MATH 166: Introductory Probability and Statistics

Dr. Eric Nimako Aidoo

en.aidoo@yahoo.com

0202901980

Probability Distribution

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.

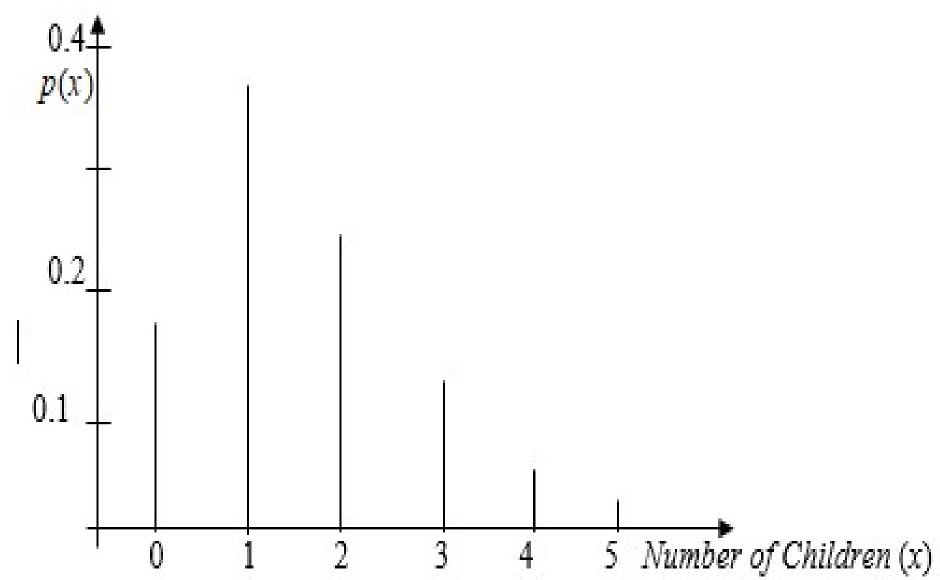
Discrete Distributions: The probability distribution for a discrete random variable *X* is a *formula, table, graph* or *any device* that specifies the probability associated with each possible value of *X*.

Probability: Discrete Distribution

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



Probability: Discrete Distribution

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. $0 \leq p(x) \leq 1$
3. $\sum p(x) = 1$

Probability: Discrete Distribution

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

Probability: Discrete Distribution

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

Probability: Discrete Distribution

Example

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

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

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

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

Probability: Discrete Distribution

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

$9k = 1$

$$\Leftrightarrow k = {}^{1}\!/\!_{9}$$

Probability: Continuous Distribution

- 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.

Probability: Continuous Distribution

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. $f(x) \geq 0$, for any value of x

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

3. $P(a \leq x \leq b) = \int_{a}^{b} f(x)\, dx$

Probability: Continuous Distribution

Example

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

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

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 $P(x > 2)$.

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

Probability: Continuous Distribution

Solution 1

Given the probability density function,

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

then,

(i) $$\int_0^3 f(x)dx = 1$$

$$\int_0^3 \left(\frac{1}{6}x + k\right)dx = 1$$

$$\frac{1}{12}x^2 + kx \Big|_0^3 = 1$$

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

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

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

Solution 1

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

$$P(1 \le x \le 2) = \int_1^2 \frac{1}{12}(2x+1)\,dx$$

$$= \frac{1}{12}\left[x^2 + x\right]_1^2$$

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

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

Probability: Continuous Distribution

Solution 2

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

$$\int_0^4 f(x)dx = 1$$

$$\int_0^3 kxdx + \int_3^4 3k(4-x)dx = 1$$

$$\left(\frac{kx^2}{2}\right)_0^3 + 3k\left(4x - \frac{x^2}{2}\right)_3^4 = 1$$

$$\frac{9k}{2} + 3k\,[(16-8)-(12-{}^9\!/_2)] = 1$$

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

$$6k = 1 \Leftrightarrow k = {}^1\!/_6$$

Probability: Continuous Distribution

Solution 2

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

$$P(1 \le x \le 2) = \int_1^2 f(x)dx$$

$$= \int_1^2 \frac{1}{6}x\,dx = \left| \frac{x^2}{12} \right|_1^2 = \frac{1}{12}\left(2^2 - 1\right) = \frac{1}{4}$$

$$P(x > 2) = \int_2^4 f(x)dx$$

$$= \int_2^3 \frac{1}{6}x\,dx + \int_3^4 \frac{1}{2}(4 - x)dx$$

Solution 2

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

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

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

Cumulative Distribution Function

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

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

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

Cumulative Distribution Function

Properties of Cumulative Distribution Function (CDF)

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

(i) $F(a) \leq F(b)$, wherever $a \leq b$, 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) $0 \leq F(x) \leq 1$

Cumulative Distribution Function

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

Cumulative Distribution Function

Solution

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

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

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

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

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

Cumulative Distribution Function

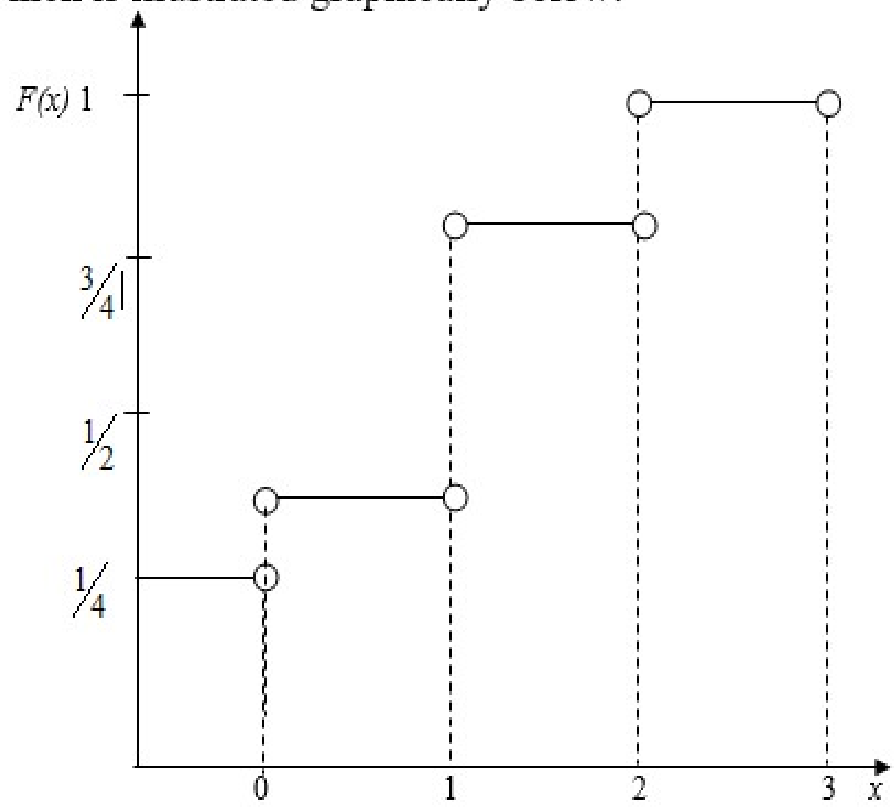
Solution

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

Hence the cumulative distribution is

X	0	1	2	3
$F(x)$	$\tfrac{1}{4}$	$\tfrac{3}{8}$	$\tfrac{7}{8}$	1

Which is illustrated graphically below:



Expectation and Variance of Random Variable

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 $-\infty \leq x \leq \infty$.

Expectation and Variance of Random Variable

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

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

(i) $$Var(x) = \sum_x (x-\mu)^2 p(x)$$

$$= \sum_x x^2 p(x) - \mu^2, \text{ if } x \text{ is discrete.}$$

Expectation and Variance of Random Variable

(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, σ = $\sqrt{Var(x)}$

Expectation and Variance of Random Variable

Example

Compute the expected value (μ) and standard deviation (σ^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) $$f(x) = \begin{cases} 6x(1-x), & 0 < x < 1 \\ 0 & , elsewhere \end{cases}$$

Expectation and Variance of Random Variable

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,

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

$$= (1-3)^2\,(0.1) + (2-3)^2\,0.3 + (3-3)^2\,(0.2) + (4-3)^2(0.3) + (5-3)^2(0.1)$$

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

Expectation and Variance of Random Variable

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,

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

Expectation and Variance of Random Variable

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

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

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

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

Expectation and Variance of Random Variable

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,

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

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

$$= \int_0^1 6x^3 (1 - x)\, dx - (0.5)^2$$

Expectation and Variance of Random Variable

Solution

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

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

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

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

Hence the standard deviation,

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

Expectation and Variance of Random Variable

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 λ and

ii) Use it to determine the mean and the standard deviation

Solution

To find λ we have,

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

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

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

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

Expectation and Variance of Random Variable

Solution

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

$$\frac{49}{8}\lambda = \frac{7}{8} \Leftrightarrow \lambda = \frac{1}{7}$$

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

Expectation and Variance of Random Variable

Solution

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

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

$$= \frac{1}{24} + \frac{1}{7} \left| \left(2y^2 - 1/3\, y^3\right) \right|_{1/2}^{4}$$

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

Expectation and Variance of Random Variable

Solution

For the standard deviation, σ, we have

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

$$= \int_0^{y_2} y^3\, dy + \int_{1/2}^{4} \frac{1}{7} y^2 (4-y)\, dy - (1.5)^2$$

$$= \left[\frac{1}{4} y^4\right]_0^{1/2} + \frac{1}{7}\left[\left(\frac{4}{3} y^3 - \frac{1}{4} y^4\right)\right]_{1/2}^{4} - 2.25$$

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

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 λ 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, $P(x > 1,050)$.

Moment and Moment Generating Function

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

$= \int_x \ g(x)f(x)$, if x is continuous

Moment and Moment Generating Function

The k^{th} Moment about the origin

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

and defined by

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

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

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.*

Moment and Moment Generating Function

The k^{th} Moment about the mean

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

$$U_k = E[(x-\mu)^k] = \sum_x (x-\mu)^k f(x) \;\; or \;\; \int_{R_x} (x-\mu)^k f(x)\,dx$$

Moment and Moment Generating Function

Uses of moments about the mean in statistical analysis

- $\boldsymbol{U_2 = E(x-\mu)^2}$ the second moment about the mean also known as variance.
- $\boldsymbol{U_3 = E(x-\mu)^3}$, 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 $a_3 < 0$)

Moment and Moment Generating Function

Uses of moments about the mean in statistical analysis

- $\boldsymbol{U_4 = E(x-\mu)^4}$ 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

Moment and Moment Generating Function

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,

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

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

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

Moment and Moment Generating Function

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 and Moment Generating Function

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:

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

Moment and Moment Generating Function

Expansion of e^{tx}

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

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

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

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

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

Moment and Moment Generating Function

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,

$$E(x^k) = \frac{\partial^k M_x(t)}{\partial t^k}\Big|_{t=0} = U_k^1 = M_x^i(0)$$

Moment and Moment Generating Function

Example 1

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

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