Random variable

A random variable X on a sample space S is a function, from set S to set R of real numbers, such that the pre image of any interval of R is an event in S.

$$X: S \to R$$

If S is *discrete* sample space in which every subset of S is an event, then every real valued function of S is a discrete random variable.

Continuous random variables are those where the range is space R_X is a Continuum of numbers such as an interval or a union of intervals.

<u>Sum and product of random variable</u>

Let X & Y be random variables on the same sample space S. The $X+Y$, $X+k$, kX, XY where k is any real number are functions on S defined as follows; for $s \in S$

i. $(X+Y)(s) = X(s) + Y(s)$
ii. $(kX)(s) = kX(s)$
iii. $(X+k)(s) = X(s) + k$
iv. $(XY)(s) = X(s)Y(s)$

More generally for any polynomial, exponential, or continuous function $h(t)$ we define $h(X)$ to be the function on S, defined as

$$[h(X)](s) = h[X(s)]$$

In short we use the notation $P(X = a)$ & $P(a \le X \le b)$ respectively, for "X *maps to* a" & "X *maps into the interval* $[a, b]$", i.e.;

$P(X = a) \Longrightarrow P\{s \in S: X(s) = a\}$

$P(a \le X \le b) \Longrightarrow P\{s \in S: a \le X \le b\}$

Probability mass function

Represented by $p(x)$ is defined as $p(x) = P(X = x)$.

If $x_i, i \ge 1$ represents the possible values of X (X is finite random variable)

$$\sum_{i=1}^{n} p(x_i) = 1$$

Probability distribution

in case of X being finite random variable on S then,

$R_X = \{x_1, x_2, \dots, x_n\}$ is called range of X.

We assume that $x_1 < x_2 < \dots < x_n$ then X induces a function $'f'$ which assigns probabilities to the points in R_X.

$$f(x_k) = P(x_k) = P\{\, s \in S : X(s) = x_k\}$$

The set of ordered pair $(x_k, f(x_k))$ is usually given in the tabular form

x	x_1	x_2	x_3	$\dots$	x_n
$f(x)$	$f(x_1)$	$f(x_2)$	$f(x_3)$		$f(x_n)$

The function f is called the probability distribution are simply distribution of the random variable X . It satisfies the following two conditions;

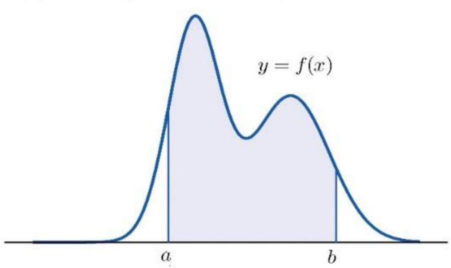
i. $f(x_k) \geq 0$
ii. $\sum_k f(x_k) = 1$

§§ It is convenient to extend a probability distribution f to all real numbers by defining $f(x) = 0$ when X does not belongs to R. Graph of such function $f(x)$ is called a *probability graph.* Sometimes probability distribution is represented as $[x_i \, , p_i]$ *or* $[x_i \, , \, p(x_i)]$ *or* $[x \, , \, p(x)]$ rather than $[x, f(x)]$

Continuous random variable

If X is a random variable whose range is space R_X is a Continuum of numbers, such as an interval then the set $a \leq X \leq b$ is an event in S and therefore the probability $P(a \leq X \leq b)$ is well defined. We assume there is a piece-wise continuous function $f : R \to R$ such that $P(a \leq X \leq b)$ is equal to the area under the graph of f between $x = a$ & $x = b$, i.e.;

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$



$P(a < X < b) =$ area of shaded region

In this case X is said to be a continuous random variable the function f is called *continuous probability function* (or density function) of X it satisfies the conditions;

i. $f(x) \geq 0 \quad \forall \; x \in R$
ii. $\int_{-\infty}^{\infty} f(x)dx = 1$

§§ $P(X = c) = 0 \, , \; c \in (a, b)$ because infinite points in (a, b) (also area under curve for one point will be zero.)

Cumulative Distribution Function

Let X be a random variable discrete (or continuous). CDF $'F'$ of X is the function $f: R \to R$ defined as

$$F(x) = P(X \leq x) = P\{X \in (-\infty, x]\}$$

i. If X be a discrete random variable with probability distribution $f(x)$ then

$$F(x) = \sum_{x_i \leq x} f(x_i) \qquad or \qquad F(x) = \sum_{x_i \leq x} p(x_i)$$

ii. On the other hand if X be a continuous random variable with probability density function $f(x)$ effects

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \dots (1)$$

In either case, F has the following properties

i. F is monotonically increasing i.e.; $F(a) \leq F(b)$ where ever $a \leq b$.

ii. $\lim_{x \to -\infty} F(x) = 0$ & $\lim_{x \to \infty} F(x) = 1$

iii. Differentiating equation (1) w.r.t. 'x' $\frac{dF(x)}{dx} = f(x)$ *exist*

iv. $P(a < X < b) = F(b) - F(a)$

Expectation $(\boldsymbol{E(X)})$ (for discrete r.v. X)

Let X be discrete then expectation $E(X)$ is defined as

$$E(X) = x_1 f(x_1) + x_2 f(x_2) + \cdots + x_n f(x_n) = \sum_{i=1}^{n} x_i f(x_i) \qquad \left(or \sum_{i=1}^{n} x_i p(x_i) \right)$$

$E(X)$ is nothing but weighted average of outcomes.

If suppose each x_i occurs with same probability then $p_i = \frac{1}{n}$, so

$$E(X) = \frac{\sum x_i}{n}$$

Which is precisely mean value of $x_1, x_2, \dots, x_n$, i.e.; of r.v. X .

Hence , $E(X) = \mu$.

In general expectation of any function $g(X)$ of a r.v. X is given as,

$$E\{g(X)\} = \sum_{i=1}^{n} g(x_i) f(x_i)$$

Variance $\left(\boldsymbol{\sigma^2(X)}\right)$ (for discrete r.v. X)

is defined as

$$Var(X) = \sigma^2(X) = (x_1 - \mu)^2 f(x_1) + (x_2 - \mu)^2 f(x_2) + \cdots + (x_n - \mu)^2 f(x_n)$$

$$= \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = E\{(X - \mu)^2\}$$

Standard deviation (σ) is defined as positive square root of variance, i.e.;

$$\sigma = \sqrt{Var(X)}$$

Theorem $\quad Var(X) = E(X^2) - \mu^2 = x_1^2 f(x_1) + x_2^2 f(x_2) + \cdots + x_n^2 f(x_n) - \mu^2$

Proof $\quad$ using $\sum f(x_i) = 1$ & $\sum x_i f(x_i) = \mu$

$$Var(X) = \sigma^2(X) = \sum_{i=1}^{n} (x_i - \mu)^2 f(x_i) = \sum (x_i^2 - 2\mu x_i + \mu^2) f(x_i)$$

$$= \sum x_i^2 f(x_i) - 2\mu \sum x_i f(x_i) + \mu^2 \sum f(x_i) \; = \sum x_i^2 f(x_i) - 2\mu^2 + \mu^2$$

$$= \sum x_i^2 f(x_i) - \mu^2 = E\{X^{\mathbf{2}}\} - \{E(X)\}^2$$

For ***continuous random variable*** **(c.r.v.)** $\boldsymbol{X}$ with density function $f(x)$,

Expectation

$$\boldsymbol{E(X)} = \int_{-\infty}^{\infty} x f(x)\, dx = \mu$$

For any function $g(X)$ of random variable X

$$\mu_{g(X)} = E\{g(X)\} = \int_{-\infty}^{\infty} g(x) f(x)\, dx \qquad (when\ it\ exist)$$

$$\boldsymbol{\sigma^2(X)} = Var(X) = E\{(X - \mu)^2\} = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2 \qquad = E(X^2) - \mu^2$$

Standard deviation is σ .

Theorem: $\qquad Var(aX + b) = a^2 Var(X)$

Proof : $\qquad$ We shall show that

$$Var(X + k) = Var(X) \qquad \& \qquad Var(kX) = k^2 Var(X)$$

we have

$$\mu_{X+k} = E(X + k) = E(X) + k = \mu_X + k \qquad \& \qquad \mu_{kX} = E(kX) = kE(X) = k\mu_X$$

$$Var(X + k) = \sum (x_i + k)^2 f(x_i) - \mu_{X+k}^2$$

$$= \sum x_i^2 f(x_i) + 2k \sum x_i f(x_i) + k^2 \sum f(x_i) - (\mu_X + k)^2$$

$$= \sum x_i^2 f(x_i) + 2k\mu_X + k^2 - \mu_X^2 - k^2 - 2k\mu_X$$

$$= \sum x_i^2 f(x_i) - \mu_X^2 = Var(X)$$

$Var(kX) \; = \sum k^2 x_i^2 f(x_i) - \mu_{kX}^2 = k^2 \sum x_i^2 f(x_i) - k^2 \mu_X^2$

$= k^2\{\sum x_i^2 f(x_i) - \mu_X^2\} = k^2 Var(X)$

Joint distribution

Let $X \; and \; Y$ are discrete random variables on same sample space S with the respective Ranger space

$R_X = \{x_1, x_2, \dots x_n\}$ & $R_Y = \{y_1, y_2, \dots y_n\}$

The joint distribution or joint probability function of $X \,\&\, Y$ represented by 'f' on the product space $R_X \times R_Y$ is defined as,

$f(x_i, y_j) = P(X = x_i, Y = y_j) = P(\{s \in S : X(s) = x_i, Y(s) = y_j\})$, clearly

$$\sum_i \sum_j f(x_i, y_j) = 1$$

Also for any region A in $xy -\; plane$,

$$P[(x, y) \in A] = \sum_A \sum f(x, y)$$

X & Y are said to be independent if,

$P(X = x, Y = y) = P(X = x)P(Y = y) \quad \forall \; x \,\&\, y$

X ↓ \ Y →	y_1	y_2	.	y_n	
x_1	$f(x_1, y_1)$	$f(x_1, y_2)$	.	$f(x_1, y_n)$	$h(x_1)$
x_2	$f(x_2, y_1)$	$f(x_2, y_2)$	.	$f(x_2, y_n)$	$h(x_2)$
.	.	.	.	.	
x_n	$f(x_3, y_1)$	$f(x_3, y_2)$	.	$f(x_3, y_n)$	$h(x_n)$
	$k(y_1)$	$k(y_2)$		$k(y_n)$	1

Here,

$$h(x_i) = \sum_j f(x_i, y_j) \qquad \& \qquad k(y_j) = \sum_i f(x_i, y_j)$$

These h & k are called *marginal distribution.*

The mean or expected value of the random variable $g(X,Y)$ is given as

$$\mu_{g(X,Y)} = E\{g(X,Y)\} = \sum_i \sum_j g(x_i, y_j) f(x_i, y_j)$$

In particular

$$\mu_{XY} = E(XY) = \sum_i \sum_j x_i y_j f(x_i, y_j)$$

If μ_X & μ_Y are mean of $X \, and \, Y$ respectively then covariance of X & Y is

$$Cov(X,Y) = \sum_i \sum_j (x_i - \mu_X)(y_j - \mu_Y) f(x_i, y_j)$$

$$= E\{(X - \mu_X)(Y - \mu_Y)\} = E(XY) - \mu_X . \mu_Y$$

$$= E(XY) - E(X)E(Y)$$

And correlation of X & Y is $\rho(X,Y) = \frac{Cov(X,Y)}{\mu_X \mu_Y}$

If X & Y are continuous random variable then $f(x,y)$ is a joint density function if

i. $f(x,y) \geq 0 \quad \forall \; x, y$

ii. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y) dy dx = 1$

iii. $P[(x,y) \in A] = \iint_A f(x,y) dy dx$, $\quad A$ is the region in $xy - plane$

The marginal distribution of X alone and Y alone are

$$h(x) = \int_{-\infty}^{\infty} f(x,y) dy \qquad \& \qquad k(y) = \int_{-\infty}^{\infty} f(x,y) dx$$

Mean or Expectation

$$\mu_{g(X,y)} = E\{g(X,Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x,y) f(x,y) dy dx$$

$$Cov(X,Y) = \sigma_{XY} = E\{(X - \mu_X)(Y - \mu_Y)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x,y) dy dx$$

$$= E(XY) - E(X)E(Y) = E(XY) - \mu_X \mu_Y$$

Properties

1. Let X & Y be two independent random variables then,

   $E(XY) = E(X)E(Y)$

   By definition, $E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dydx$

   Since X & Y are independent, hence $f(x,y) = h(x)k(y)$, using in above equation

   $E(XY) = \int_{-\infty}^{\infty} xh(x)dx \times \int_{-\infty}^{\infty} yk(y)dy = E(X)E(Y)$

2. $\sigma_{aX+b}^2 = a^2\sigma_X^2$,

   Since $\sigma_{aX+b}^2 = E\{(aX+b) - \mu_{aX+b}\}^2 = E\{(aX+b) - a\mu_X - b\}^2$

   $= a^2E\{X-\mu_X\}^2 = a^2\sigma_X^2$

3. $\sigma_{aX+bY}^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$

   $\sigma_{aX+bY}^2 = E\{(aX+bY) - \mu_{aX+bY}\}^2$

   $E(aX+bY) = a\mu_X + b\mu_Y$, hence ,

   $\sigma_{aX+bY}^2 = E\{(aX+bY) - \mu_{aX+bY}\}^2 = E\{a(X-\mu_X) - b(Y-\mu_Y)\}^2$

   $= a^2E\{X-\mu_X\}^2 + b^2E\{Y-\mu_Y\}^2 + 2abE\{(X-\mu_X)(Y-\mu_Y)\}$

   $= a^2\sigma_X^2 + b^2\sigma_Y^2 + 2ab\sigma_{XY}$

Conditional distribution

Let X & Y be two random variables discrete or continuous, the conditional distribution of the random variable Y given that $X = x$ is

$$f(Y|X=x) = \frac{f(x,y)}{h(x)}$$

Similarly, $$f(X|Y=y) = \frac{f(x,y)}{k(y)}$$

Also,

$$P(a<X<b|Y=y) = \sum_x f(x|y) \qquad for\ discrete\ X\,\&\,Y$$

$$P(a<X<b|Y=y) = \int_a^b f(x|y)\,dx \qquad for\ continuous\ X\,\&\,Y$$

Chebyshev's Inequality

Let X be a random variable with mean μ and standard deviation σ. Then for any positive $k > 0$ the probability that a value of X lies in the interval $[\mu - k\sigma, \mu + k\sigma]$ is at least $\left(1 - {}^{1}/_{k^2}\right)$, i.e.;

$$P(\mu - k\sigma \leq X \leq \mu + k\sigma) \geq 1 - \frac{1}{k^2}$$

or $$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

We prove it for the case when X is continuous with density function $f(x)$.

For a non negative random variable X

$$E(X) = \int_0^\infty xf(x)\,dx = \int_0^a xf(x)\,dx + \int_a^\infty xf(x)\,dx \qquad for\ any \quad a > 0$$

$$\geq \int_a^\infty xf(x)\,dx \geq a\int_a^\infty f(x)\,dx = a\,P(X \geq a)$$

Thus, $$P(X \geq a) \leq \frac{E(X)}{a} \qquad (Markov\ Inequality) \qquad \ldots(A)$$

Replacing X by $(X - \mu)^2$ and $a\ by\ k^2$

$$P\{(X - \mu)^2 \geq k^2\} \leq \frac{E(X - \mu)^2}{k^2}$$

$$P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2} \qquad \ldots(B)$$

or $$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

The result (B) can also be written as

$$P(|X - \mu| < \sigma k) \geq 1 - \frac{1}{k^2}$$

Moments & Moment Generating Function (m.g.f.)

If $g(X) = X^r \ for\ r = 0,1,2,\ldots$ then the r[th] moment of the random variable X about the origin, denoted by μ_r', is given as,

$$\mu_r' = E(X^r) = \begin{cases} \sum_x x^r p(x) & if\ \ X\ is\ discrete \\ \int_{-\infty}^{\infty} x^r f(x)dx & if\ \ X\ \ is\ continuous \end{cases}$$

The r[th] moment of a variable x about 'mean' denoted by μ_r is given as

$$\mu_r = E(X - \mu)^r = \begin{cases} \sum_x (x - \mu)^r p(x) & if\ \ X\ is\ discrete \\ \int_{-\infty}^{\infty} (x - \mu)^r f(x)dx & if\ \ X\ \ is\ continuous \end{cases}$$

The first and second moment about the origin

$\mu_1' = E(X) = \mu\,,\quad \&\quad \mu_2' = E(X^2)\quad \Rightarrow \sigma^2 = \mu_2' - \mu^2$

§§ The method to obtain moments other than definition, we have moment generating function.

The moment generating function of a random variable X about origin is denoted by $M_X(t)$ and is defined as

$$M_X(t) = E(e^{Xt}) = \begin{cases} \sum_x e^{tx} p(x) & if\ \ X\ \ is\ discrete \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & if\ \ X\ \ is\ coninuous \end{cases} \qquad \ldots(A)$$

The r[th] moment about the origin is obtained by differentiating mgf w.r.t. 't' $at\ t = 0$, i.e.;

$$E(X^r) = \mu_r' = \frac{d^r}{dt^r} M_X(t)|_{t=0}$$

The m.g.f. of the random variable X about an arbitrary point 'a' is define as

$$M_a(t) = E\left(e^{(X-a)t}\right) = e^{-at} E(e^{Xt})$$

Theorem: Let X be a random variable with m.g.f. $M_X(t)$, then

$$\frac{d^r}{dt^r} M_X(t)\Big|_{t=0} = \mu_r'$$

Proof: Assuming that equation (A) is differentiable

$$\frac{d^r}{dt^r} M_X(t)\Big|_{t=0} = \left\{ \begin{array}{ll} \sum_x x^r e^{xt} p(x) & if \;\; X \; is \; discrete \\ \int_{-\infty}^{\infty} x^r e^{xt} f(x) dx & if \;\; X \;\; is \; continuous \end{array} \right|_{t=0} = \mu_r'$$

§§ The mgf of the sum of the independent random variables is just a product of the individual mgf.

$$M_{X+Y}(t) = E\left(e^{(X+Y)t}\right) = E(e^{Xt} e^{Yt}) = E(e^{Xt}) E(e^{Yt}) = M_X(t) M_Y(t)$$

§§ $M_{X+a}(t) = e^{at} M_X(t)$

$$M_{X+a}(t) = E\left(e^{(X+a)t}\right) = e^{at} E(e^{Xt}) = e^{at} M_X(t)$$

§§ $M_{aX}(t) = M_X(at)$

$$M_{aX}(t) = E(e^{aXt}) = E(e^{(at)X}) = M_X(at)$$