3. THE CENTRAL LIMIT THEOREM

Addition of Random Variables

If the continuous random variable, or signals, X and Y have the pdf's $f_X(x)$ and $f_Y(y)$ respectively, it might seem straightforward that we could easily find the pdf of their sum $Z = X + Y$; i.e. $f_Z(z)$ in terms of $f_X(x)$ and $f_Y(y)$. However the 'randomness' of the variables introduces an element, which makes this task quite difficult mathematically. Although the theory is well known and is at the heart of signal processing the general purpose engineer on the other hand really only requires the essential results and these are now listed with appropriate examples and illustrations.

Example 1: Random Numbers

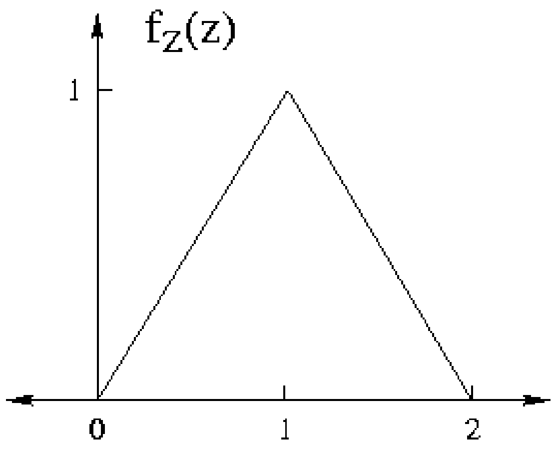
If X and Y are two uniformly distributed random numbers in the range [0, 1] then

$$f_X(x) = 1, \; 0 \le X \le 1$$
$$= 0, \text{ elsewhere}$$

and likewise for $f_Y(y)$.

X and Y thus have a Uniform Distribution over [0, 1], and their sum $Z = X + Y$ clearly satisfies $0 \le X \le 2$ but:

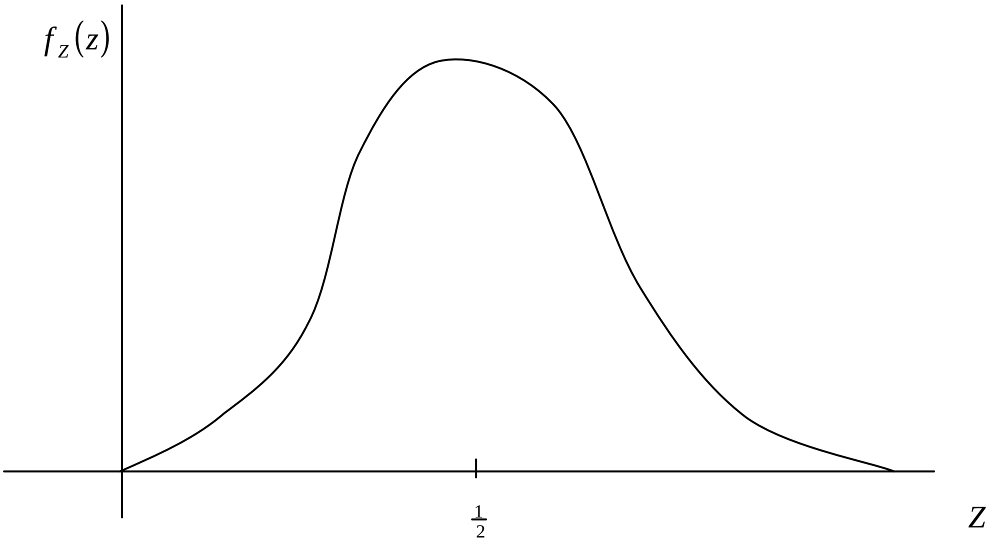
$$f_Z(z) = z \qquad 0 \le Z \le 1$$
$$= 2 - z \qquad 1 \le Z \le 2$$



The 'broken back' distribution for $f_Z(z)$ may seem more credible if we scan down tables of random numbers and see that any two in the sequence when taken in either order (e.g. to be added) are twice as likely to be on different sides of 0.5 than on the same side.

Mean of n Random Numbers

If $Z = \frac{X_1 + X_2 + X_3 + ... + X_n}{n}$, then Z, like all the X_i will satisfy $0 \le Z \le 1$, but intuitively will concentrate close to $Z = \frac{1}{2}$ and have a distribution which is symmetrical about $Z = \frac{1}{2}$.



It can be proved that if n is large $f_Z(z)$ approaches a Normal distribution with mean equal to $\frac{1}{2}$. Furthermore this is a special application of a more general result.

The Central Limit Theorem

If $X_1, X_2 ... X_n$ all have the same probability density function (pdf) then the pdf of $\overline{X} = \frac{1}{n}(X_1 + X_2 + ... + X_n)$ will, under certain conditions, approach Normality as n becomes large; the mean of the distribution of $\overline{X}$ being the same as that of any X_i.

Mean and Variance of the Sum of Random Variables

Whatever the random variables X and Y, it can be shown that if $Z = X + Y$:

$$E[Z] = E[X] + E[Y]$$

i.e. the sum of the means of two random variables equals the mean of the sum of the two random variables.
If and only if X and Y are independent, then if $Z = X + Y$:

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

i.e. the sum of the variances of two random variables equals the variance of the sum of the two random variables. This result means that variances are of more statistical interest than standard deviations.

Knowing that $E[aX] = aE[X]$ and $\mathrm{var}[aX] = a^2\,\mathrm{var}[X]$ we can extend the above results to n variables to show that if $X_1, X_2 \ldots X_n$ all have the same pdf with mean m and variance s^2 then:

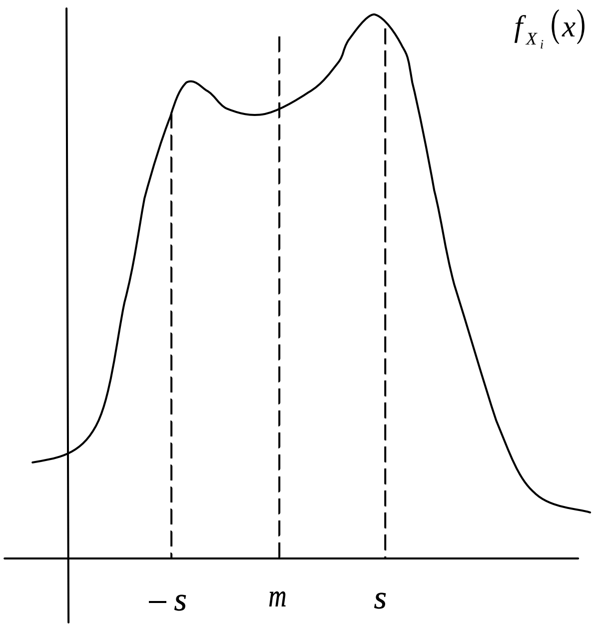
$$E[\overline{X}] = m \qquad \text{and}$$

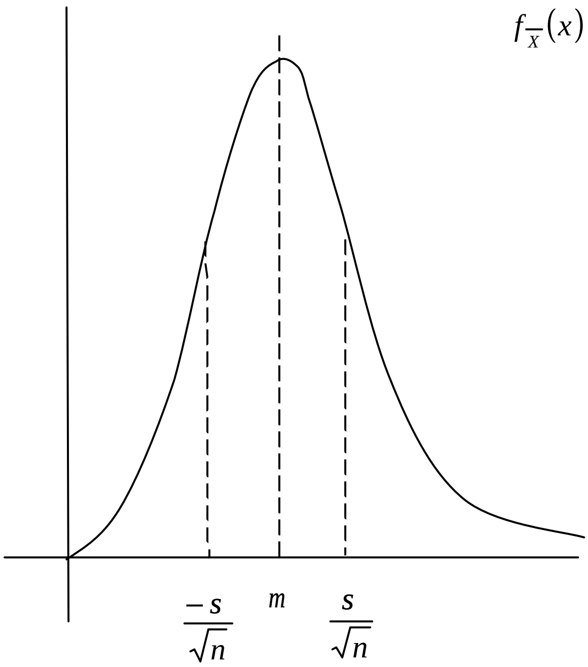
$$\mathrm{var}[\overline{X}] = \frac{s^2}{n}$$

The Central Limit Theorem tells us further that:

$$\overline{X} \sim N\left(m, \frac{s^2}{n}\right)$$

Or in other words if each X_i is an observation from a population of mean m and variance s^2, *not necessarily Normal*, then if n is large, the mean, $\overline{X}$, of a sample of size n will be distributed Normally about the population mean and with $\frac{1}{\sqrt{n}}$ of the population standard deviation, i.e.:





The Normal Distribution

A case of interest is the Normal Distribution which has the special property that if *X* and *Y* are both Normal and independent and:

$X \sim \mathrm{N}(m_x,\ s_x^2)$

$Y \sim \mathrm{N}(m_y,\ s_y^2)$

then:

$X + Y \sim \mathrm{N}(m_x + m_y,\ s_x^2 + s_y^2)$

i.e. the sum of two independent Normal random variables is also Normally distributed. This is not the case for other distributions, e.g. the Uniform distribution.