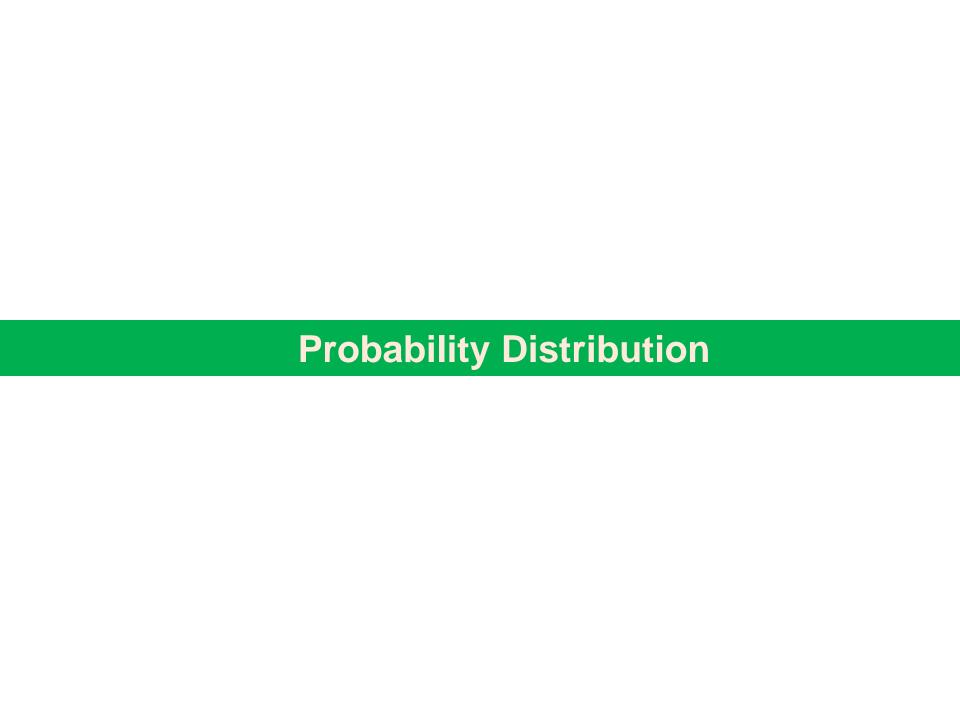
## MATH 353: STATISTICS

## Dr. Eric Nimako Aidoo

en.aidoo@yahoo.com 0202901980



### Random Variable

A Random Variable is a function that assigns a real number to each outcome in the sample space of a random experiment. Random variables are denoted by uppercase letters, such as *X*, *Y*,*Z* 

**Discrete random variable**: A random variable, *X* is said to be discrete if it can take on only a finite number or a countably infinite possible values of *X*.

**Continuous random variables**: A random variable, *X* is said to be continuous if it can assume infinitely many values within an interval of real numbers.

# **Probability**

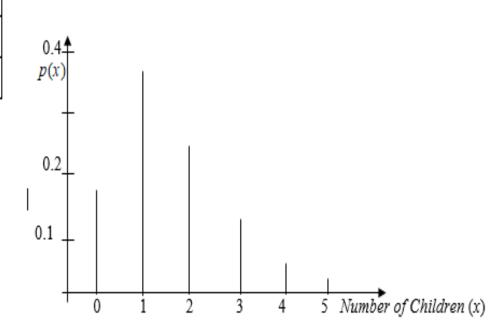
The probability distribution of a random variable X, denoted p(x) or f(x) is a description of the set of possible values of X along with the probability, p(x) or f(x) associated with each of the possible values.

**Discrete Distributions**: The probability distribution for a discrete random variable *X* is a *formula*, *table*, *graph* or *any device* that specifies the probability associated with each possible value of *X*.

### Example

A study on 300 families in a community was conducted, noting the number of children, *X* and its occurrence, *f* in a family results the following distribution and line histogram.

X	0	1	2	3	4	5
f	54	114	72	42	12	6
p(x)	0.18	0.38	0.24	0.14	0.04	0.02



#### **Definition**

The probability that X takes a discrete value, denoted P(X=x) or p(x) is called probability mass function (pmf), if the following properties are satisfied

1. 
$$p(x) = P(X = x)$$

2. 
$$0 \le p(x) \le 1 \text{ or } p(x) \ge 0$$

3. 
$$\sum p(x) = 1$$

### **Example**

The number of telephone calls received in an between 12:00 noon and 1:00 pm has the probability function given by

X	0	1	2	3	4	5	6
p(x)	0.05	0.20	0.25	0.20	0.10	0. 15	0.05

- 1. Verify that the function is a probability mass function
- 2. Find the probability that there will be 3 or more calls

#### **Solution**

i) To verify that it is a probability mass function, we have

$$p(x) > 0$$
, for  $x = 0, 1, 2, 3, 4, 5, 6$ .

$$\sum_{i=1}^{6} p(x) = 0.05 + 0.20 + 0.25 + 0.20 + 0.10 + 0.15 + 0.05 = 1$$

ii) 
$$P(x > 3) = \sum_{i=3}^{6} p(x)$$
  
= $P(3) + P(4) + P(5) + P(6)$   
= $0.20 + 0.10 + 0.15 + 0.005$   
= $0.50$ 

### **Example**

1) Verify that the following probability functions is a probability mass function (pmf).

$$p(x) = \begin{cases} \frac{1}{21}(2x+3), & x = 1, 2, 3\\ 0, & elsewhere \end{cases}$$

11) Find the value of *k* given that the function is a probability mass function.

$$p(x) = \begin{cases} k(x-1), & x = 3, 4, 5 \\ 0, & elsewhere \end{cases}$$

### **Solution**

p(x) > 0, for all x, and

$$\sum_{x=1}^{3} p(x) = \frac{1}{21} \sum_{x=1}^{3} (2x+3) = \frac{1}{21} \{ 2(1) + 3 + 2(2) + 3 + 2(3) + 3 \}$$
$$= \frac{1}{21} (5+7+9) = 1$$

(ii) We determine k by assuming p(x) is probability mass function,

$$\sum_{x=3}^{5} p(x) = \sum_{x=3}^{5} k(x-1) = 1$$

$$K(x-1) = k \{(3-1) + (4-1) + (5-1)\} = 1$$

$$9k = 1$$

$$\iff k = \frac{1}{0}$$

■The relative frequency behaviour of continuous random variable, X is modelled by a function, f(x) which is more often called probability density function (pdf).

■The graph of f(x) is a smooth curve defined over a range of interval [a, b] the random variable, X assumes.

#### **Definition**

The probability distribution for a continuous random variable X denote by f(x) is probability density function (pdf), if the following properties are satisfied

- 1.  $f(x) \ge 0$ , for any value of x
- $2. \quad \int_{-\infty}^{\infty} f(x) \, dx = 1$
- 3.  $P(a \le x \le b) = \int_a^b f(x) dx$

### **Example**

i) Let x be a continuous random variable with probability density function,

$$f(x) = \begin{cases} \frac{1}{6}x + k, & 0 \le x \le 3\\ 0, & elsewhere \end{cases}$$

Evaluate *k* and hence find  $P(1 \le x \le 2)$ 

(ii) Determine the value of and hence compute the probabilities,  $P(1 \le x \le 2)$  and P(x > 2).

$$f(x) = \begin{cases} k x & , & 0 \le x \le 3, k > 0 \\ 3k(4 - x) & , & 3 < x \le 4 \\ 0 & , & otherwise \end{cases}$$

### **Solution 1**

Given the probability density function,

$$f(x) = \begin{cases} \frac{1}{6}x + k, & 0 \le x \le 3\\ 0, & elsewhere \end{cases}$$

then,

(i) 
$$\int_{0}^{3} f(x) dx = 1$$

$$\int_{0}^{3} (\frac{1}{6}x + k) dx = 1$$

$$\frac{1}{12} x^{2} + kx \Big|_{0}^{3} = 1$$

$$\left[\frac{1}{12} (3)^{2} + 3k\right] - 0 = 1$$

$$\frac{3}{4} + 3k = 1$$

$$3k = \frac{1}{4} \Leftrightarrow k = \frac{1}{12}$$

#### **Solution 1**

Hence, 
$$f(x) = \begin{cases} \frac{1}{12} (2x+1), & 0 \le x \le 3\\ 0, & elsewhere \end{cases}$$

$$P(1 \le x \le 2) = \int_{1}^{2} \frac{1}{12} (2x+1) dx$$

$$= \frac{1}{12} \left[ x^{2} + x \right]^{2}$$

$$= \frac{1}{12} \left[ (2^{2} + 2) - (1^{2} + 1) \right]$$

$$= \frac{1}{12} (6 - 2) = \frac{1}{3}$$

### **Solution 2**

For f(x) is probability density function,  $f(x) \ge 0$  for all values of x and k > 0. We also show that,

$$\int_{0}^{4} f(x)dx = 1$$

$$\int_{0}^{3} kxdx + \int_{3}^{4} 3k(4-x)dx = 1$$

$$\left(\frac{kx^{2}}{2}\right)_{0}^{3} + 3k\left(4x - \frac{x^{2}}{2}\right)_{3}^{4} = 1$$

$$\frac{9k}{2} + 3k\left[(16 - 8) - (12 - \frac{9}{2})\right] = 1$$

$$\frac{9k}{2} + \frac{3k}{2} = 1$$

$$6k = 1 \iff k = \frac{1}{6}$$

#### **Solution 2**

Hence. 
$$f(x) = \begin{cases} \frac{1}{6}x & , 0 \le x \le 3 \\ \frac{1}{2}(4-x), 3 < x \le 4 \\ 0 & , elsewhere \end{cases}$$

$$P(1 \le x \le 2) = \int_{1}^{2} f(x) dx$$

$$= \int_{1}^{2} \frac{1}{6}x dx = \left| \frac{x^{2}}{12} \right|_{1}^{2} = \frac{1}{12} (2^{2} - 1) = \frac{1}{4}$$

$$P(x > 2) = \int_{2}^{4} f(x) dx$$

$$= \int_{2}^{3} \frac{1}{6} x dx + \int_{3}^{4} \frac{1}{2} (4-x) dx$$

### Solution 2

$$= \left| \frac{x^2}{12} \right|_2^3 + \frac{1}{2} \left| 4x - \frac{x^2}{2} \right|_3^4$$

$$= \frac{1}{12} (9 - 4) + \frac{1}{2} (16 - 8) - \frac{1}{2} (12 - \frac{9}{2})$$

$$= \frac{5}{12} + \frac{1}{4} = \frac{2}{3}$$

- The cumulative distribution function (cdf) for a random variable x, denoted F(x), is defined by  $F(x) = P(X \le x)$ .
- If x is a discrete random variable then  $F(x) = \sum_{t=0}^{x} p(t)$ , which is a step function.
- If X is a continuous random variable then

$$F(x) = \int_{-\infty}^{x} f(t)dt$$
 where  $-\infty \le x \le \infty$ ,

$$f(x) = \frac{dF(x)}{dx}$$
  $P(x_1 \le x \le x_2) = F(x_2) - F(x_1)$ 

### **Properties of Cumulative Distribution Function (CDF)**

In each case, F(x) is a monotonic increasing function with the following properties:

- (i)  $F(a) \le F(b)$ , wherever  $a \le b$ , and
- (ii) The limit of F(x) to the left is 0 and to the right is 1. That is,  $\lim_{x \to -\infty} F(x) = 0$  and  $\lim_{x \to \infty} F(x) = 1$

(iii) 
$$0 \le F(x) \le 1$$

### **Example**

Given the probability mass function,

X	0	1	2	3
p(x)	1/4	1/8	1/2	1/8

Find the cumulative distribution function

### **Solution**

$$F(x) = P(X \le x) = \sum_{x=0}^{3} p(x)$$

$$F(0) = P(X \le 0) = p(0) = \frac{1}{4}$$

$$F(1) = P(X \le 1) = p(0) + p(1) = \frac{1}{4} + \frac{1}{8} = \frac{3}{8}$$

$$F(2) = P(X \le 2) = p(0) + p(1) + p(2) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} = \frac{7}{8}$$

$$F(3) = P(X \le 3) = p(0) + p(1) + p(2) + p(3) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} + \frac{1}{8} = 1$$

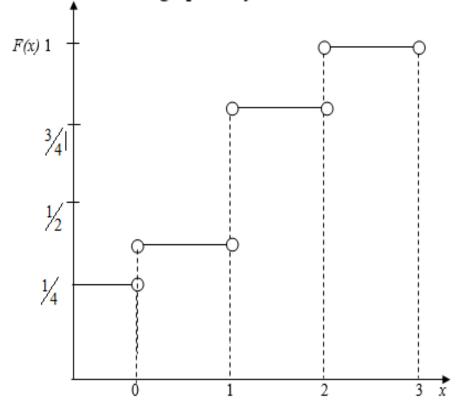
### **Solution**

$$F(3) = P(X \le 3) = p(0) + p(1) + p(2) + p(3) = \frac{1}{4} + \frac{1}{8} + \frac{1}{2} + \frac{1}{8} = 1$$

Hence the cumulative distribution is

X	0	1	2	3
F(x)	1/4	3/ /8	7/ /8	1

Which is illustrated graphically below:



The expectation or expected value (or simply the mean) of the random variable, x is defined by

(i) 
$$\mu = E(x) = \sum_{x} x p(x)$$
, if x is discrete.

(ii) 
$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx$$
, if  $x$  is continuous and  $-\infty \le x \le \infty$ 

• The *variance* of the random variable, x with probability distribution, p(x) or f(x) is defined by

$$\sigma^2 = Var(x) = E[(x - \mu)^2] = E(x^2) - \mu^2$$
, where

(i) 
$$Var(x) = \sum_{x} (x - \mu)^2 p(x)$$
  
=  $\sum_{x} x^2 p(x) - \mu^2$ , if x is discrete.

(ii) 
$$Var(x) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

$$= \int_a^b x^2 f(x) dx - \mu^2$$
, if x is continuous.

The *standard deviation* of *x* is the square root *x*... That

is, 
$$\sigma = \sqrt{Var(x)}$$

### **Example**

Compute the expected value ( $\mu$ ) and standard deviation ( $\sigma^2$ ) of the random variable, x with the following probability distribution:

(i)

X	1	2	3	4	5
p(x)	0.1	0.3	0.2	0.3	0.1

(ii) 
$$f(x) = \begin{cases} 6x(1-x), & 0 < x < 1 \\ 0, & elsewhere \end{cases}$$

#### **Solution**

The expected value of *x* or mean,

$$\mu = \sum_{x=1}^{5} x \, p(x) = 1(0.1) + 2(0.3) + 3(0.2) + 4(0.3) + 5(0.1) = 3.0$$

The variance of x,

$$Var(x) = \sum_{x=1}^{5} (x - \mu)^2 p(x) = \sigma^2$$

$$= (1 - 3)^2 (0.1) + (2 - 3)^2 0.3 + (3 - 3)^2 (0.2) + (4 - 3)^2 (0.3)$$

$$+ (5 - 3)^2 (0.1)$$

$$= 0.4 + 0.3 + 0 + 0.3 + 0.4 = 1.4, \text{ or}$$

#### **Solution**

$$Var(x) = Var(x) = \sum_{x=1}^{5} x^{2} p(x) - \mu^{2} = \sigma^{2}$$

$$= 1^{2}(0.1) + 2^{2}(0.3) + 3^{2}(0.2) + 4^{2}(0.3) + 5^{2}(0.1) - (3)^{2}$$

$$= 0.1 + 1.2 + 1.8 + 4.8 + 2.5 - 9 = 1.4$$

Hence the standard deviation,

$$\sigma = \sqrt{1.4} = 1.18$$

#### **Solution**

(ii) Given the probability density function,

$$f(x) = \begin{cases} 6x(1-x), & 0 < x < 1 \\ 0, & elsewhere \end{cases}$$

The mean of x is

$$\mu = E(x) = \int_0^1 x f(x) dx$$

$$= \int_0^1 6x^2 (1 - x) dx$$

$$= \int_0^1 (6x^2 - 6x^3) dx$$

#### **Solution**

$$= \left[\frac{6}{3}x^3 - \frac{6}{4}x^4\right]_0^1$$

$$= 2(1)^3 - \frac{3}{2}(1)^4 - 0 = 2 - \frac{3}{2} = \frac{1}{2} = 0.5$$

The variance of x,

$$\sigma^{2} = Var(x) = E(x^{2}) - \mu^{2}$$

$$= \int_{0}^{1} x^{2} f(x) dx - \mu^{2}$$

$$= \int_{0}^{1} 6x^{3} (1 - x) dx - (0.5)^{2}$$

#### **Solution**

$$= \int_0^1 (6x^3 - 6x^4) dx - 0.25$$

$$= \left[ \frac{6}{4} x^4 - \frac{6}{5} x^5 \right]_0^1 - 0.25$$

$$= \frac{3}{2} - \frac{6}{5} - 0.25$$

$$= \frac{3}{10} - \frac{1}{4} = \frac{1}{20} = 0.05$$

Hence the standard deviation,

$$\sigma = \sqrt{0.05} = 0.224$$

### Example 2

Let y have the probability distribution

$$f(y) = \begin{cases} y & , 0 \le y < \frac{1}{2} \\ \lambda(4-y) & , \frac{1}{2} \le y \le 4 \\ 0 & , elsewhere \end{cases}$$

- i) Find the value of  $\lambda$  and
- ii) Use it to determine the mean and the standard deviation

#### **Solution**

To find  $\lambda$  we have,

$$\int_0^4 f(y) \, dy = 1$$

$$\int_0^{\frac{1}{2}} y \, dy + \lambda \int_{\frac{1}{2}}^4 (4 - y) \, dy = 1$$

$$\left| \frac{1}{2} y^2 \right|_0^{\frac{1}{2}} + \lambda \left[ 4y - \frac{1}{2} y^2 \right]_{\frac{1}{2}}^4 = 1$$

$$\frac{1}{8} + \lambda \left\{ \left[ 4(4) - \frac{1}{2}(4)^2 \right] - \left[ 4\left(\frac{1}{2}\right) - \frac{1}{2}\left(\frac{1}{2}\right)^2 \right] \right\} = 1$$

#### **Solution**

$$\frac{1}{8} + \lambda \left[ (16 - 8) - \left( 2 - \frac{1}{8} \right) \right] = 1$$

$$49/8 \lambda = 7/8 \Leftrightarrow \lambda = 1/7$$

Hence, 
$$f(y) = \begin{cases} y, & 0 \le y < \frac{1}{2} \\ \frac{1}{7}(y - y), & \frac{1}{2} \le y \le 4 \\ 0, & elsewhere \end{cases}$$

#### **Solution**

$$\mu = E(y) = \int_0^{\frac{1}{2}} y^2 dy + \int_{\frac{1}{2}}^4 \frac{1}{7} y (4 - y) dy$$

$$= \frac{1}{3} y^3 \Big|_0^{\frac{1}{2}} + \int_{\frac{1}{2}}^4 \frac{1}{7} (4y - y^2) dy$$

$$= \frac{1}{24} + \frac{1}{7} \Big[ (2y^2 - \frac{1}{3}y^3)_{\frac{1}{2}}^4 \Big]$$

$$= \frac{1}{24} + \frac{1}{7} \Big[ (32 - \frac{64}{3}) - (\frac{1}{2} - \frac{1}{24}) \Big] = \frac{256}{168} = \frac{3}{2} = 1.5$$

### **Expectation and Variance of Random Variable**

#### **Solution**

For the standard deviation,  $\sigma$ , we have

$$\sigma^{2} = Var(x) = E(x^{2}) - \mu^{2}$$

$$= \int_{0}^{y_{2}} y^{3} dy + \int_{\frac{1}{2}}^{4} \frac{1}{7} y^{2} (4 - y) dy - (1.5)^{2}$$

$$= \left[ \frac{1}{4} y^{4} \right]_{0}^{\frac{1}{2}} + \frac{1}{7} \left[ \left( \frac{4}{3} y^{3} - \frac{1}{4} y^{4} \right) \right]_{\frac{1}{2}}^{4} - 2.25$$

$$= \frac{73}{24} - 2.25 = \frac{19}{24} = 0.79167$$

Hence, the standard deviation,

$$\sigma = \sqrt{0.79167} = 0.88976$$

### **Expectation and Variance of Random Variable**

#### Try it your self

The probability density of a random, y is given by

$$f(y) = \begin{cases} \lambda y^{2}(1-y) &, 0 \le y \le 1\\ 0 &, elsewhere \end{cases}$$

(i) Find the value of  $\lambda$  and the standard deviation of y

### **Expectation and Variance of Random Variable**

#### Try it your self

Given the random variable x with probability density function,

$$f(x) = \begin{cases} ke^{=0.001x}, & x > 0\\ 0, & elsewhere \end{cases}$$

Find the value of k, the mean of x and the probability, P(x > 1,050).

#### **Moments**

Let x be the random variable with probability distribution, function f(x) and g(x) be real-valued function of x. Then

$$E[g(x)] = \sum_{x} g(x) f(x)$$
, if x is discrete

$$=\int_{x}^{x} g(x)f(x)$$
, if x is continuous

### The kth Moment about the origin

If  $g(x) = x^k$ , we obtain the *kth moment about the origin*, denoted  $U_k$ . and defined by

$$U_k^i = E(x^k) = \sum x^k f(x) \ or \int_{R_x} x^k f(x) dx$$
, where

$$U_1^1 = E(x) = \sum_x x f(x) \ Or \int_{R_x} x f(x) dx,$$

which is the mean and also called the first moment about origin

$$U_2^1 = \sum_{x} x^2 f(x) \ Or \int_{R_x} x^2 f(x) dx,$$

which is called the second moment about the origin.

#### The kth Moment about the mean

If  $g(x) = (x - \mu)^k$ , we get the *kth moment about the mean*, denoted and defined by

$$U_k = E[(x-\mu)^k] = \sum_{x} (x-\mu)^k f(x) \ Or \int_{R_x} (x-\mu)^k f(x) dx$$

#### Uses of moments about the mean in statistical analysis

- $U_2 = E(x \mu)^2$  the second moment about the mean also known as variance.
- $U_3 = E(x \mu)^3$ , the third moment about the mean describes the skewness of a distribution. The measure of skewness is given by  $a_3 = \frac{U_3}{\sigma^3}$  .if  $a_3 \neq 0$ , the distribution becomes skewed (that is, tailed to the right or left depending on whether  $a_3 > 0$  or  $a_3 < 0$ )

### Uses of moments about the mean in statistical analysis

- $U_4 = E(x \mu)^4$  the fourth moment about the mean is the peakness (or kurtosis) of a distribution. The degree of peakness is  $a_4 = \frac{U_4}{\sigma^4}$ .
- If  $a_4 = 3$ , the distribution is normally distributed.
- If  $a_4 < 3$ , the distribution flattens at the centre than the normal distribution.
- If  $a_4 > 3$ , the distribution becomes more peaked at the centre than the normal distribution

#### Expansion of the moments about the mean

(i) The second moment about the mean,

$$U_{2} = E[(x - \mu)^{2}]$$

$$= E[x^{2} - 2\mu x + \mu^{2}] = E(x^{2}) - \mu^{2}$$

(ii) The third moment about the mean,

$$U_3 = E[(x - \mu)^3]$$

$$= E[x^3 - 3\mu x^2 + 3\mu^2 x - \mu^3]$$

$$= E(x^3) - 3\mu E(x^2) + 2\mu^3$$

#### Expansion of the moments about the mean

(iii) The fourth moment about the mean,

$$U_4 = E[(x - \mu)^4]$$

$$= E[x^4 - 4\mu x^3 + 6\mu^2 x^2 - 4\mu^3 x + \mu^4]$$

$$= E(x^4) - 4\mu E(x^3) + 6\mu^2 E(x^2) - 3\mu^4$$

#### **Moment Generating Function**

Moments of most distributions can also be determined by finding a function in a form of series. The coefficients of the series give the moments. The function which generates the moments is called *moment generating function*. If it exists, the mgf for the distribution function, f(x) is given by:

$$M_x(t) = E(e^{tx}) = \sum_{\forall x} e^{tx} f(x) \quad Or \quad \int_{Rx} e^{tx} f(x) dx$$

### Expansion of $e^{tx}$

Now expanding the function,  $e^{tx}$  and taking expectation,

$$e^{tx} = 1 + tx + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots + \frac{t^k x^k}{k!}$$

$$M_x(t) = E(e^{tx}) = E(1 + tx + \frac{t^2 x^2}{2!} + \frac{t^3 x^3}{3!} + \dots + \frac{t^k x^k}{k!})$$

$$= 1 + t E(x) + \frac{t^2}{2!} E(x^2) + \frac{t^3}{3!} E(x^3) + \dots + \frac{t^k}{k!} E(x^k)$$

$$= 1 + \frac{t^1}{1!} U_1^1 + \frac{t^2}{2!} U_2^1 + \frac{t^3}{3!} U_3^1 + \dots + \frac{t^k}{k!} U_k^1$$

The coefficient of  $\frac{t^k}{k!}$  is  $U_k^1$ , the *kth* moment about the origin, which is also obtained by taking the *kth* derivative of  $M_x(t)$  with respective to t and evaluating it at t = 0. That is,

$$E\left(x^{k}\right) = \frac{\partial^{k} M_{x}(t)}{\partial t^{k}} \Big|_{t=0} = U_{k}^{1} = M_{x}^{i}(0)$$

### **Application of L'Hospital Rule**

In evaluating  $M_x^1(t)$  which takes the form  $M_x^1(t) = \frac{h(t)}{q(t)}$  at t = 0 we may obtain the indeterminate,  $\frac{0}{0}$  or  $\frac{\infty}{\infty}$ . In such cases we apply the *L'Hopital Rule*, where

$$\lim_{t\to 0} M_x^1(t) = \lim_{t\to 0} \frac{h^1(t)}{q^t(t)}.$$

For example, if x is uniformly distributed over the interval [0, 1], then its moment generating function,  $M_x(t) = \frac{1}{t}(e^t - 1)$ .

Differentiating with respect to t and evaluating at t = 0;

$$M_x^1(t) = \frac{te^t - e^t + 1}{t^2} = \frac{h(t)}{q(t)}$$
 $M_x^1(0) = \frac{0}{0}$ 

which is indeterminate A applying the L'Hopital Rule we have,

$$\lim_{t \to 0} M_x^1(t) = \lim_{t \to 0} \frac{h^1(x)}{q^1(x)}$$

$$= \lim_{t \to 0} \frac{te^t + e^t - e^t}{2t}$$

$$= \lim_{t \to 0} \frac{te^t}{2t} = \lim_{t \to 0} \frac{1}{2} e^t = \frac{1}{2},$$

which is the mean value of the uniform distribution.

#### **Properties of Moment Generating Functions**

- Moment generating functions are unique.
- If x and y are random variables such that y = a + bx, then

$$M_{y}(t) \equiv E(e^{ty})$$

$$= E(e^{(a+bx)t})$$

$$= E(e^{at}.e^{btx})$$

$$= e^{at} E(e^{btx})$$

$$= e^{at}.M_{y}(bt)$$

### **Example 1**

Determine the moment generating functions for the random variables, *x* and *y* with the following distribution functions:

(i) 
$$f(x) = \begin{cases} kxe^{-2x}, & x \ge 0 \\ 0, & elsewhere \end{cases}$$