Module 2

Random Processes

Lesson
6

Some useful Distributions

After reading this lesson, you will learn about

- ***Uniform Distribution***
- ***Binomial Distribution***
- ***Poison Distribution***
- ***Gaussian Distribution***
- ***Central Limit Theorem***
- ***Generation of Gaussian distributed random numbers using computer***
- ***Error function***

Uniform (or Rectangular) Distribution

The pdf f(x) and the cdf F(x) are defined below:

f(x) = 0 for x < a and
for x > (a + b).

However, let, f(x) = 1/b for a < x < (a + b). It is easy to see that,

$$\int_{-\infty}^{+\infty} f(x)dx = \int_{a}^{a+b} \left(\frac{1}{b}\right) dx = 1 \qquad 2.7.1$$

$$\text{The cdf } F(x) = \int_{-\infty}^{x} f(x)dx = \begin{cases} 0 & \text{for x } < \text{ a ;} \\ \dfrac{x-a}{b} & \text{for a } < \text{ x } < (\text{a}+\text{b}) \text{ ;} \\ 1 & \text{for x } > (\text{a}+\text{b}) \text{ ;} \end{cases} \qquad 2.7.2$$

Binomial Distribution

This distribution is associated with discrete random variables. Let 'p' is the probability of an event (say, 'S', denoting success) in a statistical experiment. Then, the probability that this event does not occur (i.e. failure or 'F' occurs) is (1 – p) and, for convenience, let, q = 1 – p, i.e. p + q = 1. Now, let this experiment be conducted repeatedly (without any fatigue or biasness or partiality) 'n' times. Binomial distribution tells us the probability of exactly 'x' successes in 'n' trials of the experiment (x ≤ n). The corresponding binomial probability distribution is:

$$f(x) = \binom{n}{x} p^x q^{n-x} = \left[\frac{n!}{x!(n-x)!}\right] p^x (1-p)^{n-x} \qquad 2.7.3$$

Note that, $\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = 1$

The following general binomial expression and its derivative with respect to a dummy variable 'τ'are useful in verifying the above comment and also for finding expectations of 'S':

$$(p\tau + q)^n = \sum_{x=0}^{n} \binom{n}{x} (p\tau)^x q^{n-x}$$

Differentiating this expression once w.r.t. 'τ', we get,

$$n(p\tau + q)^{n-1} p = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} x \tau^{x-1}$$

The mean number of successes, $E(s) = \sum_{x=0}^{n} x\, f(x) \; = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} x$

A simple form of the last term can be obtained by putting τ = 1 in the previous expression resulting in a fairly easy-to-remember mean of the random variable:
The mean number of successes, E (S) = np

Following a similar approach it can be shown that, the variance of the number of successes,

$$\sigma^2 = npq.$$

An approximation

The following approximation of the binomial distribution of 2.7.3, known as Laplace approximation, is good to use when 'n' is large, i.e. n → ∞:

$$f(x) = \{(2\pi npq)^{-1/2}\}.\exp[-\{(x-np)^2\}/(2npq)] \qquad 2.7.4$$

Poisson Distribution

A formal discussion on Poisson distribution usually follows a description of binomial distribution because the Poisson distribution can be approximately viewed as a limiting case of binomial distribution.

The approximate Poisson distribution, $f_p(x)$ is:

$$f_p(x) = \left[\{.\lambda^x\}/x!\right].e^{-\lambda} \qquad 2.7.5$$

'λ' in the above expression is the mean of the distribution. A special feature of this distribution is that the variance of this distribution is approximately the same as its mean (λ).

Poisson distribution plays an important role in traffic modeling and analysis in data networks.

Gaussian (Normal) Distribution:

This is a commonly applied distribution often exhibited by continuous random variables in nature. A relevant example is the weak electrical noise (voltage or equivalently current) generated because of random (Brownian) movement of carriers in electronic devices and components (such as resistor, diode, transistor etc.).

A Gaussian or, normally distributed random variable X has the following probability density function:

$$f_N(x) = [\,1/\{\sigma.\,(2\pi)^{1/2}\}].\,\exp[\,-\,\{(x-m)^2\}/\,(2\sigma^2)\,] \qquad 2.7.6$$

m = E[X] : Mean or expected value of the distribution and

$\sigma^2 = E[\{X - E[X]\}^2]$: the variance of the distribution.

Fig. 2.7.1 shows two Gaussian pdf curves with means 0.0 and 1.5 but same variance 1.0. The particular curve with m = 0.0 and σ^2 = 1.0 is known as the normalized Gaussian pdf. It may be noted that a pdf curve is symmetric about its mean value. The mean may be positive or negative. Further, a change in the mean of a Gaussian pdf curve only shifts the curve horizontally without any change in shape of the curve. A smaller variance, however, increases the sharpness of the peak value of the pdf which always occurs at the average value of the random variable. This explains the significance of 'variance' of a distribution. A smaller variance means that a random variable mostly assumes values close to its expected or mean value.

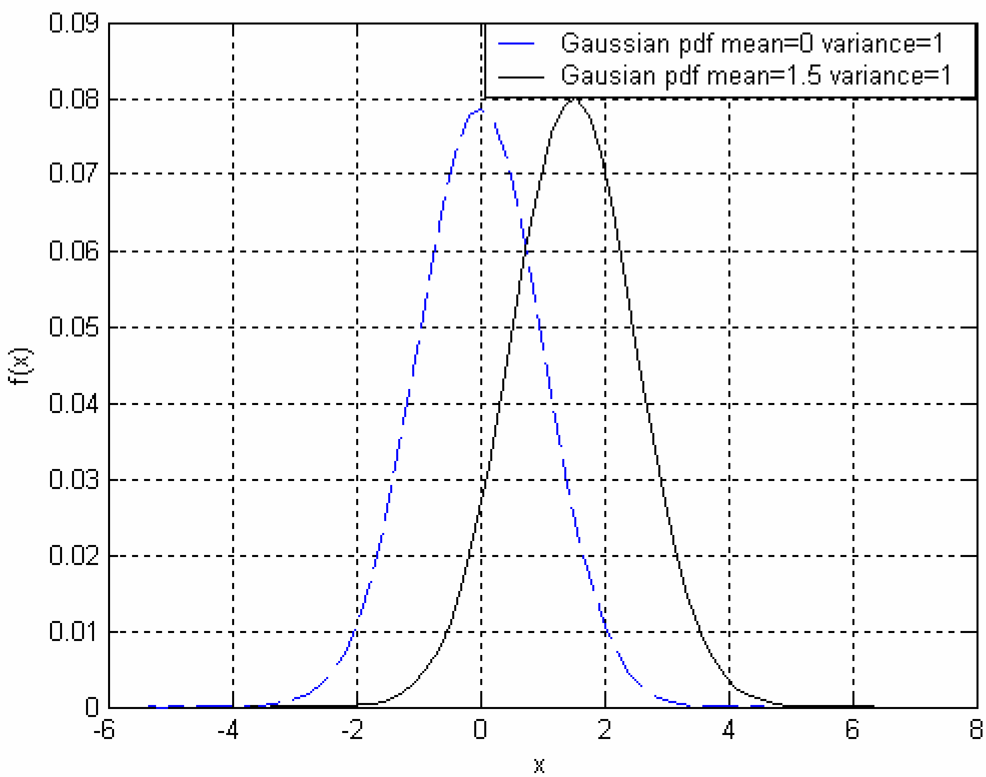

Fig.2.7.1. *Gaussian pdf with mean = 0.0 and variance = 1.0 and Gaussian pdf with mean = 1.5 and variance = 1.0*

It can be shown with some approximation that about 68% of the values (assumed by a Gaussian distributed random variable over a large number of unbiased trials) lie within $\pm \sigma$ around the mean value and about 95% of the values lie within $\pm 2\sigma$ around the mean value.

The sum of mutually independent random variables, each of which is Gaussian distributed, is a Gaussian distributed random variable.

Central Limit Theorem

Let X_1, X2, X_3, ……, X_N denote N mutually independent random variables whose individual distributions are not known and they are not necessarily Gaussian distributed. If m_i , i = 1, 2, ….,N, indicates the mean of the i-th random variable and σ_i , i = 1, 2, ….,N, indicates the variance of the i-th random variable, the central limit theorem, under a few subtle conditions, results in a very powerful and significant inference. The theorem establishes that the sum of the N random variables (say, X) is a random variable which tends to follow a Gaussian (or, normal) distribution as $N \rightarrow \infty$. Further, a) the mean of the sum random variable X is the sum of the mean values of the constituent random variables and b) the variance of the sum random variable X is the sum of the variance values of the constituent random variables X_1, X2, X_3, ……, X_N.

The Central Limit Theorem is very useful in modeling and analyzing several situations in the study of electrical communications. However, one necessary condition to look for before invoking Central Limit theorem is that no single random variable should have significant contribution to the sum random variable.

Generation of Gaussian distributed random numbers using computers

Gaussian distributed random numbers can be generated by generating random numbers having uniform distribution. A simple (but not very good always) method of doing this is by applying central limit theorem. A set of 'n' (usually $n \geq 12$) independent uniformly distributed random numbers are generated and they are summed up and scaled appropriately to result in a Gaussian distributed random variable.

A faster and more precise way for generating Gaussian random variables is known as the Box-Müller method. Only two uniform distributed random variables say, X_1 and X_2, ensure generation of two (almost) independent and identically distributed Gaussian random variables (N_1 and N_2). If x_1 and x_2 are two uncorrelated values assigned to X_1 and X_2 respectively, the two independent Gaussian distributed values n_1 and n_2 are obtained from the following relations:

$$n_1 = \sqrt{[-2.(\ln x_1).\cos 2\pi x_2]}$$
$$n_2 = \sqrt{[-2.(\ln x_2).\sin 2\pi x_1]}$$
2.7.7

Error Functions

A few frequently used functions that are closely related to Gaussian distribution are summarized below. The reader may refer back to this sub-section as necessary.

The error function, erf(x) is defined as:

$$erf\left(x\right)=\frac{2}{\sqrt{\pi}}\int_{0}^{x}e^{-z^2}dz \qquad 2.7.8$$

It may be noted that the above definite integral is not easy to calculate and tables containing approximate values for the argument 'x' are readily available. A few properties of the error function erf(x) are noted below:

a) erf(-x) = - erf(x) [Symmetry Property]

b) as x → + ∞, erf(x) → 1, since $\frac{2}{\sqrt{\pi}}\int_{0}^{\infty}e^{-z^2}dz=1$

c) if 'X' is a Gaussian random variable with mean m_x and variance σ_x^2, the probability that X lies in the interval (m_x – a, m_x +a) is:

$$P\left(m_x-a<X\le m_x+a\right)=erf\left(\frac{a}{\sqrt{2}\sigma_x}\right)=\frac{2}{\sqrt{\pi}}\int_{0}^{a/\sqrt{2}\sigma_x}e^{-z^2}dz \qquad 2.7.9$$

The complementary error function is directly related to the error function erf(x):

$$erfc\left(u\right)=\frac{2}{\sqrt{\pi}}\int_{u}^{\infty}\exp\left(-z^2\right)dz=\frac{2}{\sqrt{\pi}}\int_{u}^{\infty}e^{-z^2}dz \;=\; erfc\left(u\right)=1-erf\left(u\right) \qquad 2.7.10$$

A useful bound on the complementary error function erfc(v):

For large positive v,

$$\frac{\exp\left(-v^2\right)}{\sqrt{\pi}\cdot v}\left(1-\frac{1}{2v^2}\right)<erfc\left(v\right)<\frac{\exp\left(-v^2\right)}{\sqrt{\pi}\cdot v} \qquad 2.7.11$$

The Q- function or the Marcum function (**Fig.2.7.2**) is another frequently occurring function. Considering a standardized Gaussian random variable X with m_x = 0 & σ_x^2 = 1, the probability that an observed value of x will be greater than v is given by the following Q – function:

$$Q\left(v\right)=\frac{1}{\sqrt{2\pi}}\int_{v}^{\infty}\exp\left(-\frac{x^2}{2}\right)dx \qquad 2.7.12$$

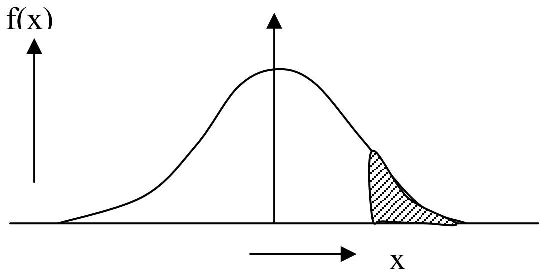

Fig.2.7.2 *The shaded area is measured by the Q-function*

We may note that,

$$Q(v) = \frac{1}{2} erfc\left(\frac{v}{\sqrt{2}}\right) \qquad 2.7.13$$

And conversely, putting $u = {}^{v}\!/\!_{\sqrt{2}}$, $\quad erfc(u) = 2Q\left(\sqrt{2}u\right)$ 2.7.14

Problems

Q2.7.1) If a fair coin is tossed ten times, determine the probability that exactly two heads occur.

Q2.7.2) Mention an example/situation where Poisson distribution may be applied.

Q2.7.3) Explain why Gaussian distribution is also called as Normal distribution.