Random Variables

Random variable (discrete or continuous) is used to derive the output statistical properties of a system whose input is a random variable or random in nature.

Definition

Consider an experiment whose outcomes are elements of the Universal set S. To every outcome e (elementary event), we assign a real number $X(e)$ to the outcome e, so that corresponding to each element of S, there is a real number. The probability of an elementary event occurring is not changed by the mapping from e to $X(e)$, so that

$$P(x) = P[X(e)] = P(e). \tag{1}$$

The rule of assignment (outcome e maps to a real number $X(e)$) is called a random variable. The mapping from the outcomes to the points on a real line is quite arbitrary and any required correspondence may be chosen. The reason for this mapping is that real numbers are much more easier to analysis.

The value of $X(e)$ is called the value of the random variable. We will use the notation X to denote the random variable and $x = X(e)$ to denote a particular value of the random variable X. When $x = X(e)$ may have a discrete set of values, X is a discrete random variable. When $x = X(e)$ may have a continuous set of values over an interval, then X is said to be a continuous random variable. The interval may be finite or infinite.

Distribution and Density Function

The Cumulative Distribution Function (CDF) of the random variable X is denoted by $F_X(x)$ such that

$$P(X \le x) = F_X(x). \tag{2}$$

$P(X \le x)$ denotes the probability of the random variable X whose value is less or equal to x after the rule of assignment. When $x = a$, a is a fixed number, we can simply substitute a into equation (2).

The important properties of $F_X(x)$ are as follows :

1. $F_X(x)$ is non-decreasing, so that

$$F_X(x+h) \ge F_X(x) \tag{3}$$

for $h > 0$.

2. $F_X(x)$ is a right continuous, so that

$$\lim_{\substack{h \to 0 \\ h > 0}} F_X(x+h) \ge F_X(x) \tag{4}$$

3. $$\lim_{x \to -\infty} F_X(x) = 0 \tag{5}$$

4. $$\lim_{x \to \infty} F_X(x) = 1, \tag{6}$$

assuming that the random varaible X can have only finite vaules. X can be discrete or continuous.

It will be assumed throughout the following discussion that X takes on only finite values, and the following notation will be used :

Equation (4) becomes

$$\lim_{\substack{h \to 0 \\ h > 0}} F_X(x+h) \geq F_X(x^+) \tag{7}$$

or becomes

$$\lim_{\substack{h \to 0 \\ h > 0}} F_X(x-h) \geq F_X(x^-). \tag{8}$$

Equation (5) becomes

$$\lim_{x \to -\infty} F_X(x) = F_X(x = -\infty) = F_X(-\infty). \tag{9}$$

Equation (6) becomes

$$\lim_{x \to \infty} F_X(x) = F_X(x = +\infty) = F_X(+\infty) \tag{10}$$

The followings are some important relationships :

1. $0 \leq F_X(a) \leq 1$ (11)

2. $F_X(b) - F_X(a) = P(a < x \leq b)$ (12)

3. $F_X(a) - F_X(a^+) = P(x = a)$ (13)

4. $F_X(a^-) = P(x < a)$ (14)

5. $F_X(a) = P(x \leq a)$ (15)
   $F_X(a) = P(X = x$, where $x \leq a)$.

A discrete random variable X can take on only the discrete values x_i, for $i = 1, 2, ..., n$. Its cumulative distribution function (CDF) remains constant between two adjacent values of x (i.e., between x_i and x_{i+1}) and increases suddenly at each x_i.

Figure 1

For a discrete random variable X whose possible values are x_i for $i = 1, 2, \ldots, n$.

$$P(X = x_i) \geq 0 \tag{16}$$

$$\sum_{i=1}^{n} P(X = x_i) = 1 \tag{17}$$

$$P(X \leq x_i) = \sum_{\text{all } x_i \leq x} P(X = x_i) \tag{18}$$

$$P(X \leq x_i) = F_X(x)$$

$$P(X = x_i) = F_X(x_i) - F_X(x_i^-). \tag{19}$$

A continuous random variable X can take on any value over one or more intervals, and its cumulative distribution function (CDF) $F_X(x)$ is continuous in nature.

Figure 2

For a continuous random variable,

$$P(X = x_i) = 0. \tag{20}$$

Furthermore, there are now an infinite number of possible values of the random varaible X. Thus, instead of considering the probability of random variable having a discrete value, we must consider the probability of it having one of a range of values.

Let $f_X(x_i)dx$ be the probability that the continuous random variable lies in the range of values from x_i - dx to x_i, such that

$$P(x_i - dx < x \leq x_i) = f_X(x_i)dx \tag{21}$$

x_i may here have any one of an infinite set of values. $f_X(x_i)$ is called the probability density function of the random variable X.

The probability that the continuous random variables lies in one of the n separate ranges x_1- dx to x_1, x_2- dx to x_2, …, x_n- dx to x_n, is

$$f_X(x_1)dx + f_X(x_2)dx + ...+ f_X(x_n)dx = [f_X(x_1) + f_X(x_2) + ...+ f_X(x_n)]dx. \tag{22}$$

This is due to the fact that in any experiment, the random variable can have only one value. Thus, the value cannot lie in the range x_i- dx to x_n if it lies in the range x_j- dx to x_j, where $i \neq j$. Thus, the n events are mutually exclusive and the probability of any one of the n events occurring is the sum of their individual probabilities.

It can now be seen that

$$P(a < x \leq b) = \int_a^b f_X(x)\, dx. \tag{23}$$

But from equation (12),

$$P(a < x \le b) = F_X(b) - F_X(a) \tag{24}$$

where $F_X(x)$ is the cumulative distribution function of the continuous random variable X.

Thus

$$F_X(b) - F_X(a) = \int_a^b f_X(x)\, dx. \tag{25}$$

From Figure 2, it can be deduced that

$$F_X(-\infty) = 0.$$

Therefore,

$$\begin{aligned} P(-\infty < x \le b) &= F_X(b) - F_X(-\infty) \\ P(-\infty < x \le b) &= F_X(b) \\ P(-\infty < x \le b) &= \int_{-\infty}^b f_X(x)\, dx. \end{aligned} \tag{26}$$

Similarly,

$$F_X(a) = \int_{-\infty}^a f_X(x)\, dx. \tag{27}$$

Differentiating $F_X(a)$ with respect to a, we obtain

$$f_X(a) = dF_X(a) / da$$

so that

$$f_X(x) = dF_X(x) / dx \tag{28}$$

and

$$F_X(x) = \int_{-\infty}^x f_X(x)\, dx. \tag{29}$$

From equation (21)

$$P(a - dx < x \le a) = f_X(a)dx \tag{30}$$

so that

$$f_X(a) = P(a - dx < x \le a) / dx. \tag{31}$$

Hence $f_X(a)$ is the probability density of the random variable X at the value a.

In summary, a function $f_X(x)$ is a continuous density function if it has the following properties :

1. $f_X(x) \geq 0, \quad -\infty < x < \infty$ (32)

2. $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$ (33)

3. $P(x \leq a) = F_X(a) = \int_{-\infty}^{a} f_X(x)\, dx.$ (34)

4. $f_X(x) = dF_X(x) / dx$ (35)

Transformation of Random Variables

Let $Y = h(X)$. The continuous random variable Y is a function of the continuous random variables X. We solve the equation $y = h(x)$ for x in terms of y, where x is the value of X and y is the value of the random variable Y.

A function $h(x)$ is monotonically increasing if $h(x = x_2) > h(x = x_1)$ when $x_2 > x_1$. A function $h(x)$ is monotonically decreasing if $h(x = x_2) < h(x = x_1)$ when $x_2 > x_1$.

For monotonic functions, there is a one-to-one correspondence between $h(x)$ and x, which means that for each value of the random variable X there is a unique value of $h(x)$.

1. Consider the case where $Y = h(x)$ is monotonically increasing and differentiable. Let the random variable X has a density function $f_X(x)$ and CDF of $F_X(x)$, while the random variable Y has a density function $f_Y(y)$ and CDF of $F_Y(y)$.

   Let $b = h(a)$. Since $Y = h(x)$ is monotonically increasing, one-to-one mapping, therefore there is no change in probability.

   Then

   $$P(Y \leq b) = P(X \leq a) \tag{36}$$

   or

   $$F_Y(b) = F_X(a) \tag{37}$$

   Also,

   $$dF_Y(b) / db = dF_X(a) / db \tag{38}$$

   or

   $$dF_Y(b) / db = dF_X(a) / da\, (da / db). \tag{39}$$

Therefore,

$$f_Y(b) = f_X(a)\ (da\ /db) \qquad (40)$$

$$f_Y(y) = f_X(x)\ (dx\ /dy) \qquad (41)$$

2. Consider the case where $Y = h(X)$ is monotonically decreasing and differentiable.

If $\quad b = h(a) \qquad (42)$

then $\ P(Y \leq b) = P(X \geq a) \qquad (43)$

and $F_Y(b) = 1 - F_X(a). \qquad (44)$

Thus $dF_Y(b)\ /\ db = -\ F_X(a)\ /\ da\ (da\ /\ db) \qquad (45)$

or $\quad f_Y(b) = -\ f_X(a)\ (da\ /\ db) \qquad (46)$

Replacing a by x and b by y

$$f_Y(y) = -\ f_X(x)\ (dx\ /\ dy) \qquad (47)$$

However the derivative $dx\ /\ dy$ is negative for monontically decreasing and differentiable random variable. But in equation (41), $dx\ /\ dy$ is always positive for monotically increasing and differentiable random variable. Thus equations (41) and (47) may be combined to give the follwoing general equation for monotonic functions:

$$f_Y(y) = f_X(x)\ |dx\ /\ dy| \qquad (48)$$

where $Y = h(X)$ is a one-to-one continuously differentiable function of X.

3. Consider the case where $Y = h(X)$ is not a monotone, the value x of the random variable X has the values $a_1, a_2, \ldots, a_n$, all of which are real. Thus

$$b = h(a_1) = h(a_2) = \ldots = h(a_n). \qquad (49)$$

Several values of X may correspond to a single value of Y, then

$$P(Y \leq b) = P(\text{all values of } X \text{ for which } Y \leq b) \qquad (50)$$

and

$$F_Y(b) = \int_{\{X\ /(Y\ \leq b\ )\}} f_X(x)\ dx \qquad (51)$$

where $f_X(x)$ is integrated over all values of X for which $Y \leq b$.

$F_Y(b)$ may be expressed as sum of n definite integrals of $f_X(x)$, the i-th of which is

$$\int_{k_i}^{a_i} f_X(x)\,dx \quad \text{or} \quad \int_{a_i}^{k_i} f_X(x)\,dx \tag{52}$$

where k_i is a constant such that $dk_i / db = 0$. This is shown in Figure 3.

Figure 3

Suppose $k_1 = +\infty$, then

$$\int_{k_1=-\infty}^{a_1} f_X(x)\,dx = F_X(a_1) \tag{53}$$

and

$$dF_X(a_1) / db = f_X(a_1)\,(a_1 / db) \tag{54}$$

When $dF_X(a_1)$ is differentiated to give $f_X(a_1)$ at the value $x = a_1$.

Since

$$\int_{a_2}^{k_2} f_X(x)\,dx = \int_{-\infty}^{k_2} f_X(x)\,dx - \int_{-\infty}^{a_2} f_X(x)\,dx$$

$$\int_{a_2}^{k_2} f_X(x)\,dx = F_X(k_2) - F_X(a_2). \tag{55}$$

Therefore,

$$dF_X(k_2) / db - dF_X(a_2) / db = 0 - dF_X(a_2) / db \tag{56}$$

where

$$dF_X(a_2) / db = f_X(a_2)\,(da_2 / db). \tag{57}$$

Since da_2 / db is negative (see Figure 3), it follows that

$$f_Y(b) = f_X(a_1)\,|da_1 / db| + \ldots + f_X(a_n)\,|da_n / db|. \tag{58}$$

In general,, for a given value y of Y, X has the values $x_1, x_2, \ldots, x_n$, so that

$$y = h(x_1) = h(x_2) = \ldots = h(x_n) \tag{59}$$

then

$$f_Y(y) = f_X(x_1)\ |dx_1 / dy| + \dots + f_X(x_n)\ |dx_n / dy|. \qquad (60)$$

Example 1

Consider the transformation $Y = cX + d$. We write the value of Y as y and the value of X as x. Thereofre $y = cx + d$.

Thus $x = (y - d) / c$

and $dx / dy = 1 / c$.

From equation (48),

$$f_Y(y) = f_X(x)\ |dx / dy|$$
$$f_Y(y) = f_X[\ (y - d) / c]\ |1 / c|. \qquad \text{Q.E.D.}$$

Example 2

Find $f_Y(y)$ where $Y = X^2$.

Summary

1. Transformation of random variable is useful in finding the characteristics of the output signal in a communication system where the input signal is random in nature.

2. $f_Y(y) = \sum_{i=1}^{n} f_X(x_i)\ |dx_i / dy|$.

3. The technique can be extended to two random variables X and Y (continuous) whose joint probability density function is $f_{X,Y}(x, y)$ and $Z = g(X, Y)$, $W = h(X, Y)$.

$$f_{Z,W}(z, w) = \sum_{i=1}^{n} f_{X,Y}(x_i, y_i)\ |J_i|,$$

where the Jacobian

$$J_i = \begin{vmatrix} \partial x_i / \partial z & \partial x_i / \partial w \\ \partial y_i / \partial z & \partial y_i / \partial w \end{vmatrix}.$$

Expectation

$P(X = x_i)$ and $f_X(x)$ give a total description of a discrete and continuous random varaibles, respectively. Alternatively, it is simpler to describe a random variable in terms of its statistical average (expectation) and moment.

Consider the discrete random variable X whose possible values are $\{x_i\}$. The expected value of X is defined to be

$$E[X] = \sum_i \; h(X = x_i)\, P(X = x_i) \tag{61}$$

i.e.,

$$\mu_X = h(x_i)\, P(x_i) \tag{62}$$

as the simplified notation.

Consider now the continuous random variable X. The expected value of X is defined to be

$$E[X] = \int_{-\infty}^{\infty} x\, f_X(x)\, dx. \tag{63}$$

If $h(X)$ is a function of the continuous random variable X, then the expected value of $h(X)$ is defined to be

$$E[h(X)] = \int_{-\infty}^{\infty} h(X = x)\, f_X(x)\, dx. \tag{64}$$

i.e.,

$$\mu_X = \int_{-\infty}^{\infty} h(X\,)\, f_X(x)\, dx. \tag{65}$$

as the simplified notation.

The expected value of a random variable has some obvious properties:

1. $E[h_1(X) + h_2(X)] = E[h_1(X)\,] + E[h_2(X)]$ (66)

2. $E[k\, h(X)\,] = k\, E[h(X)]$ (67)

3. $E[k] = k$ (68)

4. $E[h(X)] \geq 0$ (69)

if $h(X = x)$ for all x in the continuous random variable case or $h(X = x_i)$ for all $\{x_i\}$ in the discrete random variable case.

Moments

In the case of a discrete random variable X, the k-th moment about the mean is defined to be

$$\eta_k = E[(X - \mu_X)^k] = \sum_i \; (x_i - \mu_X)^k\, P(X = x_i). \tag{70}$$

The second moment about the mean is called the variance $\sigma_X{}^2$, and its square root σ_X is known as the standard deviation.

$$\sigma_X^2 = E[(X - \mu_X)^2] \tag{71}$$

Example 1

$$\eta_1 = E[X - \mu_X] = E[X\ ] - E[\mu_X]$$
$$\eta_1 = E[X\ ] - \mu_X$$
$$\eta_1 = \mu_X - \mu_X$$
$$\eta_1 = 0 \tag{72}$$

i.e., the first moment about the mean μ_X is zero.

Example 2

$$\sigma_X^2 = E[(X - \mu_X)^2]$$
$$\sigma_X^2 = E[X^2 - 2X\mu_X + \mu_X^2]$$
$$\sigma_X^2 = E[X^2] - 2\mu_X\, E[X] + \mu_X^2$$
$$\sigma_X^2 = E[X^2] - 2\mu_X^{\ 2} + \mu_X^2$$
$$\sigma_X^2 = E[X^2] - \mu_X^{\ 2}.$$

Therefore,

$$E[X^2] = \sigma_X^2 + \mu_X^{\ 2}. \tag{73}$$

i.e., the second moment about the zero (the origin) is equal to the variance plus the square of the mean $\mu_X^{\ 2}$. $E[X^2]$ is called the mean square value.

$E[X^2]$ - mean square value
$\mu_X^{\ 2}$ - square of the mean
σ_X^2 - variance
σ_X - standard deviation

In the case of a continuous random variable X, the k-th moment about the mean is defined to be

$$\eta_k = E[(X - \mu_X)^k] = \int_{-\infty}^{\infty} (x - \mu_X)^k f_X(x)\, dx. \tag{74}$$

The second moment about the mean is

$$\sigma_X^2 = E[(X - \mu_X)^2] = \int_{-\infty}^{\infty} (x - \mu_X)^2 f_X(x)\, dx. \tag{75}$$

Equations (72) and (73) also hold for the continuous random variable case. In general, the followings are two inportant properties of moments.

1. The first moment about the mean is zero.

$\eta_1 = E[X - \mu_X]$
$\eta_1 = E[X] - E[\mu_X]$
$\eta_1 = \mu_X - \mu_X$
$\eta_1 = 0.$

2. The second moment about the origin is equal to the variance plus the square of the mean.

$\sigma_X^2 = E[(X - \mu_X)^2]$
$\sigma_X^2 = E[X^2] - \mu_X^2.$
$E[X^2] = \sigma_X^2 + \mu_X^2.$

Example 3

Given $Y = a X + b$ where X, Y are continuous random variables and a, b are constants; find μ_Y and σ_Y^2 in terms of a, b, μ_X and σ_X^2.

$E[Y] = E[a X + b]$

$\mu_Y = E[a X] + E[b]$
$\mu_Y = a E[X] + E[b]$
$\mu_Y = a \mu_X + b$

$\sigma_Y^2 = E[(Y - \mu_Y)^2]$
$\sigma_Y^2 = E[(a X + b - a \mu_X - b)^2]$
$\sigma_Y^2 = E[a^2 (X - \mu_X)^2]$
$\sigma_Y^2 = a^2 E[(X - \mu_X)^2]$
$\sigma_Y^2 = a^2 \sigma_X^2.$

Tchebycheff's Inequality

Consider a continuous random variable X with mean μ_X and variance σ_X^2. Define a random variable Y by

$$Y = \begin{cases} 0 & \text{for } |X - \mu_X| < \lambda\sigma_X \\ \lambda^2 \sigma_X^2 & \text{for } |X - \mu_X| \geq \lambda\sigma_X \end{cases} \qquad (76)$$

for any real $\lambda > 0$.

The expected value of Y is

$$E[Y] = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy \tag{77}$$

$$E[Y] = 2 \int_{0}^{x-\mu_X} y\, f_Y(y)\, dy + 2 \int_{x-\mu_X}^{\infty} y\, f_Y(y)\, dy$$

$$E[Y] = \underbrace{0P(Y=0)}_{P(|X-\mu_X|<\lambda\sigma_X)} + \underbrace{\lambda^2\sigma_X^2\, P(Y=\lambda^2\sigma_X^2)}_{P(|X-\mu_X|\geq\lambda\sigma_X)}$$

$$E[Y] = \lambda^2\sigma_X^2\, P(|X - \mu_X| \geq \lambda\sigma_X). \tag{78}$$

When the value y of the random variable Y takes on $\lambda^2\sigma_X^2$

$$Y = \lambda^2\sigma_X^2$$

thus

$$|X - \mu_X| \geq \lambda\sigma_X$$

i.e.,

$$(X - \mu_X)^2 \geq \lambda^2\sigma_X^2$$
$$(X - \mu_X)^2 \geq Y.$$

Thus

$$E[(X - \mu_X)^2] \geq E[Y]$$
$$\sigma_X^2 \geq E[Y].$$

Use equation (78),

$$\sigma_X^2 \geq \lambda^2\sigma_X^2\, P(|X - \mu_X| \geq \lambda\sigma_X). \tag{79}$$

Therefore,

$$1/\lambda^2 \geq P(|X - \mu_X| \geq \lambda\sigma_X). \tag{80}$$

Let $K = \lambda\sigma_X$, then equation (80) becomes

$$\sigma_X^2/K^2 \geq P(|X - \mu_X| \geq K) \tag{81}$$

which is the Tchebycheff’s Inequality. It is also valid for a discrete random variable.

This inequality can be used to obtain bounds on th eprobability of finding X outside an interval μ_X +/- K.

Weak Law of Large Numbers

Let x_1, x_2, ..., x_n be statistically independent sample values of the continuous random variable X, which has a mean value μ_X and a variance $\sigma_X{}^2$.

Let y be the mean of these n sample values, so that

$$y = (1/n) \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i / n. \qquad (82)$$

Because X is a continuous random variable, $\{x_i\}$ are not the possible values of a discrete random variable.

Each x_i / n has a mean value μ_X / n and a variance $(\sigma_X / n)^2$. $\{x_i / n\}$ are statistically independent random variables.

$$\mu_Y = E[Y] = n\,[\mu_X / n\,] \qquad (83)$$

i.e., sum of the means of $\{x_i / n\}$.

$$\sigma_Y{}^2 = n\,(\sigma_X / n\,)^2 \qquad (84)$$

i.e., sum of the variances of $\{x_i / n\}$.

From equations (83) and (84), the larger the value of n, the smaller is $\sigma_Y{}^2$ and closer is the value y of Y likely to be to n. This can also be shown by Tchebycheff's Inequality as follows.

$$1/\lambda^2 \geq P(|Y - \mu_Y| \geq \lambda\sigma_Y). \qquad (85)$$

Let $\varepsilon = \lambda\sigma_Y$

so that $\lambda^2 = \varepsilon^2 / \sigma_Y{}^2 = n\varepsilon^2 / \sigma_X{}^2$.

Equation (85) becomes

$$\sigma_X{}^2 / (n\varepsilon^2) = P(|Y - \mu_Y| \geq \varepsilon).$$

From equation (83), $\mu_Y = \mu_X$

$$\sigma_X{}^2 / (n\varepsilon^2) = P(|Y - \mu_X| \geq \varepsilon). \qquad (86)$$

For a fixed value of ε (positive quality), the probability that the mean value of y differs from μ_X by an amount greater than ε is inversely proportional to n. y becomes a better estimate of μ_X as n increases. hence the discrepancy between y and μ_X reduces.

Charecteristic Functions

The characteristic function of a random variable X is defined to be

$$\varphi_X(w) = E[e^{jwx}], \quad j^2 = -1. \tag{87}$$

For a discrete random variable X,

$$\varphi_X(w) = \sum_i e^{jwx_i} P(X = x_i) \tag{88}$$

and for a continuous random variable X,

$$\varphi_X(w) = \int_{-\infty}^{\infty} e^{jwx} f_X(x)\, dx. \tag{89}$$

for a continuous random variable X

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(w) e^{-jwx}\, dw. \tag{90}$$

$f_X(x)$ and $\varphi_X(w)$ are Fourier transform pair.

Application : To determine the probability density of the sum of two independent random variables, it is often first/easier to determine the characteristic function of the sum of the random variables and then the probability density function.

Example

$$Y = X_1 + X_2$$
$$\varphi_Y(w) = E[e^{jwy}]$$
$$\varphi_Y(w) = E[e^{jwx_1 + jwx_2}]$$
$$\varphi_Y(w) = E[e^{jwx_1}]\, E[e^{jwx_2}]$$
$$\varphi_Y(w) = \varphi_{X_1}(w) \varphi_Y(w)$$

$$f_Y(y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_Y(w) e^{-jwy}\, dw. \qquad \text{Q.E.D.}$$

Joint Distribution

We will now consider two random variables X and Y that take on the values x_1, x_2, ..., x_n and y_1, y_2, ..., y_n. The joint distribution function of the two random variables is defined as

$$F_{XY}(x, y) = P(X \leq x \text{ and } Y \leq y). \tag{91}$$

$F_{XY}(x, y)$ has the following properties.

1. $F_{XY}(x, y)$ is monotonic non-decreasing in both variables.

2. $F_{XY}(x, y) \to 0$ if either x or y approaches $-\infty$.

3. $F_{XY}(x, y) \to 1$ if both x and y approach ∞.

4. $F_{XY}(x, \infty) = F_X(x)$.

5. $F_{XY}(\infty, y) = F_Y(x)$,

   where $F_X(x)$ and $F_Y(y)$ are the distribution functions for X and Y, respectively.

For discrete random variables, the joint probability is

$$P(X \le x,\ Y \le y) = P(X \le x \text{ and } Y \le y). \quad (92)$$
$$P(X \le x,\ Y \le y) = \sum_{x_i = x} \sum_{y_j = y} P(X = x_i,\ Y = y_j). \quad (93)$$

and satisfies the following equations.

1. $$\sum_{i=1}^{n} \sum_{j=1}^{n} P(X = x_i,\ Y = y_j) = 1. \quad (94)$$

2. When the value of X is x_i,

$$P(X = x_i) = \sum_{j=1}^{n} P(X = x_i,\ Y = y_j) \quad (95)$$

$$P(X = x_i) = \sum_{j=1}^{n} P(X = x_i / Y = y_j)\, P(Y = y_j). \quad (96)$$

3. $$P(X = x_i / Y = y_j) = P(X = x_i,\ Y = y_j) / P(Y = y_j). \quad (97)$$

For continuous random variables, the joint probability density function $f_{XY}(x, y)$ satisfies the following equations.

1. $$f_{XY}(x, y) \ge 0, \quad (98)$$

   where $-\infty < x < \infty$ and $-\infty < y < \infty$.

2. $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dx\, dy = 1. \quad (99)$$

3. $F_{XY}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(x, y)\, dx\, dy.$ (100)

4. $f_{XY}(x, y) = \frac{\delta^2}{\delta x\ \delta y} F_{XY}(x, y)],$ (101)

where $F_{XY}(x, y)$ is continuous and differentiable.

Expectation

The expectation of a function $g(X, Y)$ of the discrete random variables X and Y is

$$E[g(X, Y)] = \sum_i \sum_j g(x_i, y_j)\, P(X = x_i\ ,\ Y = y_j). \tag{102}$$

The expectation of a function $g(X, Y)$ of the continuous random variables X and Y is

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{XY}(x, y)\, dx\, dy. \tag{103}$$

In the following analysis only continuous random variables will be considered, but the corresponding relationships apply throughout to discrete random variables.

The expected value (expectation) of two random variables has the following properties :

1. $E[g_1(X, Y) + g_2(X, Y)] =] = E[g_1(X, Y)] + E[g_2(X, Y)]$ (104)

2. $E[k\, g(X, Y)] = k\, E[g(X, Y)]$ (105)

3. $E[k\ ] = k$, if k is a constant (106)

4. $E[k\, g(X, Y)] \geq 0$, if $g(X, Y) \geq 0$ for all X and Y. (107)

Two random variables X and Y are independent if

$$P(X \leq x_i\ ,\ Y \leq y_j) = P(X \leq x_i\ )\, P(Y \leq y_j). \tag{108}$$

Thus

$$F_{XY}(x, y) = F_X(x)\, F_Y(y). \tag{109}$$

The above equation implies that

$$f_{XY}(x, y) = f_X(x)\, f_Y(y). \tag{110}$$

where $f_{XY}(x, y)$ is the joint density function. When X and Y are independent,

$$E[g(X)\, g(Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x) g(y) f_{XY}(x, y)\, dx\, dy. \tag{111}$$

$$E[g(X)\,g(Y)] = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx \int_{-\infty}^{\infty} g(y) f_Y(y)\,dy. \quad (112)$$

$$E[g(X)\,g(Y)] = E[g(X)]\,E[g(Y)]. \quad (113)$$

In particular, if x and y are values of two independent random variables X and Y,

$$E[X]\,E[Y] = E[XY]. \quad (114)$$

Moments

The kl-th moment about the mean is defined to be

$$\eta_{kl} = E[(X - \mu_X)^k (Y - \mu_Y)^l] \quad (115)$$

where μ_X is the mean of X and μ_Y is the mean of Y. In particular,

$$\eta_{20} = E[(X - \mu_X)^2] = \sigma_X{}^2 \quad (116)$$

and the variance of Y is

$$\eta_{02} = E[(Y - \mu_Y)^2] = \sigma_Y{}^2. \quad (117)$$

The covariance of X and Y is defined to be

$$\eta_{11} = E[(X - \mu_X)(Y - \mu_Y)] = \sigma_{XY}. \quad (118)$$

and the correlation coefficient of X and Y is defined to be

$$\rho_{XY} = \frac{\eta_{11}}{\sqrt{\eta_{20}}\sqrt{\eta_{02}}} = \frac{\eta_{XY}}{\eta_X\;\eta_Y}. \quad (119)$$

It follows that $-1 \le \rho_{XY} \le 1$. ρ_{XY} gives a measure of dependence between random variables X and Y. For example, $\rho_{XY} = 0.9$ or $\rho_{XY} = -0.9$ give the same measure of dependence between X and Y.

In general,

$$\eta_{11} = E[(X - \mu_X)(Y - \mu_Y)]$$
$$\eta_{11} = E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y]$$
$$\eta_{11} = E[XY] - \mu_X\mu_Y. \quad (120)$$

Therefore, $E[XY] = \eta_{11} + \mu_X\mu_Y$ where $E[XY]$ gives the ($k = l = 1$)-st moment about zero.

If X and Y are independent, then

$$E[XY] = E[X]E[Y] \quad (121)$$

Thus

$$\eta_{11} = E[X]E[Y] - \mu_X\mu_Y \tag{122}$$
$$\eta_{11} = 0.$$

ρ_{XY} is also equal to zero.

Summary

Two random variables X and Y are

statistically independent if $f_{XY}(x, y) = f_X(\mathrm{x})\, f_Y(y)$

uncorrelated if $E[XY] = E[X]\, E[Y]$

orthogonal if $E[XY] = 0$; i.e., either $E[X] = 0$ or $E[Y] = 0$.

Hence

Figure 4

If X and Y are statistically independent, then functions of X and Y (i.e., $g(X)$ and $g(Y)$_ are also independent.

$$f[g(X), g(Y)] = f[g(X)]\, f[g(Y)] \,. \tag{123}$$

$f[g(X), g(Y)]$ is the joint probability density of $g(X)$ and $g(Y)$.

If X and Y are statistically independent, then functions of X and Y (i.e., $g(X)$ and $g(Y)$) are uncorrelated.

$$E[g(X)\, g(Y)] = E[g(X)]\, E[g(Y)] \,. \tag{124}$$

If X and Y are uncorrelated, then their covariance and correlation coefficient are both zero:

$$\mu_{XY} = \sigma_{XY} = E[(X - \mu_X)(Y - \mu_Y)]$$
$$\mu_{XY} = \sigma_{XY} = E[XY - \mu_X Y - \mu_Y X + \mu_X\mu_Y]$$
$$\mu_{XY} = \sigma_{XY} = E[X]E[Y] - \mu_X E[Y] - \mu_Y E[X] + \mu_X\mu_Y$$
$$\mu_{XY} = \sigma_{XY} = \mu_X\mu_Y - \mu_X\mu_Y - \mu_X\mu_Y + \mu_X\mu_Y$$
$$\mu_{XY} = \sigma_{XY} = 0.$$

$$\rho_{XY} = 0 = \sigma_{XY} / \sigma_X\sigma_Y.$$

This implies that the random variables $(X - \mu_X)$ and $(Y - \mu_Y)$ are orthogonal.

For any two random variables X and Y,

$$E[X + Y] = E[X] + E[Y] \tag{125}$$
$$E[X + Y] = \mu_X + \mu_Y$$

so that the mean of their sum is equal to the sum of their means.

The variance

$$\sigma^2_{X+Y} = E[\{(X + Y) - (\mu_X + \mu_Y)\}^2]$$
$$\sigma^2_{X+Y} = E[X^2 + Y^2 + 2XY + \mu_X^2 + \mu_Y^2 + 2\mu_X\mu_Y - 2\mu_X X - 2\mu_X Y - 2\mu_Y X - 2\mu_Y Y]$$
$$\sigma^2_{X+Y} = E[(X - \mu_X)^2 + (Y - \mu_Y)^2 + 2(X - \mu_X)(Y - \mu_Y)]$$
$$\sigma^2_{X+Y} = \sigma_X^2 + \sigma_Y^2 + 2\sigma_{XY}. \tag{126}$$

If X and Y are uncorrelated, $\sigma_{XY} = 0$. Equation (126) becomes

$$\sigma^2_{X+Y} = \sigma_X^2 + \sigma_Y^2. \tag{127}$$

Thus if two random variables are uncorrelated, the the variance of their sum equals the sum of their variance.

If X and Y are orthogonal, it can be seen that

$$E[(X + Y)^2] = E[X^2] + E[Y^2] \tag{128}$$

because $E[XY] = 0$ for orthogonality, i.e., $E[X]$ or $E[Y] = 0$. Since $E[XY] = E[X]E[Y]$ for uncorrelation, this implies that X and Y are uncorrelated if they are orthogonal.

In general, if X and Y are statistically independent then X and Y must also be uncorrelated. However, uncorrelated random variables need not be statistically independent.

Example

X and Y are statistically independent,

$$E[XY] = E[X]E[Y]$$

but if thay are uncorrelated, then, in general,

$$E[XY] \neq E[X]E[Y].$$

Examples of Various Distributions

1. **Binomial**

   Consider a sequence of repeated independent trials of a given random experiment with only two possible outcomes on each trial and with the probabilities of two outcomes fixed throughout the trials. Such sequences are called Bernoulli traials.

It is conventional to call one of the two outcome S (success) and the other F (failure). The probability of a success is p and the probability of a failure q, where

$$p + q = 1. \tag{129}$$

Thus, in any one trial,

$$P(S) = p \tag{130}$$

$$P(F) = 1 - p. \tag{131}$$

Since the individual trials are independent, the probability of obtaining a given sequence of successes and failure in n trials, where there are k successes and $n - k$ failures, is

$$P_k = p^k (1 - p)^{(n - k)}. \tag{132}$$

In a sequence of n trials, there are $\binom{n}{k}$ different sequences containing k successes and $n - k$ failures. These sequences are mutually exclusive and each has P_k. Thus the probability of obtaining k successes in a sequence of n independent trials, where any sequence of k sequence and $n - k$ failures is acceptable, is

$$b_k = \binom{n}{k} p^k (1 - p)^{(n - k)}. \tag{133}$$

b_k is the (k+1)-term in the binomial expansion of $(p + q)^n$ where

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

$$(p + q)^n = \sum_{k=0}^{n} b_k \tag{134}$$

Since $q = 1 - p$, equation (134) becomes

$$\sum_{k=0}^{n} b_k = [p + (1 - p)]^n = 1.$$

In addition, b_k is non-negative, it is therefore a probability function, written as P_k, of the discrete random variable k.

The variation of b_k with k, for a given n, gives the corresponding Binomial Distribution with a probability

$$P(k) = \binom{n}{k} p^k (1 - p)^{(n - k)} \tag{135}$$

for $k = 0, 1, 2, \ldots, n$.

It may be shown that the mean value $E[k]$ of the binomial distribution is

$$E[k] = \mu_k = n\,p \tag{136}$$

and the variance $E[(k - \mu_k)^2]$ is

$$\begin{aligned} \sigma_k^2 &= n\,p\,(1 - p) \\ \sigma_k^2 &= E[(k - \mu_k)^2]. \end{aligned} \tag{137}$$

2. **Poisson**

Consider the binomial distribution where n becomes very large and p very small but $n\,p$ remains of a reasonable size. Suppose that as $n \to \infty$, $p \to 0$ and

$$n\,p = \lambda \tag{138}$$

where λ is a finite non-zero positive constant.

The first term of the binomial distribution is

$$b_0 = \binom{n}{0} p^0 (1 - p)^n = (1 - p)^n. \tag{139}$$

Thus

$$\begin{aligned} \lim_{n \to \infty} b_0 &= \lim_{n \to \infty} (1 - [\lambda / n])^n \\ \lim_{n \to \infty} b_0 &= e^{-\lambda}. \end{aligned} \tag{140}$$

The second term of the binomial distribution is

$$\begin{aligned} b_1 &= \binom{n}{1} p^1 (1 - p)^{(n - 1)} = \frac{n}{1!}\frac{p}{(1 - p)}(1 - p)^n \\ b_1 &= \frac{\lambda}{1!\,(1 - [\lambda/n])}\, b_0 \end{aligned} \tag{141}$$

and

$$\begin{aligned} \lim_{n \to \infty} b_1 &= \lim_{n \to \infty} \frac{\lambda}{1!\,(1 - [\lambda/n])}\, e^{-\lambda} \\ \lim_{n \to \infty} b_1 &= \lim_{n \to \infty} \frac{\lambda}{1!}\, e^{-\lambda}. \end{aligned} \tag{142}$$

In general,

$$\lim_{n \to \infty} b_k = \lim_{n \to \infty} \frac{\lambda^k}{k\,!}\, e^{-\lambda}, \tag{143}$$

for $k = 0, 1, 2, \ldots, n$. Thus if $n \to \infty$ and $p \to 0$ in such a way that $n\,p = \lambda$, where λ is a finite non-zero positive constant, in the limit b_k becomes the $(k + 1)$-th term p_k of the Poisson Distribution, where

$$p_k = \frac{\lambda^k}{k\,!}\, e^{-\lambda}, \tag{144}$$

for $k = 0, 1, 2, \ldots, \infty$. In addition, p_k is non-negative, and

$$\sum_{k=0}^{\infty} p_k = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k\,!}$$
$$\sum_{k=0}^{\infty} p_k = e^{-\lambda}\, e^{\lambda} = 1. \tag{145}$$

Thus p_k is the probability function $P(k)$ of the discrete random variable k.

It can also be shown that the mean μ_k and the variance $\sigma_k{}^2$ of p_k are each equal to λ. Thus

$$E[k] = \mu_k = \lambda \tag{146}$$

$$E[(k - \mu_k)^2] = \sigma_k{}^2 = \lambda. \tag{147}$$

3. Gaussian or Normal Distribution

The random variable X is binomially distributed if the probability that it has the value x is

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{(n - x)}, \tag{148}$$

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

As n increases to a very large value and assuming that $n\,p\,(1 - p) >> 1$, the probability that the random variable has the value x becomes

$$P(X = x) \simeq \frac{1}{\sqrt{[2\pi np\,(1 - p\,)]}}\ \exp\left[-\frac{(x - np\,)^2}{2np\,(1 - p\,)}\right]. \tag{149}$$

The random variable is here discrete and not continuous. In the limit as $n \to \infty$, the random variable X can be for practical purposes be considered to be continuous from $-\infty$ to ∞, with a probability density function

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}}\ \exp\left[-\frac{(x - \mu_X)^2}{2\sigma_X^2}\right], \tag{150}$$

$-\infty < x < \infty$.

Figure 4.1

The approximation is amazingly good even for small n ($n \geq 5$) provided that p is not too far from 0.5 and $|x|$ not much greater then σ_X. Equation (150) is now defined as the Gaussian probability density function of the random variable X.

Its mean is μ_X and its variance is σ_X^2. $f_X(x)$ is symmetrical about μ_X.

$$P(X > a) = \int_a^\infty \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{(x-\mu_X)^2}{2\sigma_X^2}\right] dx \qquad (151)$$

$$P(X > a) = \int_{\frac{(a-\mu_X)}{\sigma_X}}^\infty \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\, dz. \qquad (152)$$

To evaluate this integral, we will use tabulated values of the Q function, which is defined as

$$Q(y) = \frac{1}{2\pi} \int_y^\infty \exp(-z^2/2)\, dz; \quad y > 0 \qquad (153)$$

Thus

$$P(X > a) = Q(y) \qquad (154)$$

where $y = (a - \mu_X) / \sigma_X$. In fact, a Gaussian probability density function is completely defined by μ_X and σ_X^2, where X is the random variable itself.

Important Results :

1. Sum of two or more statistically independent Gaussian random variables is also a Gaussian random variable.

2. If two Gaussian random variables are uncorrelated then thay are also statistically independent. Gaussian random variables are the only case where uncorrelation gives independence.

3. $\mu_X = 0$ gives a WHITE Gaussian Random Variable (WGRV).

4. $\mu_X = 0$, $\sigma_X^2 = 0$, then $f_X(x)$ is said to have a unit normal distribution.

Central Limit Theorm

The theorem states that if $X_1,\ X_2,\ ...\ X_n$ are statisticvally independent random variables and $Y = \sum_i X_i$, then, as n becomes very large, the distribution of Y approaches a Gaussian distribution. It can now be seen that a random noise can be modelled by the Gaussian probability density function because noise in communicatin sytstems is often made up of a very large number of independent noise sources.

References

[1] Schanmugan, K. S., 'Digital and Analog Communication Systems', John Wiley & Sons, 1979.

[2] Peebles, P. Z., 'Probability, Random Variables and Random Signal Principles', 2/e, McGraw-Hill, 1987.

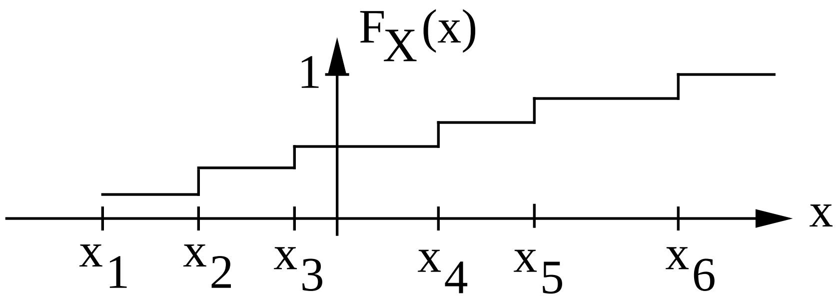

Figure 1

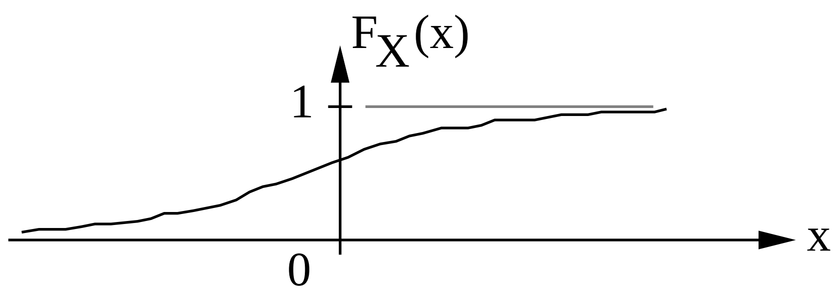

Figure 2

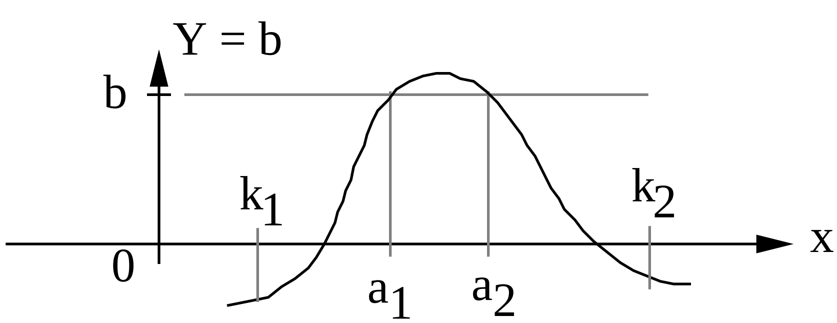

Figure 3

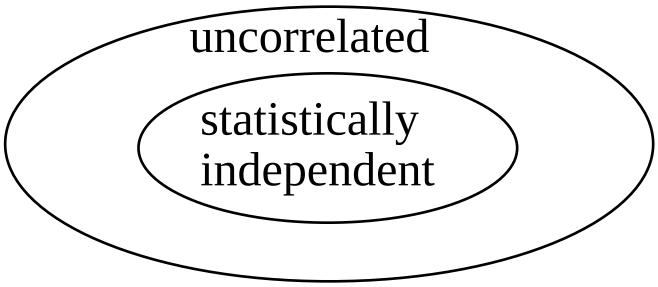

Figure 4

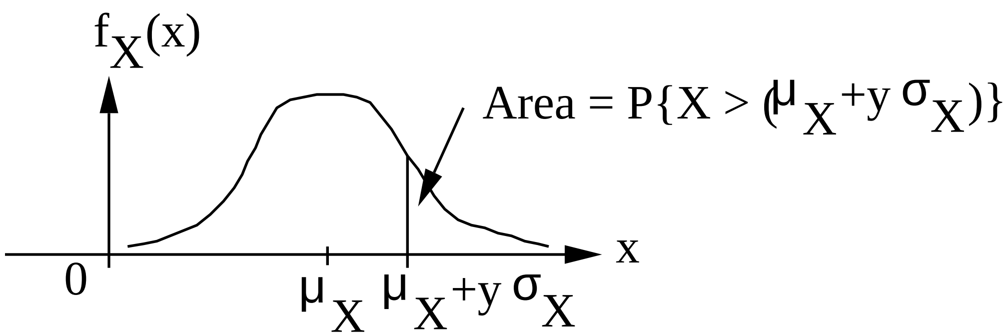

Figure 4.1