COURSE CODE: STS 331

COURSE TITLE: DISTRIBUTION THEORY 1

NUMBER OF UNIT: 3 UNITS

COURSE DURATION: THREE HOURS PER WEEK.

COURSE COORDINATOR: PROFESSOR O.E. ASIRIBO. asiribo@yahoo.com

LECTURER OFFICE LOCATION: DEPARTMENT OF STATISTICS

COURSE CONTENT:

Distribution function of random variables (r.v), Probability density function (p.d.f) – continuous and discrete, Cumulative Distribution function (CDF), Marginal and Conditional distributions, Joint distributions, Stochastic Independence, Derived distributions, Moments and Cumulants. Mathematical expectations, Moment generating function, Weak and strong laws of large numbers and Central limit theorem.

COURSE REQUIREMENTS:

This is a compulsory course for all statistics students. Students are expected to have a minimum of 75% attendance to be able to write the final examination.

READING LIST:

(1) Introduction to the Theory of Statistics by Mood, A.M, Graybill, F.A. and Boes, D.C.

(2) Introduction to Mathematical Statistics by Hogg R.V. and Craig A. T.

(3) Probability and Statistics by Spiegel, M. R., Schiller, J and Alusrinivasan, R..

LECTURE NOTES

Distribution of Random variable

Definition I:

Given a random experiment with a sample space □ , a function X which assign to each element c □ □, one and only one real number X(c) = x is called a <u>Random Variable.</u>

The space of X is the set of real numbers A = {x: x = X(c); c □ □}.

<u>Example:</u> Let the random experiment be the tossing of a single coin and let the sample space associated with the experiment be □ = {c: c is Tail or c is Head}.

Then X is a single value, real-value function defined on the sample space □

such that

X(c) = 0 if c is Tail

1 if c is Head

i.e A = {x: x = 0,1}.

X is a r.v. and the associated sample space is A.

<u>Definition 2:</u>

Given a random experiment with the sample space □. Consider two random variables X_1 and X_2 which assign to each element c of □ one and only ordered pair of numbers: X_1 (c) = x_1 , X_2 (c)= x_2.

The space of X_1 and X_2 is the set of ordered pairs.

A = {(x_1,x_2) : x_1 = X_1(c), x_2 = X_2(c), c □ □}

Definition 3:

Given a random experiment with the sample space □. Let the random variable X_i assign to each element c □ □, one and only one real no. X_i(c) = x_i, i = 1, . . ., n. the space of these random variables is the set of ordered n – turples.

$A = \{(x_1, x_2, . . . , x_n): x_1 = x_1(c), . . . , x_n = x_n(c), c$ □ □}

Probability Density function

Let X denote a r.v. with space *A* and let A C *A*, we can compute p(A) = p(x□A) for each A under consideration. That is, how the probability is distributed over the various subsets of Д. This is generally referred to as the probability density function (pdf).

There are two types of distributions, viz; discrete and continuous.

Discrete Density Function

Let X denote a r.v. with one dimensional space *A*. Suppose the space is a set of points s. t. there is at most a finite no. of points of *A* in any finite interval, then such a set *A* will be called a set of discrete points. The r.v. X is also referred to as a discrete r.v.

Note that X has distinct values $x_1, x_2, . . . , x_n$ and the function is denoted by f(x)

Where $f(x) = p\{X = x_i\}$ if $x = x_i$, $i = 1, 2, . . . , n$

$= 0$ if $x \neq x_i$

i.e $f(x) \geq 0$ x□ *A*

$\Sigma f(x) = 1$

CONTINUOUS DENSITY FUNCTION.

Let *A* be a one dimensional r.v; then a r.v. X is called continuous if there exist a function f(x)

s.t. $\int_A f(x)dx = 1$

where (1) $f(x) > 0$ x □ *A*

(2) f(x) has atmost a finite no. of discontinuity in every finite interval (subset of *A*)

or if $\mathcal{A}$ is the space for r.v. X and if the probability set function p(A), $A \subset \mathcal{A}$ can be expressed in terms of f(x)

s.t: $p(A) = p(x \in A) = \int_A f(x)dx$;

then x is said to be a r.v. of the continuous type with a continuous density function.

Cumulative distribution function

Let the r.v. X be a one dimensional set with probability set function p(A). Let x be a real value no. in the interval $-\infty$ to x which includes the point x itself, we have

$P(A) = P(x \in A) = P(X \leq x)$.

This probability depends on the point x, a function of x and it is denoted by

$F(x) = P(X \leq x)$

The function F(x) is called a cumulative distribution function (CDF) or simply referred to as distribution function of X.

Thus $F(x) = \sum_{w \leq x} f(w)$ (for discrete r.v. X)

And $F(x) = \int_{-\infty}^{x} f(w)dw$ (for continuous r.v. X)

Where f(x) is the probability density function.

Properties of the Distribution function, F(X).

a. $0 \leq F(x) \leq 1$ and $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$

b. $F_X(x)$ is a monotonic, non decreasing function i.e. $F(a) \leq F(b)$ for $a < b$

c. F(x) is continuous from the right i.e. lim F (x+h) = F(x) for h> 0 and h small

Note: ($F(x) = P(X \leq x)$ the equality makes it continuous from the right while without equality, it is from the left)

Assignment

1. The Prob. Dist. Function of time between successive customer arrival to a petrol station is given by

$$f(x) = 0 \qquad x < 0$$
$$= 10e^{-10x} \qquad 0 \leq x < \infty.$$

Find:

a. $P(0.1 < x < 0.5)$

b. $P(X < 1)$

c. $P(0.2 < x < 0.3 \text{ or } 0.5 < x < 0.7)$

d. $P(0.2 < x < 0.5 \text{ or } 0.3 < x < 0.7)$

MARGINAL AND CONDIDTIONAL DISTRIBUTION

Definition 1: Joint Discrete Density Function: If $(X_1, X_2, . . . , X_k)$ is a k-dimensional discrete r.v., then the joint discrete density function of $(X_1, X_2, . . . , X_k)$ denoted by $f_{X_1, X_2, \ldots X_k}(x_1, x_2, \ldots, x_k)$ and defined as

$$f_{X_1, X_2, \ldots X_k}(x_1, x_2, \ldots, x_k) = p(X_1 = x_1, X_2 = x_2, \ldots X_k = x_k,)$$

Note $\sum f_{X_1, X_2, \ldots X_k}(x_1, x_2, \ldots x_k) = 1$

Where the summation is over all possible value of $(X_1, X_2, . . . , X_k)$

Definition 2: Marginal Discrete Density Function: If X and Y are joint discrete r.v., then $f_X(x)$ and $f_Y(y)$ are called marginal discrete density functions. That is, if $f_{X,Y}(x, y)$ is a joint density function for joint discrete r.v. X and Y. then

$$f_X(x) = \sum_{y_i} f_{X,Y}(x, y) \quad \text{and} \quad f_Y(y) = \sum_{x_i} f_{X,Y}(x, y)$$

Also let (X, Y) be joint continuous r.v. with joint probability density function $f_{X,Y}(x, y)$, then

$$P[(X, Y) \in A] = \iint_A f_{X,Y}(x, y)\,dx\,dy$$

If $A = \{(x, y); a_1 < x \le b_1; a_2 < y \le b_2\}$, then

$$P[a_1 < x \le b_1; a_2 < y \le b_2] = \int_{a_2}^{b_2} \left[\int_{a_1}^{b_1} f_{X,Y}(x, y)\,dx \right] dy$$

Assignment

Given

$f(x, y) = x + y$ $\quad\quad$ (0 < x < 1; 0 < y < 1)

(a) Find p(0 < x < ½ , 0 < y < ¼)

If X and Y are joint continuous r.v. then, $f_X(x)$ and $f_Y(y)$ are called marginal probability functions, given by

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy \quad \text{and} \quad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx$$

(b) Find the marginal density function of Y and hence, obtain p(Y = 2).

CONDITIONAL DISTRIBUTION FUNCTION

Conditional discrete density function: Let X and Y be joint discrete r.v. with joint discrete density function $f_{X,Y}(x, y)$. The conditional discrete density function of Y given X = x denoted by $f_{Y/X}(y/x)$ is defined as

$$f_{Y/X}(y/x) = \frac{f_{X,Y}(x,y)}{f_X(x)}$$

where $f_X(x)$. is the marginal density of X at the point X= x.

similarly $$f_{X/Y}(x/y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}$$

$$= \frac{P[X = x, Y = y]}{P(Y = y)} = P\left(\frac{X = x / Y = y}{}\right)$$

Note that $\sum_y f_{Y/X}(y/x) = \sum_y \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$

$\Rightarrow$ that it is a probability density function.

The above definition also holds for the continuous case.

$$\int_{-\infty}^{\infty} f_{Y/X}(y/x)dy = \int_{-\infty}^{\infty} \frac{f_{X,Y}(x,y)dy}{f_X(x)} = \frac{f_X(x)}{f_X(x)} = 1$$

Stochastic Independence (S.I.)

Definition: Let $X_1, X_2, \ldots, X_k$ be a k- dimensional continuous (or discrete) r.v. with joint density function

$f_{X_1,X_2,\ldots X_k}(x_1,x_2,\ldots x_k)$ and marginal density function $f_{X_i}(x_i)$ then $X_1, X_2, \ldots, X_k$ are said to be stochastically independent iff

$$f_{X_1,X_2,\ldots X_k}(x_1,x_2,\ldots x_k) = \prod_{i=1}^{k} f_{X_i}(x_i) \qquad \forall x_i$$

for example, if r.v. X_i and X_2 have the joint density function $f_{X_1,X_2}(x_1,x_2)$ with marginal pdf $f_{X_1}(x_1)$ and $f_{X_2}(x_2)$ respectively, then X_1 and X_2 are said to be stochastically independent iff $f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$

Note that

$f(x_1,x_2) = f\left({}^{x_2}\!/_{x_1}\right) f\left(x_1\right)$ by earlier definition of conditional density

$\Rightarrow f\left({}^{x_2}\!/_{x_1}\right) = f\left(x_{\backslash 2}\right)$ iff X_1 and X_2 are independent.

Also recall that

$$\Rightarrow$$

$$f(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)dx$$

$$= \int_{-\infty}^{\infty} f\left({}^{x_2}\!/_{x_1}\right) f(x_1)dx_1$$

$$= f\left({}^{x_2}\!/_{x_1}\right) \int_{-\infty}^{\infty} f(x_1)dx_1$$

$$= f\left({}^{x_2}\!/_{x_1}\right) \qquad \textit{if } f\left({}^{x_2}\!/_{x_1}\right) \textit{ does not depend on } x_1$$

$$\Rightarrow f\left(x_1, x_2\right) = f\left({}^{x_2}\!/_{x_1}\right) f(x_1)$$

$$= f(x_2)\, f(x_1)$$

Exercise

Let the joint pdf of X_1 and X_2 be given as

$$f\left(x_1, x_2\right) = x_1 + x_2 \qquad 0 < x_1 < 1;.\ \ 0 < x_2 < 1$$
$$= 0 \qquad \textit{otherwise}$$

Show that X_1 and X_2 are stochastically dependent

Theorem: Let the r.v. X_1 and X_2 have the joint density function $f(x_1,x_2)$, then X_1 , X_2 are said to be stochastically independent iff $f(x_1,x_2)$ can be written as the product of non-negative function of x_1 alone and non-negative function of x_2 alone. i.e

$f(x_1,x_2) = g(x_1)h(x_2)$ where $g(x_1) >0$, $h(x_2) >0$

Proof:

If X_1 and X_2 are S.I, then $f(x_1,x_2)= f_1(x_1)f_2(x_2)$ where $f(x_1)$ and $f(x_2)$ are marginal density function of X_1 and X_2 respectively, i.e

$f(x_1,x_2) = g(x_1)h(x_2)$ is true

Conversely

If $f(x_1,x_2) = g(x_1)h(x_2)$, then for the r.v. of the continuous type, we have

$$f_1(x_1) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_2 = g(x_1)\int_{-\infty}^{\infty} h(x_2)dx_2 = c_1 g(x_1)$$

$$f_2(x_2) = \int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1 = h(x_2)\int_{-\infty}^{\infty} g(x_1)dx_1 = c_2 h(x_2)$$

Where c_1 and c_2 are constants and not functions of x_1 or x_2

But

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1)h(x_2)dx_1dx_2 = 1 \qquad \sin ce \quad a \quad pdf$$

$$\Rightarrow \int_{-\infty}^{\infty} g(x_1)dx_1 \int_{-\infty}^{\infty} h(x_2)dx_2 = c_1c_2$$

$$\Rightarrow c_1c_2 = 1$$

i.e $f(x_1,x_2) = g(x_1)h(x_2) = c_1\, c_2\, g(x_1)h(x_2) = c_1\, g(x_1)\, c_2\, h(x_2) = f_1(x_1)\, f_2(x_2)$ i.e X_1 and X_2 are S.I

<u>Theorem 2</u>: If X_1 and X_2 are S.I. with marginal pdf $f_1(x_1)$ and $f_2(x_2)$ respectively, then

$P(a<x_1<b, c<x_2<d) = p(a<x_1<b)p(c<x_2<d)$ for $a<b$ and $c<d$ and a, b, c, d are constants.

<u>Proof:</u> from definition of S.I of X_1 and X_2,

$$f(x_1, x_2) = f_1(x_1) f_2(x_2)$$

$$P(a < x_1 < b; c < x_2 < d) = \int_c^d \int_a^b f(x_1, x_2) dx_1 dx_2$$

$$= \int_c^d \int_a^b f_1(x_1) f_2(x_2) dx_1 dx_2$$

$$= \left[\int_a^b f_1(x_1)\right]\left[\int_c^d f_2(x_2)\right]$$

$$= P(a < x_1 < b) P(c < x_2 < d)$$

<u>Exercise</u>:

(a) Given f(x,y) = x + y

obtain $P\left(0 < x < \frac{1}{2}; 0 < y < \frac{1}{2}\right)$, $P\left(0 < x < \frac{1}{2}\right)$ *and* $P\left(0 < y < \frac{1}{2}\right)$ and hence show that X and Y are not S.I.

(b) Given f(x,y) = $e^{-(x+y)}$ $0<x<\infty$, $0<y<\infty$

Show that X and Y are independent.

<u>DERIVED DISTRIBUTIONS</u>

Consider a continuous r.v. X and the relation

Y = a + bx -------------------------------------- (1)

Since X is a r.v., so is Y.

Suppose we which to find the density function of Y. let f(x) be the density function of X where

$$f(x) \quad > 0 \qquad \propto < x < \beta$$
$$\quad = 0 \qquad elsewhere$$

If b > 0, then Y assumes values between a + bα and a + bβ, hence

$$P(Y \le y) = P(Y \le a + bx)$$

$$or \quad P(Y \le y) = P(a + bX \le y)$$

$$= P\left(X \le \frac{y-a}{b} \right) \text{..}(2)$$

If F(x) and G(y) are distribution functions of X and Y respectively, then

$$G(y) = F\left(\frac{y-a}{b} \right) \text{..}(3)$$

Since the density of Y, g(y) is given by $g(y) = \frac{dG}{dy}$

$$\Rightarrow g(y) = \frac{dF\left(\frac{y-a}{b} \right)}{dy} = \frac{d}{dy} \int_{-\infty}^{\frac{y-a}{b}} f(x)dx = \frac{1}{b} f\left(\frac{y-a}{b} \right)$$

The transformation given in (1) is known as one to one transformation.

Generalization of (1):

Let Y = ϕ(x) ------------------------------------(4)

since Y is a function of X, we can solve equation (4) for X to obtain X as a function of Y denoted by

X = Ψ(Y) ------------------------------------ (5)

= ϕ^{-1}(y)

The transformation in equations (4) and (5) are said to be 1 – 1 if for any value of x, ϕ(x) yields one and one value of Y and if for any value of Y, Ψ(Y) yields one and only one value of X.

<u>Theorem:</u> Let X and Y be continuous r.v. defined by the transformation

Y = ϕ(x) and X = $\varphi(Y)$

Let these transformations be either increasing or decreasing functions of X and Y and 1-1. If f(x) is the pdf of X where f(x) > 0 α<x<β and f(x) = 0 elsewhere

Then Pdf of Y is

$$g(y) = \left|\frac{d\phi(y)}{dy}\right| f(\phi(y)) \qquad \alpha_1 < y < \alpha_2$$

where $\alpha_1 = \min[\varphi(\alpha), \varphi(\beta)]$

$$\alpha_2 = \max[\varphi(\alpha), \varphi(\beta)]$$

Proof: Let $\alpha_1 = \phi(\alpha)$ and $\alpha_2 = \phi(\beta)$, in this case, $\phi(x)$ is an increasing function of X since $\alpha < \beta$ and

G(y) = p(Y ≤ y) = p($\phi(x) \le y$)

$= \text{P}[\text{X} \le \varphi(y)] = F(\varphi(y)) = \int_{-\infty}^{\varphi(y)} f(x)dx$

The density of Y is therefore given by

$$g(y) = \frac{dG(y)}{dy} = \frac{d}{dy}\int_{-\infty}^{\varphi(y)} f(x)dx$$

or

$$g(y) = \frac{d\varphi(y)}{dy} F(\varphi(y)) \qquad \phi(\alpha) \le y \le \phi(\beta)$$

Since $\phi(x)$ is an increasing function of x, hence $d/_{dy}\,\varphi(y) > 0$ which makes $g(y) \ge 0$.

Now suppose that $\phi(x)$ is an decreasing function of X, i.e. as X increasing $\phi(x)$ decreases. Thus the min of Y is $\phi(\beta)$ and maximum value of Y is $\phi(x)$.

$G(y) = p(Y \le y) = p(\phi(x) \le y)$
$= p(X \ge \varphi(y)) = 1 - F(\varphi(y))$

Hence the pdf of Y is given as

$$g(y) = \frac{d}{dy} G(y) = -\frac{d\varphi(y)}{dy} F(\varphi(y)) \qquad \phi(\beta) \le y \le \phi(\alpha)$$

Since $\phi(x)$ is a decreasing function of X, thus $\varphi(y)$ is a decreasing function of y and the partial derivative of $\varphi(y) < 0$.

i.e. $\frac{d}{dy}\varphi(y) < 0$

$\Rightarrow g(y) \ge 0$.

i.e $g(y) = \left|\frac{d\varphi(y)}{dy}\right| f(\varphi(y)) \qquad \alpha_1 < y < \alpha_2$

TRANSFORMATION OF VARIABLES OF DISCRETE TYFPE

Let X be a r.v. of discrete type with a pdf f(x). Let ***A*** denote the set of discrete points for which f(x) > 0 and let Y = $\nu(x)$ be a 1-1 transformation that mapped ***A*** onto β. Let $x = \omega(y)$ be the solution of $y = \nu(x)$, then for each $y \varepsilon \beta$, we have $x = \omega(y) \in \boldsymbol{A}$

$\Rightarrow$ event $Y = y[or\, \nu(x) = y]$ and $X = \omega(y)$ are equivalent

Thus

$$g(y) = p[Y = y] = p[X = \omega(y) = F(\omega(y))] \qquad y \in \beta$$

ASSIGNMENT

Given X to be a discrete r.v. with a poison distribution function, obtain pdf of Y = 4X

Let f(x_1, x_2) be the joint pdf of two discrete r.vs. X_1 and X_2 with set of points at which f(x_1, x_2) > 0. Define a 1-1 transformation such that $Y_1 = U_1(X_1, X_2)$ and $Y_2 = U_2(X_1, X_2)$, for which the joint pdf jis given by

$$g(y_1, y_2) = f(\omega_1(y_1, y_2), \omega_2(y_1, y_2)), y_1 y_2 \in \beta$$

$x_1 = \omega_1(y_1, y_2)$ and $x_2 = \omega_2(y_1, y_2)$ are the inverse of $y_1 = U_1(x_1, x_2)$ and $y_2 = U_1(x_1, x_2)$.

from the joint pdf $g(y_1, y_2)$, we then obtain the marginal pdf of y_1 by solving over y_2 and vice-versa.

TRANSFORMATION OF VARIABLES OF CONTINUOUS TYPE

Let X be a r.v. of continuous type with a pdf of f(x). Let $\boldsymbol{A}$ be a one dimensional space for f(x) > 0. Consider a 1-1 transformation which maps the set $\boldsymbol{A}$ onto set β. Let the inverse of $Y = v(x)$ be denoted by $x = w(y)$ and let the derivative $\frac{dx}{dy} = \omega'(y)$ be continuous and not vanishing for all points $Y \in \beta$. Then the points of $Y = U(x)$ is given by

$$g(y) = f(\omega(y))\left|\omega^1(y)\right| \qquad y \in \beta$$

$$= 0 \qquad \text{elsewhere}$$

$\left|\omega^1(y)\right|$ is called the Jacobian of the linear transformation $x = \omega(y)$ is denoted by $|J|$.

Exercise

Given X to be continuous with

Ca) $f(x) = 1 \qquad 0 < x < 1$

$= 0 \qquad elsewhere$

Show that $Y = -2\ln x$ has χ^2 distribution with 2 df.

(b) $f(x) = 2x \qquad 0 < x < 1$

$= 0 \qquad elsewhere$

find pdf of $Y = 8x^3$

(c) $f(x) = e^{-x}$ $\quad x > 0$

$\quad = 0 \quad$ elsewhere

find pdf of $Y = \sqrt{x}$

The method of finding the pdf of a function one r.v. can be extended to two or more r.v.s of continuous type.

Let $Y_1 = U_1\,(x_1, x_2)$

$Y_2 = U_1(x_1, x_2)$

Define a 1-1 transformation which maps a 2-dimensional set of A in the x_1, x_2 plane into 2-dimensional set of B in the y_1 *and* y_2 plane. If we express each of x_1, x_2 in terms of y_1 *and* y_2, we can write $x_2 = w_1(y_1, y_2)$ *and* $x_2 = w_2(y_1, y_2)$, and the determinant of order 2 can be obtained

$$J = \begin{vmatrix} \frac{dx_1}{dy_1} & \frac{dx_1}{dy_2} \\ \frac{dx_2}{dy_1} & \frac{dx_2}{dy_2} \end{vmatrix}$$ *Known as the Jacobian of transformation*

It is assumed that these first order partial derivatives are continuous and that J is not identically equal to zero in B.

Exercise 1;

Given the r.v. X with

$f(x) = 1 \qquad\qquad 0 < x < 1$

$0 = \qquad\qquad elsewhere$

Let X_1 and X_2 denote random samples from the distribution. Obtain the marginal density function of $Y_1 = X_1 + X_2 \quad and \quad Y_2 = X_1 - X_2$ respectively.

Exercise 2 :

Let X_1 and X_2 be a r.s. from an exponential distribution of the form

$f(x) = e^{-x} \qquad\qquad 0 < x < \infty$

$= 0 \qquad\qquad elsewhere$

Given $Y_1 = X_1 + X_2$

$$Y_2 = \frac{X_1}{X_1 + X_2}$$

Show that Y_1 and Y_2 are S.I.

Mathematical Expectation

Let X be a r.v. with pdf f(x) and let V(x) be a function of x such that $\int_{-\infty}^{\infty} V(x) f(x) dx$ exists $\forall$ x (continuous r.v.) and $\sum V(x) f(x)$ exists if X is a discrete r.v. The integral or summation as the case may be is called the mathematical expectation or expected value of V(x) and it is denoted by E$[(x)]$. It is required that the integral or sum converge absolutely. More generally, let $x_1, x_2, \ldots x_n$ be a r.v. with pdf f($x_1, x_2, \ldots x_n$) and let

V($x_1, x_2,...x_n$) be a function of the variable such that the n-fold integrals exist, i.e. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}....\int_{-\infty}^{\infty}V(x_1,x_2,...x_n)f(x_1,x_2,...x_n)dx_1,dx_2,...dx_n$ exists, if the r.vs. are of continuous type and $\sum_{x_1}...\sum_{x_n}V(x_1,x_2,...x_n)f(x_1,x_2,...x_n)$ exists if the r.vs. are discrete.

The n-fold integrals or the n-fold summation is called the mathematical expectation denoted by $E[(V(x_1,x_2,...x_n))]$ of function f($x_1, x_2,...x_n$).

Properties of Mathematical Expectation

1) If k is a constant, then $E(k) = k$

2) if k is a constant and V is a function, then $E(kV) = kE(V)$

3) if k_1 and k_2 are constants and V_1 and V_2 are functions the $E(k_1V_1 + k_2V_2) = k_1E(V_1) + k_2E(V_2)$

Example:

f(x) = 2(1-x) 0<x<1

$$E(x) = \int_{-\infty}^{\infty} xf(x)dx = \int_0^1 2x(1-x)dx = 1/3$$

$$E(x^2) = \int_{-\infty}^{\infty} x^2 f(x)dx = \int_0^1 2x^2(1-x)dx = 1/6$$

$$V(x) = E(x^2) - [E(x)]^2 = \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{6} - \frac{1}{9} = \frac{3}{54}$$

$$E(6x + 3x^2) = 6E(x) + 3E(x^2) = 6(1/3) + 3(1/6)$$

$$= 2 + \frac{1}{2} = 2\,1/2$$

$$f(x) = \frac{x}{6} \qquad x=1,2,3,$$

$$E(X^3) = \sum_{x=1,2,3} x^3 f(x) = \frac{1}{6}(1+16+81) = \frac{98}{6}$$

f(x,y) = x + y   0<x<1, 0<y<1

$E(xy^2)$=?

$$\iint xy^2 f(x,y)dxdy$$

$$\int_0^1\int_0^1 xy^2(x+y)dxdy = \frac{17}{72}$$

Weak Law of Large Number(WLLN)

Let X_1, X_2, … be a set of independent r.v. distribution in the same form with mean μ.

Let $\overline{X_n}$ be the mean of the first n observation.

i.e. $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} x_i$

then $\overline{X_n}$ who has the mean μ.

The weak law of large numbers states that $\overline{X_n}$ becomes more and more narrowly dispersed about μ as n increase

i.e. $\lim_{n\to\infty} P\{|\overline{X_n} - \mu| > \in\} = 0$ $\in > 0$

if we assume that the variance of any X exist and equal to 6^2

then $v(\overline{X_n}) = \frac{\sigma^2}{n}$

chebyshev inequality,

$$P\left\{\left|\overline{X}_n - \mu\right| > \in\right\} \leq \frac{\sigma^2}{n \in^2}$$

or

$$P\left\{\left(\overline{X}_n - \mu\right)^2 > \in^2\right\} \leq \frac{\sigma^2}{n \in^2}$$

or

$$p\left\{\left|X_n - \mu\right| > \in \sigma\right\} = P\left\{\left(X_n - \mu\right)^2 \geq \in^2 \sigma^2\right\} \leq \frac{1}{\in^2}$$

$$P\left\{\left|X_n - \mu\right| < \in \sigma\right\} \geq 1 - \frac{1}{\in^2}$$

note

$$P\left[g(x) \geq k\right] \leq \frac{E(g(x))}{k} \forall k > 0$$

<u>Theorem</u>: Let g(x) be a non negative function of a r.v. X. If $E(g(x))$ exist, then for any +ve constant $\in$

$$P\left[g(x) \geq \in\right] \leq \frac{E(g(x))}{\in}$$

<u>Proof</u>:

Let $A = \{x : g(x) \geq \in\}$ and let f(x) be the pdf of X, then

$$E(g(x)) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$= \int_{A} g(x) f(x) dx + \int_{A^c} g(x) f(x) dx$$

$$A^c = \{x : g(x) < \in\}$$

But each integral on the RHS is non –ve then

$$E(g(x)) \geq \int_{A} g(x) f(x) dx$$

$$\geq \int_{A} \in f(x) dx \qquad \text{since } g(x) \geq \in$$

$$= \in \int_{A} f(x) dx = \in P\left[g(x) \geq \in\right]$$

i.e. $\frac{E[g(x)]}{\in} \geq P[g(x) \geq \in]$

or $P[g(x) \geq \in] \leq \frac{E[g(x)]}{\in}$ $\qquad \forall \in > 0$

Proof of Chebyshev inequality

$$P[g(x) \geq k] \leq \frac{E(g(x))}{K}$$

Let $g(x) = (X - \mu)^2$

$K = \in^2 \sigma^2$

i.e $P\left[(X - \mu)^2 \geq \in^2 \sigma^2\right] \leq \frac{\sigma^2}{\in^2 \sigma^2} = \frac{1}{\in^2}$

Proof of weak law of large number

let $g(x) = (\overline{X}_n - \mu)^2$

$K = \in^2$

$$P(|\overline{X}_n - \mu| > \in) = P\left[(\overline{X}_n - \mu)^2 > \in^2\right]$$

$$\leq \frac{E(\overline{X}_n - \mu)^2}{\in^2} = \frac{\sigma^2}{n \in^2}$$

$$\lim_{n\to\infty} P\left[\left|\hat{X}_n - \mu\right| > \in\right] \leq \lim_{n\to\infty} \frac{\sigma^2}{n \in^2} = 0$$

Note that WLLN can also be stated as

$$P\left\{\left|\overline{X}_n - \mu\right| < \in\right\} \geq 1 - \delta$$
or
$$P\left[- \in < \overline{X}_n - \mu < \in\right] \geq 1 - \delta$$

Again using Chebyshev inequality

Let $g(x) = (X_n - \mu)^2$ and $k = \in^2$

$$P\left[- \in < \overline{X}_n - \mu < \in\right] = P\left[\left|\overline{X}_n - \mu\right| < \in\right]$$
$$= P\left[\left|\overline{X}_n - \mu\right|^2 < \in^2\right] \geq 1 - \frac{E\left(\overline{X}_n - \mu\right)^2}{\in^2}$$
$$= 1 - \frac{\left(\frac{1}{n}\right)\sigma^2}{\in^2} \geq 1 - \delta$$

Where $\delta > \dfrac{\frac{1}{n}\sigma^2}{\in^2}$ or $n > \dfrac{\delta^2}{\delta \in^2}$.

Exercise 1:

Suppose that a sample is drawn from some distribution with an unknown mean and variance equal to unity. How large a sample must be taken in order that the probability will atleast 0.95 that the sample mean $\overline{X}_n$ will lie within 0.5 of the population mean?

Exercise 2:

How large a sample must be taken in order that you are 99% certain that $\overline{X}_n$ is within 0.56 of μ?

Strong Law of Large Numbers(SLLN)

The weak law of large numbers state a limiting property of sums of r.v. but the strong law of large numbers state something about the behavior of a sequence, $S_n = \sum_{i=1}^{n} x_i \quad \forall n$

If $X_1, X_2, \cdots, X_n$ are independent and identical with finite mean μ,

$$\frac{S_n}{n} \xrightarrow[n\to\infty]{a.s} \mu \qquad \text{(almost surely)}$$

This is known as the Strong Law of Large Numbers (SLLN).

SLLN implies WLLN.

Central Limit Theorem

Let $X_1, X_2, \cdots, X_n$ be a r.s. from a normal distribution with mean μ and variance σ^2, then the r.v.,

$$Z_n = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$$

approaches the standard normal distribution as n approaches infinity

$$Z_n \xrightarrow[n\to\infty]{} N(0,1)$$

Proof

Note that the moment generating function of a standard normal distribution is given as

$$M_X(t) = \ell^{\frac{1}{2}t^2}$$

Let $M_{Z_n}(t)$ denoted the mgf of Z_n. Hence we need to show that

$$M_{Z_n}(t) \xrightarrow[n\to\infty]{} M(t)$$

$$M_{Z_n}(t) = E\left(\ell^{tZ_n}\right) = E(\exp tZ_n) = E\left[\exp\left(t\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right)\right]$$

$$= E\left[\exp\left(t\frac{\frac{\sum X}{n}-\mu}{\sigma/\sqrt{n}}\right)\right] = E\left[\exp\tfrac{1}{n}\sum t\frac{(X_i-\mu)}{\frac{\sigma}{\sqrt{n}}}\right]$$

$$= E\left[\exp\left(\tfrac{t}{\sqrt{n}}\sum\frac{X_i-\mu}{\sigma}\right)\right] = E\left[\prod_{i=1}^{n}\exp\left(\tfrac{t}{\sqrt{n}}\frac{X_i-\mu}{\sigma}\right)\right]$$

Let $y_i = \dfrac{X_i-\mu}{\sigma}$

Then $M_{Z_n}(t) = E\left[\prod_{i=1}^{n}\exp\left(\tfrac{t}{\sqrt{n}}y_i\right)\right] = \prod_{i=1}^{n}E\left[\exp\left(\tfrac{t}{\sqrt{n}}y_i\right)\right]$

$$= \prod_{i=1}^{n}M_{y_i}\left(\tfrac{t}{\sqrt{n}}\right) = \prod_{i=1}^{n}M_{Y}\left(\tfrac{t}{\sqrt{n}}\right)$$

$$= \left[M_Y\left(\tfrac{t}{\sqrt{n}}\right)\right]^n$$

But r^{th} derivative of $M_y\left(\tfrac{t}{\sqrt{n}}\right)$ evaluated at t=0 gives the r^{th} moment about the mean of the density function divided by $(\sqrt{n})^r$

Note

$$M_X(t) = E\left(\ell^{tx}\right)$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots,$$

$$= \sum_{j=0}^{\infty}\frac{x^j}{j!}$$

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

$$M(t) = E[e^{tx}] = 1 + \frac{tE(x)}{1!} + \frac{t^2E(x^2)}{2!} + \frac{t^3E(x^3)}{3!} + \cdots,$$

$$M^1(0) = E(x) \qquad M^{11}(0) = E(x^2) \quad e.t.c$$

$$\Rightarrow M_Y\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{ty}{\sqrt{n}} + \frac{t^2y^2}{2n} + \frac{t^3y^3}{3!\left(\sqrt{n}\right)^3} + \cdots,$$

$$= 1 + \frac{tE(x-\mu)}{\sigma\sqrt{n}} + \frac{t^2E(x-\mu)^2}{2n\sigma^2} + \frac{t^3E(x-\mu)^3}{3!\left(\sigma\sqrt{n}\right)^3} + \cdots,$$

$$= 1 + \frac{t\mu_1}{\sigma\sqrt{n}} + \frac{t^2\mu_2}{2n\sigma^2} + \frac{t^3\mu_3}{3!\left(\sigma\sqrt{n}\right)^3} + \cdots,$$

$$= 1 + \frac{t\mu_1}{\sigma\sqrt{n}} + \frac{t^2\mu_2}{2\left(\sigma\sqrt{n}\right)^2} + \frac{t^3\mu_3}{3!\left(\sigma\sqrt{n}\right)^3} + \frac{t^4\mu_4}{4!\left(\sigma\sqrt{n}\right)^4} + \cdots,$$

$$But \ \mu_1 = E(X-\mu) = 0 \,;\, \mu_2 = E(X-\mu)^2 = \sigma^2$$

$$M_Y\left(\frac{t}{\sqrt{n}}\right) = 1 + \frac{t^2}{2n} + \frac{t^3\mu_3}{3!\left(\sigma\sqrt{n}\right)^3} + \frac{t^4\mu_4}{4!\left(\sigma\sqrt{n}\right)^4} + \cdots,$$

$$= 1 + \frac{1}{n}[\frac{t^2}{2} + \frac{t^3\mu_3}{3!\,\sigma^3\sqrt{n}} + \frac{t^4\mu_4}{4!\,n\sigma^4} + \cdots,$$

$$= 1 + \frac{Q}{n}$$

$$But \ \lim_{n\to\infty}\left(1 + \frac{Q}{n}\right)^n = e^{\frac{t^2}{2}}$$

$$Also \ M_{Z_n}(t) = \left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n = \left(1 + \frac{Q}{n}\right)^n$$

$$\lim_{n\to\infty} M_{Z_n}(t) = \lim_{n\to\infty}\left[M_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n$$

$$= \lim_{n\to\infty}\left(1 + \frac{Q}{n}\right)^n = e^{\frac{t^2}{2}}$$

Which is the same mgf for a standard normal distribution, hence $Z_n \xrightarrow[n\to\infty]{} N(0,1)$

Let X be a r.v. with pdf given by $f_X(x) = \lambda e^{\ \lambda x}$ $\quad 0 < x < \infty$

find the mgf of X and hence mean and variance of X.

$$M_X(t) = \mathrm{E}(e^{tx}) = \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx$$

$$= \int_0^{\infty} \lambda e^{-(\lambda - t)x} dx = \frac{\lambda}{\lambda - t} \qquad for\ t < \lambda$$

$$M'(t) = \frac{\lambda}{(\lambda - t)^2} \qquad hence \quad M'(0) = E(x) = \frac{1}{\lambda}$$

$$M(t) = \frac{2\lambda}{(\lambda - t)^3} \quad \Rightarrow M''(0) = E(X^2) = \frac{2}{\lambda^2},\ \sigma^2 = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \left[\frac{1}{\lambda}\right]^2 = \frac{1}{\lambda^2}$$

Assignment

(a) Find the mgf of the following distributions and hence the mean and variance.

Bernoulli, Binomial, Poisson, Geometric, Uniform, Exponential, Gamma, Beta, Normal, Chi-Square.

(b) Suppose $X \sim N(0, 1)$

Let $Y = X^2$. Find the distribution of Y.

(c) Assume X_1, X_1, . . ., X_n are i.i.d and distributed as exponential. Find the distribution of $S_n = \sum_{j=1}^{n} X_i$