}{\sqrt{n}}$	$\bar{\bar{x}} - 1.96\frac{\bar{R}}{d_n\sqrt{n}}$
LAL	$D_{0.999}\bar{R}$	$\mu - 3.09\frac{\sigma}{\sqrt{n}}$	$\bar{\bar{x}} - 3.09\frac{\bar{R}}{d_n\sqrt{n}}$

Other charts are in common use. These include

- average and standard deviation ($n > 12$)
- cumulative sum
- moving averages and ranges
- exponentially weighted and moving averages

The rules that we are using to assess lack of control are

- mean or range values which lie outside the action limits;
- more than about 1 in 40 values between the warning and action limits;
- two consecutive mean or range values outside the same warning limit;
- a run or trend of 5 or more values which also infringes a warning or action limit;
- a run of more than 6 means which lie either above or below the grand mean;
- trends of more than 6 sample means which are either rising or falling.

Cumulative Standard Normal Distribution

$$P(x > z^*) = \frac{1}{\sqrt{2\pi}} \int_{z^*}^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx = \alpha$$, where α is the confidence level

z	P	z	P	z	P	z	P	z	P	z	P
0.00	0.5000										
0.01	0.4960	0.31	0.3783	0.61	0.2709	0.91	0.1814	1.21	0.1131	1.51	0.0655
0.02	0.4920	0.32	0.3745	0.62	0.2676	0.92	0.1788	1.22	0.1112	1.52	0.0643
0.03	0.4880	0.33	0.3707	0.63	0.2643	0.93	0.1762	1.23	0.1093	1.53	0.0630
0.04	0.4840	0.34	0.3669	0.64	0.2611	0.94	0.1736	1.24	0.1075	1.54	0.0618
0.05	0.4801	0.35	0.3632	0.65	0.2578	0.95	0.1711	1.25	0.1056	1.55	0.0606
0.06	0.4761	0.36	0.3594	0.66	0.2546	0.96	0.1685	1.26	0.1038	1.56	0.0594
0.07	0.4721	0.37	0.3557	0.67	0.2514	0.97	0.1660	1.27	0.1020	1.57	0.0582
0.08	0.4681	0.38	0.3520	0.68	0.2483	0.98	0.1635	1.28	0.1003	1.58	0.0571
0.09	0.4641	0.39	0.3483	0.69	0.2451	0.99	0.1611	1.29	0.0985	1.59	0.0559
0.10	0.4602	0.40	0.3446	0.70	0.2420	1.00	0.1587	1.30	0.0968	1.60	0.0548
0.11	0.4562	0.41	0.3409	0.71	0.2389	1.01	0.1562	1.31	0.0951	1.61	0.0537
0.12	0.4522	0.42	0.3372	0.72	0.2358	1.02	0.1539	1.32	0.0934	1.62	0.0526
0.13	0.4483	0.43	0.3336	0.73	0.2327	1.03	0.1515	1.33	0.0918	1.63	0.0516
0.14	0.4443	0.44	0.3300	0.74	0.2296	1.04	0.1492	1.34	0.0901	1.64	0.0505
0.15	0.4404	0.45	0.3264	0.75	0.2266	1.05	0.1469	1.35	0.0885	1.65	0.0495
0.16	0.4364	0.46	0.3228	0.76	0.2236	1.06	0.1446	1.36	0.0869	1.66	0.0485
0.17	0.4325	0.47	0.3192	0.77	0.2206	1.07	0.1423	1.37	0.0853	1.67	0.0475
0.18	0.4286	0.48	0.3156	0.78	0.2177	1.08	0.1401	1.38	0.0838	1.68	0.0465
0.19	0.4247	0.49	0.3121	0.79	0.2148	1.09	0.1379	1.39	0.0823	1.69	0.0455
0.20	0.4207	0.50	0.3085	0.80	0.2119	1.10	0.1357	1.40	0.0808	1.70	0.0446
0.21	0.4168	0.51	0.3050	0.81	0.2090	1.11	0.1335	1.41	0.0793	1.71	0.0436
0.22	0.4129	0.52	0.3015	0.82	0.2061	1.12	0.1314	1.42	0.0778	1.72	0.0427
0.23	0.4090	0.53	0.2981	0.83	0.2033	1.13	0.1292	1.43	0.0764	1.73	0.0418
0.24	0.4052	0.54	0.2946	0.84	0.2005	1.14	0.1271	1.44	0.0749	1.74	0.0409
0.25	0.4013	0.55	0.2912	0.85	0.1977	1.15	0.1251	1.45	0.0735	1.75	0.0401
0.26	0.3974	0.56	0.2877	0.86	0.1949	1.16	0.1230	1.46	0.0721	1.76	0.0392
0.27	0.3936	0.57	0.2843	0.87	0.1922	1.17	0.1210	1.47	0.0708	1.77	0.0384
0.28	0.3897	0.58	0.2810	0.88	0.1894	1.18	0.1190	1.48	0.0694	1.78	0.0375
0.29	0.3859	0.59	0.2776	0.89	0.1867	1.19	0.1170	1.49	0.0681	1.79	0.0367
0.30	0.3821	0.60	0.2743	0.90	0.1841	1.20	0.1151	1.50	0.0668	1.80	0.0359

z	P	z	P	z	P	z	P	z	P	z	P
1.81	0.0351	2.11	0.0174	2.41	0.0080	2.71	0.0034	3.01	0.0013	3.31	0.0005
1.82	0.0344	2.12	0.0170	2.42	0.0078	2.72	0.0033	3.02	0.0013	3.32	0.0005
1.83	0.0336	2.13	0.0166	2.43	0.0075	2.73	0.0032	3.03	0.0012	3.33	0.0004
1.84	0.0329	2.14	0.0162	2.44	0.0073	2.74	0.0031	3.04	0.0012	3.34	0.0004
1.85	0.0322	2.15	0.0158	2.45	0.0071	2.75	0.0030	3.05	0.0011	3.35	0.0004
1.86	0.0314	2.16	0.0154	2.46	0.0069	2.76	0.0029	3.06	0.0011	3.36	0.0004
1.87	0.0307	2.17	0.0150	2.47	0.0068	2.77	0.0028	3.07	0.0011	3.37	0.0004
1.88	0.0301	2.18	0.0146	2.48	0.0066	2.78	0.0027	3.08	0.0010	3.38	0.0004
1.89	0.0294	2.19	0.0143	2.49	0.0064	2.79	0.0026	3.09	0.0010	3.39	0.0003
1.90	0.0287	2.20	0.0139	2.50	0.0062	2.80	0.0026	3.10	0.0010	3.40	0.0003
1.91	0.0281	2.21	0.0136	2.51	0.0060	2.81	0.0025	3.11	0.0009	3.41	0.0003
1.92	0.0274	2.22	0.0132	2.52	0.0059	2.82	0.0024	3.12	0.0009	3.42	0.0003
1.93	0.0268	2.23	0.0129	2.53	0.0057	2.83	0.0023	3.13	0.0009	3.43	0.0003
1.94	0.0262	2.24	0.0125	2.54	0.0055	2.84	0.0023	3.14	0.0008	3.44	0.0003
1.95	0.0256	2.25	0.0122	2.55	0.0054	2.85	0.0022	3.15	0.0008	3.45	0.0003
1.96	0.0250	2.26	0.0119	2.56	0.0052	2.86	0.0021	3.16	0.0008	3.46	0.0003
1.97	0.0244	2.27	0.0116	2.57	0.0051	2.87	0.0021	3.17	0.0008	3.47	0.0003
1.98	0.0239	2.28	0.0113	2.58	0.0050	2.88	0.0020	3.18	0.0007	3.48	0.0003
1.99	0.0233	2.29	0.0110	2.59	0.0048	2.89	0.0019	3.19	0.0007	3.49	0.0002
2.00	0.0228	2.30	0.0107	2.60	0.0047	2.90	0.0019	3.20	0.0007	3.50	0.0002
2.01	0.0222	2.31	0.0104	2.61	0.0045	2.91	0.0018	3.21	0.0007	3.51	0.0002
2.02	0.0217	2.32	0.0102	2.62	0.0044	2.92	0.0018	3.22	0.0006	3.52	0.0002
2.03	0.0212	2.33	0.0099	2.63	0.0043	2.93	0.0017	3.23	0.0006	3.53	0.0002
2.04	0.0207	2.34	0.0096	2.64	0.0041	2.94	0.0016	3.24	0.0006	3.54	0.0002
2.05	0.0202	2.35	0.0094	2.65	0.0040	2.95	0.0016	3.25	0.0006	3.55	0.0002
2.06	0.0197	2.36	0.0091	2.66	0.0039	2.96	0.0015	3.26	0.0006	3.56	0.0002
2.07	0.0192	2.37	0.0089	2.67	0.0038	2.97	0.0015	3.27	0.0005	3.57	0.0002
2.08	0.0188	2.38	0.0087	2.68	0.0037	2.98	0.0014	3.28	0.0005	3.58	0.0002
2.09	0.0183	2.39	0.0084	2.69	0.0036	2.99	0.0014	3.29	0.0005	3.59	0.0002
2.10	0.0179	2.40	0.0082	2.70	0.0035	3.00	0.0013	3.30	0.0005	3.60	0.0002

Hartley's Constants

Sample size n	d_n	$D_{0.999}$	$D_{0.975}$	$D_{0.025}$	$D_{0.001}$
2	1.128	0	0.04	2.81	4.12
3	1.693	0.04	0.18	2.17	2.98
4	2.059	0.1	0.29	1.93	2.57
5	2.326	0.16	0.37	1.81	2.34
6	2.534	0.21	0.42	1.72	2.21
7	2.704	0.26	0.46	1.66	2.11
8	2.847	0.29	0.5	1.62	2.04
9	2.97	0.32	0.52	1.58	1.99
10	3.078	0.35	0.54	1.56	1.93
11	3.173	0.38	0.56	1.53	1.91
12	3.258	0.4	0.58	1.51	1.87

$\boldsymbol{\Gamma(x)}$

x	Gamma		x	Gamma		x	Gamma		x	Gamma		x	Gamma
1.01	0.9943		1.31	0.8960		1.61	0.8947		1.91	0.9652		2.21	1.1078
1.02	0.9888		1.32	0.8946		1.62	0.8959		1.92	0.9688		2.22	1.1140
1.03	0.9835		1.33	0.8934		1.63	0.8972		1.93	0.9724		2.23	1.1202
1.04	0.9784		1.34	0.8922		1.64	0.8986		1.94	0.9761		2.24	1.1266
1.05	0.9735		1.35	0.8912		1.65	0.9001		1.95	0.9799		2.25	1.1330
1.06	0.9687		1.36	0.8902		1.66	0.9017		1.96	0.9837		2.26	1.1395
1.07	0.9642		1.37	0.8893		1.67	0.9033		1.97	0.9877		2.27	1.1462
1.08	0.9597		1.38	0.8885		1.68	0.9050		1.98	0.9917		2.28	1.1529
1.09	0.9555		1.39	0.8879		1.69	0.9068		1.99	0.9958		2.29	1.1598
1.1	0.9514		1.4	0.8873		1.7	0.9086		2	1.0000		2.3	1.1667
1.11	0.9474		1.41	0.8868		1.71	0.9106		2.01	1.0043		2.31	1.1738
1.12	0.9436		1.42	0.8864		1.72	0.9126		2.02	1.0086		2.32	1.1809
1.13	0.9399		1.43	0.8860		1.73	0.9147		2.03	1.0131		2.33	1.1882
1.14	0.9364		1.44	0.8858		1.74	0.9168		2.04	1.0176		2.34	1.1956
1.15	0.9330		1.45	0.8857		1.75	0.9191		2.05	1.0222		2.35	1.2031
1.16	0.9298		1.46	0.8856		1.76	0.9214		2.06	1.0269		2.36	1.2107
1.17	0.9267		1.47	0.8856		1.77	0.9238		2.07	1.0316		2.37	1.2184
1.18	0.9237		1.48	0.8857		1.78	0.9262		2.08	1.0365		2.38	1.2262
1.19	0.9209		1.49	0.8859		1.79	0.9288		2.09	1.0415		2.39	1.2341
1.2	0.9182		1.5	0.8862		1.8	0.9314		2.1	1.0465		2.4	1.2422
1.21	0.9156		1.51	0.8866		1.81	0.9341		2.11	1.0516		2.41	1.2503
1.22	0.9131		1.52	0.8870		1.82	0.9368		2.12	1.0568		2.42	1.2586
1.23	0.9108		1.53	0.8876		1.83	0.9397		2.13	1.0621		2.43	1.2670
1.24	0.9085		1.54	0.8882		1.84	0.9426		2.14	1.0675		2.44	1.2756
1.25	0.9064		1.55	0.8889		1.85	0.9456		2.15	1.0730		2.45	1.2842
1.26	0.0944		1.56	0.8896		1.86	0.9487		2.16	1.0786		2.46	1.2930
1.27	0.9025		1.57	0.8905		1.87	0.9518		2.17	1.0842		2.47	1.3019
1.28	0.9007		1.58	0.8914		1.88	0.9551		2.18	1.0900		2.48	1.3109
1.29	0.8990		1.59	0.8924		1.89	0.9584		2.19	1.0959		2.49	1.3201
1.3	0.8975		1.6	0.8935		1.9	0.9618		2.2	1.1018		2.5	1.3293

t-distribution table

Critical values of the t-distribution are given for areas α in one tail.

α	0.100	0.050	0.025	0.010	0.005	0.001	0.0005
DF							
1	3.078	6.314	12.706	31.821	63.657	318.30	636.619
2	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.4150	1.8950	2.3650	2.998	3.499	4.7850	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.3500	1.771	2.1600	2.6500	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.768
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.0600	2.4850	2.787	3.4500	3.72500
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
50	1.299	1.676	2.009	2.403	2.678	3.261	3.496
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
80	1.292	1.664	1.990	2.374	2.639	3.195	3.416
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
infinity	1.282	1.645	1.960	2.326	2.576	3.091	3.291