VU LEARNING HUB

vulearninghub.xyz

HIGHLIGHTED HANDOUT

STA642

Probability Distributions

STA642 by VU Learning HUB

Prepared for VU students • Free study resources

https://vulearninghub.xyz

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

Topic No. 1

Random Experiment,

Outcomes, Sample Space & Events

Random Experiment

Any experiment or process in which the outcome is unpredictable that is eligible to be called a random

experiment.

Example 1:

Tossing of a Coin => a Head or a Tail.

However, until it lands on the floor, we do not know which of these two will turn up.

Example 2:

Survey of the Employees of an Organization:

Question on Marital Status:

Possible outcomes:

single, married, separated, divorced, widowed

However, until we obtain a reply from the respondent, we do not know what is his/her marital status.

Random Experiment

Any experiment or process in which the outcome is unpredictable that is eligible to be called a random experiment.

Outcomes

The various possible results of a random experiment.

Example 1: Head, Tail

Example 2: Single, Married, Separated, Divorced, Widowed

Sample Space

It is the set of all possible outcomes of a random experiment.

Example 1: Head , Tail

Example 2: Toss a single die 1,2,3,4,5,6}

Sample space will be the set of all of these taken together written as

S= {1,2,3,4,5,6}

In case of two fair dice, the sample space consists of the 36 ordered pairs:

i.e.{(1,1), (1,2), (1,3),…,(6,5), (6,6)}.

Event

Any subset of a sample space is called an Event.

Example 1

The set of ordered pairs which yield a sum which is an even number

denoted as A= {(1,1), (1,3), (1,5), (2,2), (2,4), (2,6),…,(6,6)}.

Example 2

The set of ordered pairs in which both the numbers are same

denoted as B= {(1,1), (2,2), (3,3), (4,4), (5,5), (6,6)}

A= {(1,1), (1,3), (1,5), (2,2),
(2,4), (2,6),…,(6,6)}.

B= {(1,1), (2,2), (3,3), (4,4),
(5,5), (6,6)}

Event B is a subset of event A.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 2

Definition of

Random Variable

Concept of a Random Variable

It is that variable which is associated with a random experiment.

Example:

Let the random experiment be the toss of a fair coin and the possible outcomes are Head and Tail. And the sample space associated with the experiment denoted by C be $C = \{H,T\}$, where H and T represent heads and tails, respectively.

Let X be a function such that $X(T) = 0$ and $X(H)=1$. Thus X is a real-valued function defined on sample space C.

That X is a function defined on the sample space C, $X(T) = 0$ and $X(H)=1$ which takes us from the sample space C to a space of real numbers $D= \{0,1\}$.

C is *H* and *T* and applying the function the image we obtain i.e the domain and range concept, So, if the domain here is H and T then the range will be 0 and 1.

Definition

Consider a random experiment with a sample space C. A function X, which assigns to each element $c \in C$ one and only one number $X(c) = x$, is called a **random variable**.

The **space** or **range** of X is the set of real numbers

$D = \{ x : x = X(c), c \in C \}$.

Given a random variable X, its range D becomes the space of interest. Besides inducing the space D, X also induces a probability which we call the distribution of X.

We are interested in knowing the probability of getting a 0 and 1, and as we already know that if it's a fair coin then the probability of getting a 0 is half and also for the probability of getting a 1 is half as well.

But, generally in this case, we can say that the set D is a finite or countable set or an interval of real numbers.

When a set is finite then, we are referring to discrete random variables. We call random variables of the first type **discrete** random variables, while those of the second type **continuous** random variables.

Discrete random variable has two situation, one is exactly the same I mentioned earlier where a set D is a finite set that has the elements {0, 1} or the other, if we toss a fair die we have the elements{1,2,3,4,5,6}.

The second situation ,which also gives you a discrete random variable that is when it is the **Countably infinite set.**

Example:

D is the set consisting of the numbers 1,2,3, and so on, which mean that they are countable but not finite. So, in this situation, we say that this is a Discrete random variable or if they are finite the elements may be called $d_1, d_2, \ldots, d_m$ then also it is a discrete random variable.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 3

Example of
Random Variable

Example:

Let a card be selected from an ordinary deck of playing cards. The outcome is one of these 52 cards. Suppose that P assigns a probability of 1/52 to each outcome c.

Let $X(c)=4$ if c is an ace, let $X(c)=3$ if c is a king, let $X(c)=2$ if c is a queen, let $X(c)=1$ if c is a jack and let $X(c)=0$ otherwise.

Now, we can develop a table with two columns.

The first column consisting of the values of the random variable X i.e. 0,1,2 , 3,4 and the second are the probabilities.

The classical definition of probability i.e. favorable over the total.
The sum of probability should be equal to 1.

x	$p_X(x)$
0	$\frac{36}{52}$
1	$\frac{4}{52}$
2	$\frac{4}{52}$
3	$\frac{4}{52}$
4	$\frac{4}{52}$
	$\sum_{x \in D} p_X(x) = \frac{52}{52} = 1$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 4

Discrete Random Variable

Definition

(Discrete Random Variable)

We say a random variable is a discrete random variable if its space is either finite or countable.

Example:

Consider a sequence of independent flips of a coin, each resulting in a head (H) or a tail (T). Moreover, on each flip, we assume that H and T are equally likely; that is, P(H)=P(T)=1/2.

x	$p(x)$
0	**1/2**
1	**1/2**
sum	**1**

This is that case, when the space is Finite.

Example of a Countably infinite set:

Let the random variable X equal the number of tosses needed to obtain the first head. The space of X is

D= {1,2,3,4,…}.

Consider that X=1 when the sequence begins with a H.

Likewise, X=2 when the sequence begins with TH.
In the same way when X=3 the sequence is TTH. See we are getting a head in the third toss.

Suppose that we are interested in finding the probabilities of the various possible values of X.

As shown in the table:

Although it is an infinite sequence of probabilities the sum has to be 1.

x	P(x)
1	$\left(\frac{1}{2}\right)^1$
2	$\left(\frac{1}{2}\right)^2$
3	$\left(\frac{1}{2}\right)^3$
4	$\left(\frac{1}{2}\right)^4$
.	.
.	.
.	.
∞	$\Sigma =$ 1

$$P(X = x) = \left(\frac{1}{2}\right)^{x-1}\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{x}, \qquad x = 1, 2, 3, \ldots$$

Here, X is the total no. of tosses that was required in order to get the first Head.

Now, let *X*=1 we see,

$$P(X=x)=\left(\frac{1}{2}\right)^{x-1}\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{x}, \qquad x=1,2,3,\ldots$$

$$x\in\{1 \text{ or } 2 \text{ or } 3 \text{ or} \ldots\}$$

$$P(X=1)=\left(\frac{1}{2}\right)^{1-1}\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{1}$$

$$P(X=2)=\left(\frac{1}{2}\right)^{2-1}\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{2}$$

$$P(X=3)=\left(\frac{1}{2}\right)^{3-1}\left(\frac{1}{2}\right)=\left(\frac{1}{2}\right)^{3}$$

Hence, the sequence of the probability will be

$$\left\{\left(\frac{1}{2}\right)^1+\left(\frac{1}{2}\right)^2+\left(\frac{1}{2}\right)^3+\ldots\right\}$$

Obviously, that is an infinite *geometric series*

$$a=\frac{1}{2},\ r=\frac{\left(\frac{1}{2}\right)^2}{\frac{1}{2}}=\left(\frac{1}{2}\right)$$

as we know that

$$Geometric\ series=\frac{a}{1-r}$$

$$P\left[X\in\{1,2,3,\ldots\}\right]=\frac{1/2}{1-(1/2)}=1$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 5

Concept of Probability Mass Function

What is Probability Mass Function?

Consider a random experiment with a sample space, finite or countably infinite. A function X, which assigns to each element $c \in C$ one and only one number $X(c) = x$, is called a random variable. If we express the probabilities against each random values of X algebraically, then, we can say that this is the probability mass function of X.

Example:

Toss a fair die, the possible values are 1,2,3,4,5,6.

Algebraically:

$$P(X = x) = \frac{1}{6}, \qquad x = 1, 2, 3, 4, 5, 6.$$

Definition

The probability mass function (pm*f*) of X is given by

$$p_X(x) = P[X = x], \quad for \quad x \in D. \qquad (1)$$

The PMFs satisfy two properties:

1) $0 \leq p_X(x) \leq 1, \quad x \in D$

2) $\sum_{x \in D} p_X(x) = 1 \qquad (2)$

CDF of a Discrete Random Variable

Example:

Suppose we roll a fair die with the numbers 1 through 6 on it. Let X be the upface of the roll. Then the space of X is $\{1,2,\ldots.,6\}$ and its pmf is

$$p_X(i) = 1/6, \quad for\ i = 1,2,...,6.$$

$$F(6) = P(X \leq 6) = 1/6$$

or

$$F(4) = P(X \leq 4) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6}$$

x	$p(x)$	$F(x)$
1	1/6	1/6
2	1/6	=1/6+1/6=2/6
3	1/6	=2/6+1/6=3/6
4	1/6	=3/6+1/6=4/6
5	1/6	=4/6+1/6=5/6
6	1/6	5/6+1/6=6/6=1

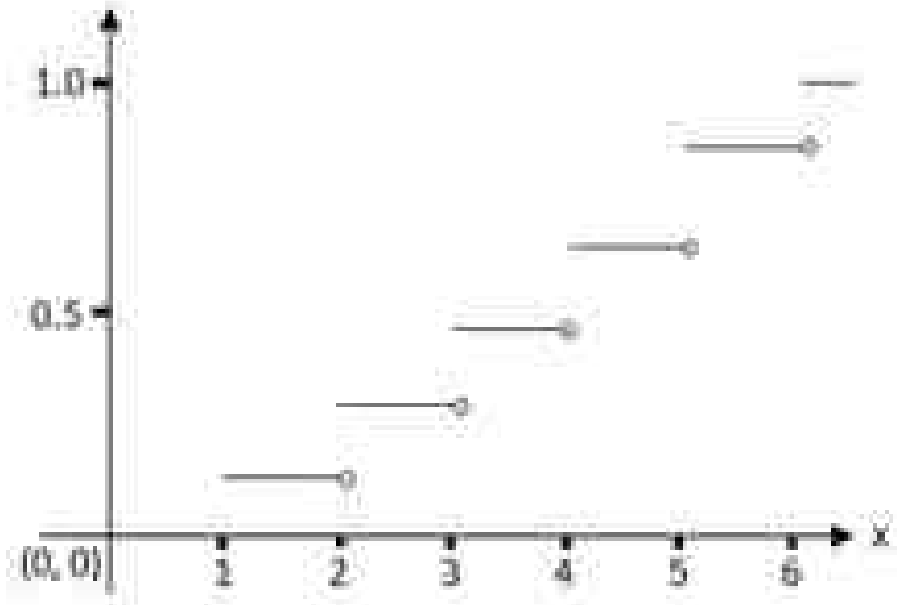


Figure 1: Distribution Function

The height of each 'Jump' yields the probability of that particular value of X.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 6

Example of Probability Mass Function

Example:

Consider and urn which contains slips of paper each with one of the numbers 1,2,…,100 on it. suppose these are *i* slips with the number *i* on it for *i* = 1,2,…, 100. For example, there are 25 slips of paper with the number 25. assume that the slips are identical except for the numbers. Suppose one slip is drawn at random. Let X be the number of slip.

a) Suppose that *X* has the pm*f* $p(x) = x/5050$, x =1,2,3,…,100, zero elsewhere

b) Compute $P(X \leq 50)$.

Solution:

a) Suppose that X has the pmf $p(x)$ $= x/5050, x = 1,2,3,\ldots,100,$ zero elsewhere.

i	
1	1 slip containing the no. 1
2	2 slip containing the no. 2
3	3 slip containing the no. 3
.	
.	
.	
100	100 slip containing the no. 100

Total number of slips= 1+2+3+…+100

‘Classical definition’ = m/n

The sum of the first n natural numbers $= \frac{n(n+1)}{2}$

Here, total no. of slips $= \frac{100(101)}{2} = 5050$

As the slips are identical i.e. equally likely to be drawn

Therefore, the probability $= \frac{m}{5050}$

X	Frequency (f)	$\frac{f}{\sum f} = p(x) = \frac{m}{n}$
1	1	1/5050
2	2	2/5050
3	3	3/5050
4	4	4/5050
.	.	.
.	.	.
.	.	.
100	100	100/5050
	5050	5050/5050=1

Now, to express mathematically,

$$P(X = x) = \frac{x}{5050}, \quad x = 1, 2, 3, ..., 100$$

b) Compute P(X≤ 50).

Solution:

$$P\left(X \le 50\right) = \frac{1+2+3+...+50}{5050}$$

or

$$P\left(X \le 50\right) = \frac{\frac{n\left(n+1\right)}{2}}{5050} = \frac{\frac{50\left(51\right)}{2}}{5050} = \frac{51}{202}$$

$$\therefore P\left(X \le 50\right) = 0.2525$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 7

Continuous Random Variable

Concept of a Continuous Random Variable

In the continuous case, we cannot define probability on one particular point, instead, it is the probability defined on an interval.

Example:

If we are measuring a height of a student. It depends on the refinement of that measuring instrument.

Suppose if we say that this person is 5 feet 4 inches tall, or in other words we are saying he is 64 inches tall. This is the measurement at a crude level.

- In reality it is a some value between 63.5 and 64.5, if we use a more refined instrument and after measuring we say that it is actually 64.2 inches. Even this is not the actual truth, it is the height somewhere between 64.15 and 64.25 inches.
- It all depend on the measuring instrument that how much precision can that instrument measures.
- Theoretically there can be an infinite number of decimal places after the decimals. This is the basic concept that variable can assume any value.

Cumulative Distribution Function $F(x)$

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$$

The function $f_X(t)$ is called a **probability density function (pdf)** of X. If $f_X(x)$ is also continuous, then the Fundamental Theorem of Calculus implies that

$$\frac{d}{dx} F_X(x) = f_X(x)$$

The PDFs satisfy the two properties:

1) $\int_{-\infty}^{\infty} f_X(x)\,dt = 1$ *and*

2) $f_X(x) \geq 0$

In the Continuous case, the area under the curve of that function gives you the probability.

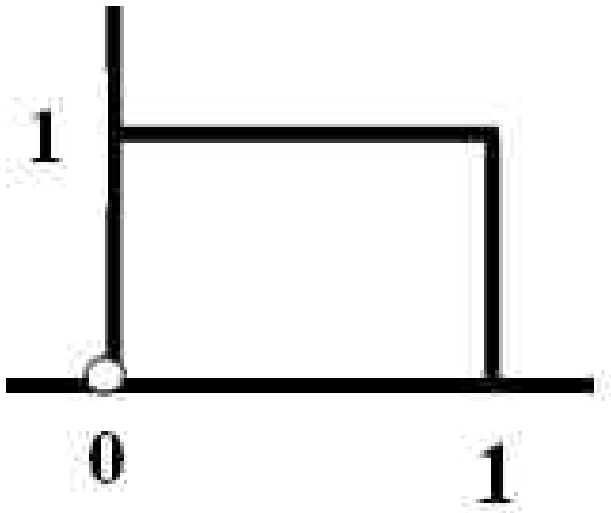
If you are interested in computing the probabilities, then, they can be obtained by integration, i.e.,

$$P\left(a < X \leq b\right) = \int_{a}^{b} f_X\left(t\right) dt = F_X\left(b\right) - F_X\left(a\right)$$

Example:

Suppose that, we randomly select a number between 0 and 1. Recall that the cdf of X is $F_X(x) = x$, for $0 < x < 1$. Hence the pdf of X is given by

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & elsewhere. \end{cases}$$



This is called a Uniform or rectangular distribution where, $0<x<1$.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 8

Example of Probability Density Function

Example:

In continuous case, consider the following simple experiment: choose a real number at random from the interval (0,1).

Let X be the number chosen. In this case the space of X is $D=(0,1)$.

The number is chosen at random it is reasonable to assign

$$P_X\left[(a,b)\right] = b - a,$$

$$for \quad 0 < a < b < 1.$$

It follows that the pdf of X is

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

For example, the probability that X is less than an eighth or greater than seven by eighths is

$$P\left[\left\{X < \frac{1}{8}\right\} \cup \left\{X > \frac{7}{8}\right\}\right] = \int_0^{\frac{1}{8}} dx + \int_{\frac{7}{8}}^{1} dx$$

$$= x\Big|_0^{\frac{1}{8}} + x\Big|_{\frac{7}{8}}^{1}$$

$$= \left(\frac{1}{8} - 0\right) + \left(1 - \frac{7}{8}\right)$$

$$= \frac{1}{8} + \frac{1}{8}$$

$$= \frac{2}{8}$$

$$P\left[\left\{X < \frac{1}{8}\right\} \cup \left\{X > \frac{7}{8}\right\}\right] = \frac{1}{4}$$

$$\therefore$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 9

Probability Generating Function

Probability Generating Function

- In probability theory, the **probability generating function** of a discrete random variable is a power series representation (the generating function) of the probability mass function of the random variable.

Definition

Univariate case

If X is a discrete random variable taking values in the non negative integers {0,1, ...}, then the *probability generating function* of X is defined as

$$G(z) = E\left(z^X\right) = \sum_{x=0}^{\infty} p(x) z^x,$$

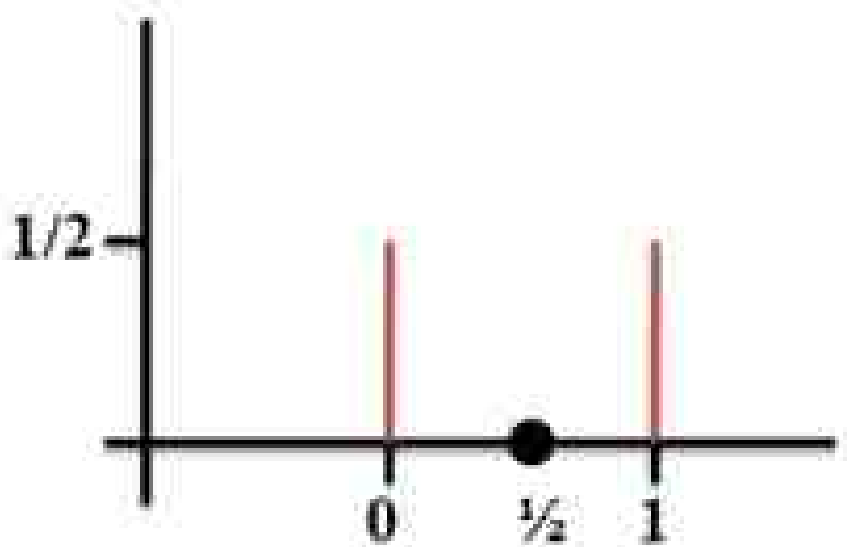
where p is the probability mass function of X.

Example:

The probability generating function of a Bernoulli random variable with parameter *p*. So the probability generating function of *X*, where *X* represents the number of heads that we obtain when tossing a fair coin once is given by

$$G(z) = \frac{1}{2} + \frac{z}{2}$$

x	p(x)	z^x	z^x p(x)
0	1/2	1	1/2
1	1/2	z	z/2
Sum			$\frac{1}{2} + \frac{z}{2}$



Example:

For Poisson Distribution $G(z) = e^{\lambda(z-1)}$

x	$p(x)=\frac{e^{-\lambda}\lambda^x}{x!}$	z^x	$z^x\,p(x)$
0	$\frac{e^{-\lambda}\lambda^0}{0!}=e^{-\lambda}$	$z^0=1$	$e^{-\lambda}$
1	$\frac{e^{-\lambda}\lambda^1}{1!}$	$z^1=z$	$e^{-\lambda}\lambda z$
2	$\frac{e^{-\lambda}\lambda^2}{2!}$	$z^2=z^2$	$\frac{e^{-\lambda}(\lambda z)^2}{2!}$
3	$\frac{e^{-\lambda}\lambda^3}{3!}$	$z^3=z^3$	$\frac{e^{-\lambda}(\lambda z)^3}{3!}$
.	.	.	.
.	.	.	.
.	.	.	.
∞	∞	∞	∞

$$\sum p(x)z^x = e^{-\lambda} + \frac{e^{-\lambda}(\lambda z)^1}{1!} + \frac{e^{-\lambda}(\lambda z)^2}{2!} + \ldots$$

$$\sum p(x)z^x = e^{-\lambda}\left[1 + (\lambda z)^1 + \frac{(\lambda z)^2}{2!} + \frac{(\lambda z)^3}{3!} \ldots\right]$$

$$G(z) = e^{-\lambda}\left[e^{\lambda z}\right] \qquad \therefore e^{\theta} = 1 + \theta + \frac{\theta^2}{2!} + \frac{\theta^3}{3!} + \ldots$$

$$= e^{-\lambda + \lambda z} = e^{\lambda(z-1)}$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 10

How to utilize Probability Generating Function

and

Mean and Variance through PGF

The **probability mass function** of X is recovered by taking derivatives of G,

$$p(k) = P(X = k) = \frac{G^{(k)}(0)}{k!} \qquad (1)$$

Example:

The probability generating function of a Bernoulli random variable with parameter *p*. So the probability generating function of *X*, where *X* represents the number of heads that we obtain when tossing a fair coin once is given by

$$G(z) = 1/2 + z/2$$

With the help of eq. (1), we take the derivative

$$\lim_{x\to\infty} G(z) = \frac{1}{2} + \frac{z}{2}$$

$$P(0) = \mathrm{P}(X = 0) = \frac{G^{(0)}(0)}{0!} = \frac{\left(\frac{1}{2} + \frac{z}{2}\right)}{0!}\Bigg|_{z=0} = \frac{1}{2}$$

$$P(1) = \mathrm{P}(X = 1) = \frac{G^{(1)}(0)}{1!} = \frac{\left(\frac{1}{2} + \frac{z}{2}\right)}{1!}\Bigg|_{z=0} = \frac{1}{2}$$

MEAN

and VARIANCE

through the

Probability generating function

The expectation of X is given by

$$E(X) = G'(1^-) \quad \textbf{(1)}$$

where,

$G'(1^-) = \lim_{z \to 1} G'(z)$ from below.

Example:

Poisson Distribution (Mean)

$$G(z) = e^{\lambda(z-1)}$$

$$\therefore G'(z) = \lambda e^{\lambda(z-1)}$$

$$G'(1^-) = \lim_{z \to 1} G'(z)$$

$$= \lambda e^{\lambda(1-1)} = \lambda e^{\lambda(0)} = \lambda e^{0}$$

$$\therefore G'(1^-) = \lambda$$

So the variance of X is given by

$$Var(X)$$
$$= G''(1^-) + G'(1^-) - \left[G'(1^-)\right]^2 \qquad \textbf{(2)}$$

Example:

Poisson Distribution (Variance)

$$G''(z) = \lambda.\lambda e^{\lambda(z-1)} = \lambda^2 e^{\lambda(z-1)}$$

$$\therefore G''(1^-) = \lambda^2 e^{\lambda(1-1)} = \lambda^2 e^{\lambda(0)} = \lambda^2$$

$$G'(1^-) = \lambda$$ ←From eq. (2)

$$\therefore V(X) = \lambda^2 + \lambda - [\lambda]^2$$

$$\therefore V(X) = \lambda$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 11

Algebraic Expressions of Some Well-known PGFs

For some of the more common distributions, the PGFs are as follows:

(i) Constant r.v. -

if $p_c = 1,\ p_k = 0, k \neq c,$ *then*

$$G_X(z) = E\left(z^X\right) = z^c$$

(ii) *Bernoulli* r.v. - if

$$p_1 = p, \quad p_0 = 1 - p = q, \quad p_k = 0, \quad k \neq 0 \text{ or } 1,$$

then

$$G_X(z) = E\left(z^X\right) = q + pz$$

(iii) *Grometric* r.v. - if

$$p_k = pq^{k-1}, \quad k = 1, 2, \ldots; q = 1 - p, \quad then$$

$$G_X(z) = \frac{pz}{1 - qz} \; if \; |z| < q^{-1}$$

$$(iv)\ \ Binomial\ \ r.v. -\ \ if$$

$$X \sim Bin(n, p),\ \ then$$

$$G_X(z) = (q + pz)^n, \qquad (q = 1 - p)$$

(v) *Poisson r.v. – if*

$X \sim Poisson(\lambda)$, *then*

$$G_X(z) = \sum_{k=0}^{\infty} \frac{1}{k!} \lambda^k e^{-\lambda} z^k = e^{\lambda(z-1)}$$

(vi) *Negative binomial r.v. – if*

$X \sim \text{NegBin}(n, p)$, *then*

$$G_X(z) = \sum_{k=0}^{\infty} \binom{k-1}{n-1} p^n q^{k-n} z^k$$

$$= \left(\frac{pz}{1-qz} \right)^n \text{ if } |z| < q^{-1} \text{and } p+q=1$$

Discrete Uniform Distribution

Example:

Random selection of one of the ten digits 0,1,2,...,9

x	p(x)	x^2	x^2 p(x)
0	1/10	x^2	1/10
1	1/10	$x^2 = x$	x/10
2	1/10	x^2	x^2/10
3	1/10	x^2	x^2/10
.	.	.	.
.	.	.	.
.	.	.	.
9	1/10	x^2	x^2/10
	10/10=1		

06:50

Therefore,

$$G(z)=\frac{1}{10}\left[1+z+z^2+z^3+\ldots+z^9\right]$$

This is a finite geometric series in which the first term a is 1 and the common ratio $r=z$ and $n=10$

So,

$$Sum=\frac{a\left(1-r^n\right)}{1-r}=\frac{1\left(1-z^{10}\right)}{1-z}$$

$$G(z)=\frac{1}{10}\left(\frac{1-z^{10}}{1-z}\right)$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 12

A Linear Combination
of
PDFs is a PDF

In mathematics , a linear combination is an expression constructed from a say of terms by multiplying each term by a constant and adding those results

(e.g. a linear combination of x and y would be any expression of the form ax+by, where a and b are constants).

Let,

1: Consider k continuous-type distributions with the following characteristics:

pdf $f_i(x)$,

mean μ_i, i=1,2,...k.

If $c_i \geq 0,\ i = 1, 2, ..., k,$

and

$c_1 + c_2 + ... + c_k = 0$

$w(x) = c_1 f_1(x) + c_2 f_2(x) + ... + c_k f_k(x) \geq 0$

To prove

Let, $\int_{-\infty}^{\infty} w(x)\,dx = 1$

Proof :

$$L.H.S = \int_{-\infty}^{\infty} \left[c_1 f_1(x) + c_2 f_2(x) + \ldots + c_k f_k(x) \right] dx$$

$$= \int_{-\infty}^{\infty} c_1 f_1(x)\,dx + \int_{-\infty}^{\infty} c_2 f_2(x)\,dx + \ldots + \int_{-\infty}^{\infty} c_k f_k(x)\,dx$$

$$\int_{-\infty}^{\infty} w(x)\,dx = c_1 \int_{-\infty}^{\infty} f_1(x)\,dx + c_2 \int_{-\infty}^{\infty} f_2(x)\,dx + \ldots + c_k \int_{-\infty}^{\infty} f_k(x)\,dx \quad (1)$$

$$But \quad f_i(x) \, is \, a \, pdf$$

$$\int_{-\infty}^{\infty} f_i(x)\,dx = 1$$

$$eq.(1) \, re-written$$

$$L.H.S = c_1(1) + c_2(1) + \ldots + c_k(1)$$

$$= c_1 + c_2 + \ldots + c_k$$

$$But \, by \, our \, choice$$

$$c_1 + c_2 + \ldots + c_k = 1$$

$$\therefore L.H.S = 1 = R.H.S$$

Show that the mean of the distribution having

pdf $c_1 f_1(x) + c_2 f_2(x) + \ldots + c_k f_k(x)$ *are* mean $\mu = \sum_{i=1}^{k} c_i \mu_i$

$E(\textit{new density}) \int_{-\infty}^{\infty} w(x)\,dx = 1$

$$E(X) = \int_{-\infty}^{\infty} x\left[c_1 f_1(x) + c_2 f_2(x) + \ldots + c_k f_k(x)\right] dx$$

$$= \int_{-\infty}^{\infty} \left[c_1 x f_1(x) + c_2 x f_2(x) + \ldots + c_k x f_k(x)\right] dx$$

$$= c_1 \int_{-\infty}^{\infty} x f_1(x)\,dx + c_2 \int_{-\infty}^{\infty} x f_2(x)\,dx + \ldots + c_k \int_{-\infty}^{\infty} x f_k(x)\,dx$$

$$E(X) = c_1\mu_1 + c_2\mu_2 + \ldots + c_k\mu_k$$

Hence, we get

$$\mu = \sum_{i=1}^{k} c_i \mu_i$$

In other words, the mean of a linear combination of PDFs is that same Linear combination of the means of those PDFs

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 13

Concept of

Cumulative Distribution Function

(discrete and continuous)

Definition: (Cumulative Distribution Function)

Let X be a random variable. Then its cumulative distribution function (cdf) is defined by $F_X(x)$, given in eq. (1) which you can see on the screen

$$F_X(x) = P_X\left((-\infty, x]\right) = P\left(\left\{c \in C : X(c) \leq x\right\}\right). \quad (1)$$

As above, we shorten

$$P\left(\left\{c \in C : X\left(c\right) \leq x\right\}\right) \text{ to } P\left(X \leq x\right).$$

Also,

$F_X(x)$ is often called simply the **distribution function (df)**.

- The graph for the discrete case it is like a staircase, which is also called a **Step function**.

- On the other hand, in the continuous case it is a continuous **curve or a line**.

For example:

Let $p_X(x)$ be the pmf of a random variable X. Find the cdf $F(x)$ of X and sketch its graph with that of $p_X(x)$ if:

a) $p_X(x) = 1/3$, x = -1, 0, 1, zero elsewhere.

x	p(x)	F(x)
-1	$\frac{1}{3}$	$\frac{1}{3}$
0	$\frac{1}{3}$	$\frac{1}{3}+\frac{1}{3}=\frac{2}{3}$
1	$\frac{1}{3}$	$\frac{2}{3}+\frac{1}{3}=\frac{3}{3}=1$

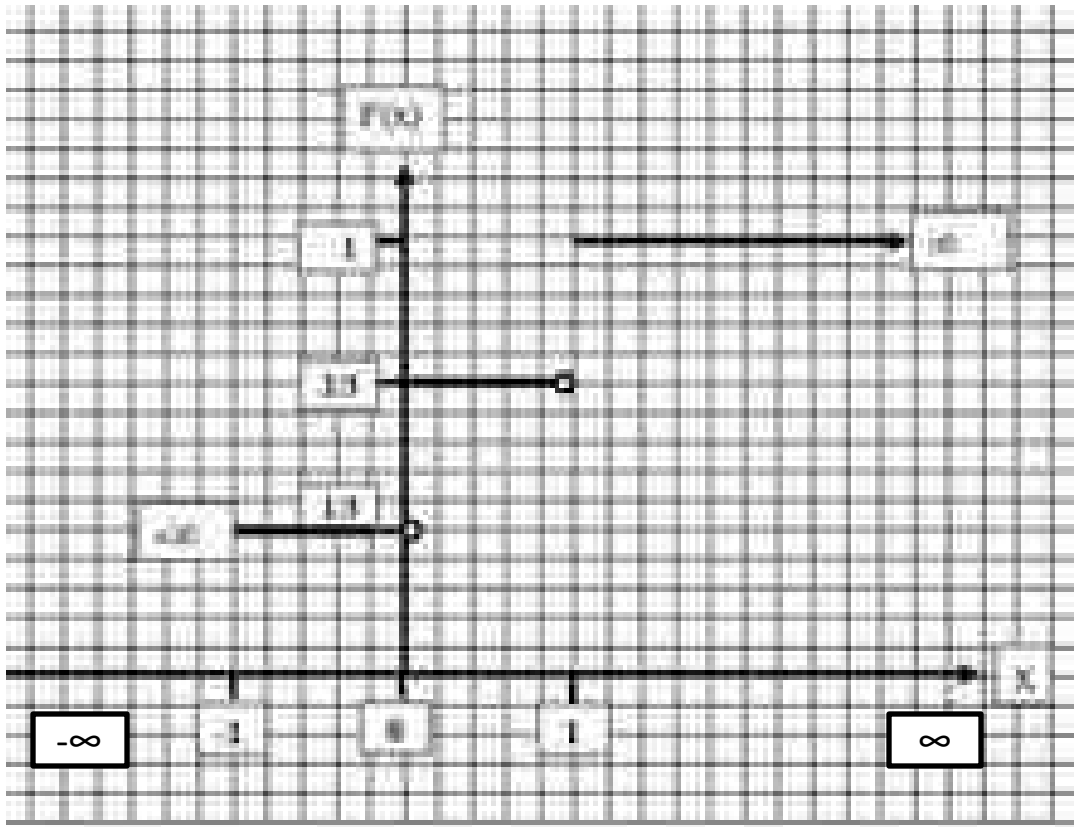


Figure 1: Distribution Function

Example 2:

Let X denotes a real number chosen at random between 0 and 2. we now obtain the cdf of X. First, if $x < 0$, then $P(X \leq x) = 0$.

Next, $x \geq 2$, then $P(X \leq x) = 1$. Finally, if $0 < x < 2$ it follows from equation 2 that $P(X \leq x) = P(0 < X \leq x) = x - 0 = x$. Hence the cdf of X is

$$F_X(x) = \begin{cases} 0 & if \quad x < 0 \\ x/2 & if \quad 0 \leq x < 2 \\ 1 & if \quad x \geq 2 \end{cases} \qquad (3)$$

$$P_X\left[(a,b)\right] = b - a, \qquad for \;\; 0 < a < b < 1. \qquad (2)$$

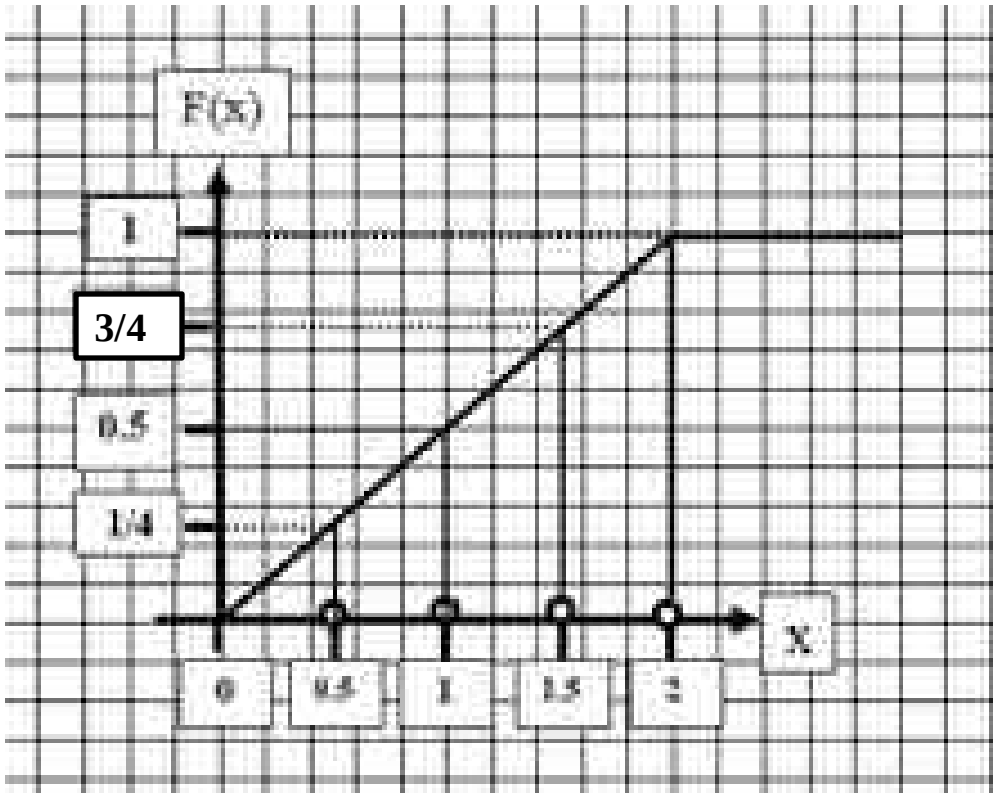


Figure 2: Distribution Function

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 14

Example of the CDF of a Discrete Random Variable

Example:

Consider an urn which contains slips of paper each with one of the numbers $1,2,\ldots,100$ on it. suppose there are i slips with the number i on it for $i = 1,2,\ldots,100$. or example, there are 25 slips of paper with the number 25.

Assume that the slips are identical except for the numbers. Suppose one slip is drawn at random.

Let X be the number on the slip

Show that the cdf of X is $F(x) = [x]([x]+1)/10100$, for $1 \leq x \leq 100$, where $[x]$ is the greatest integer in x.

Let X be the number on the slip (a)

x	$p(x)$
1	1/5050
2	2/5050
3	3/5050
- - -	- - -
100	100/5050
Sum	5050/5050=1

x	p(x)	F(x)
1	$\frac{1}{5050}$	$\frac{1}{5050}$
2	$\frac{2}{5050}$	$\frac{(1+2)}{5050}$
3	$\frac{3}{5050}$	$\frac{(1+2+3)}{5050}$
4	$\frac{4}{5050}$	$\frac{(1+2+3+4)}{5050}$
.	.	.
.	.	.
.	.	.
100	$\frac{100}{5050}$	$\frac{(1+2+\ldots+100)}{5050}$

Show that the cdf of X is $F(x) = [x]([x]+1)/10100$, for $1 \leq x \leq 100$, where $[x]$ is the greatest integer in x.

Therefore,

$$F(x) = \frac{[x]([x]+1)}{2 \times 5050}$$

$$F(x) = \frac{[x]([x]+1)}{10100}$$

Virtual University of Pakistan
Probability Distributions

by

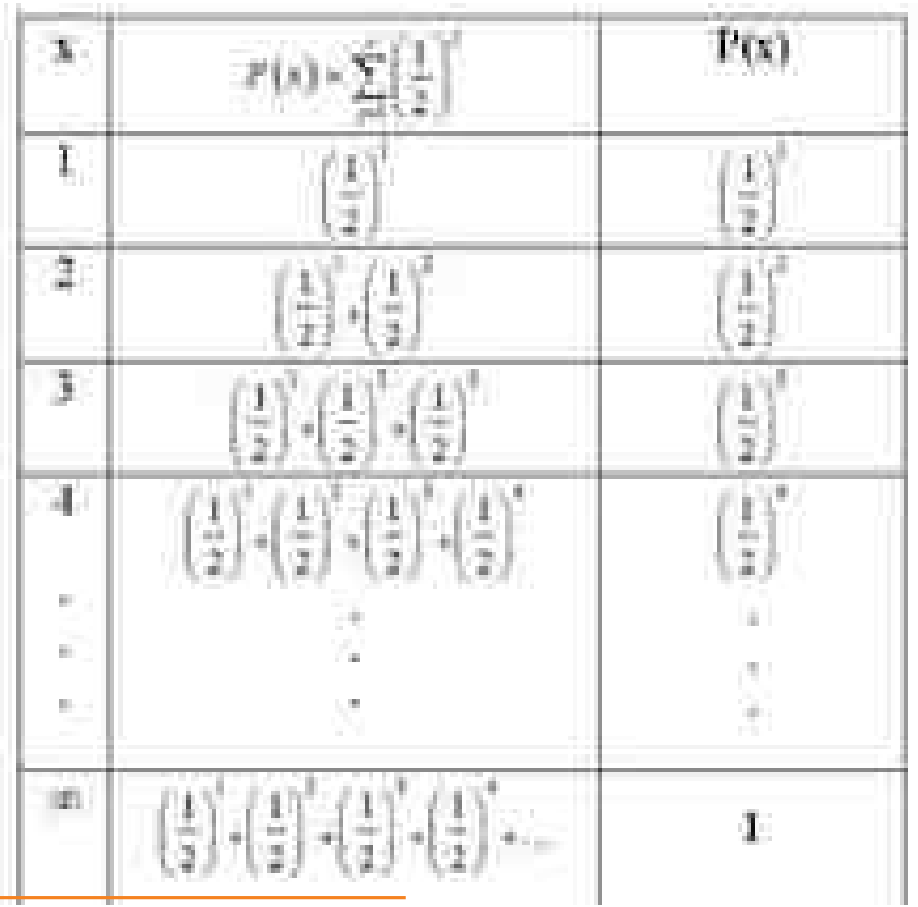
Dr. Saleha Naghmi Habibullah

Topic No. 15

Obtaining the PMF from the CDF

For each of the following CDFs $F(x)$, find the PDF of $f(x)$:

(a) $$F(x) = \sum_{j=1}^{x} \left(\frac{1}{2}\right)^j .$$



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 16

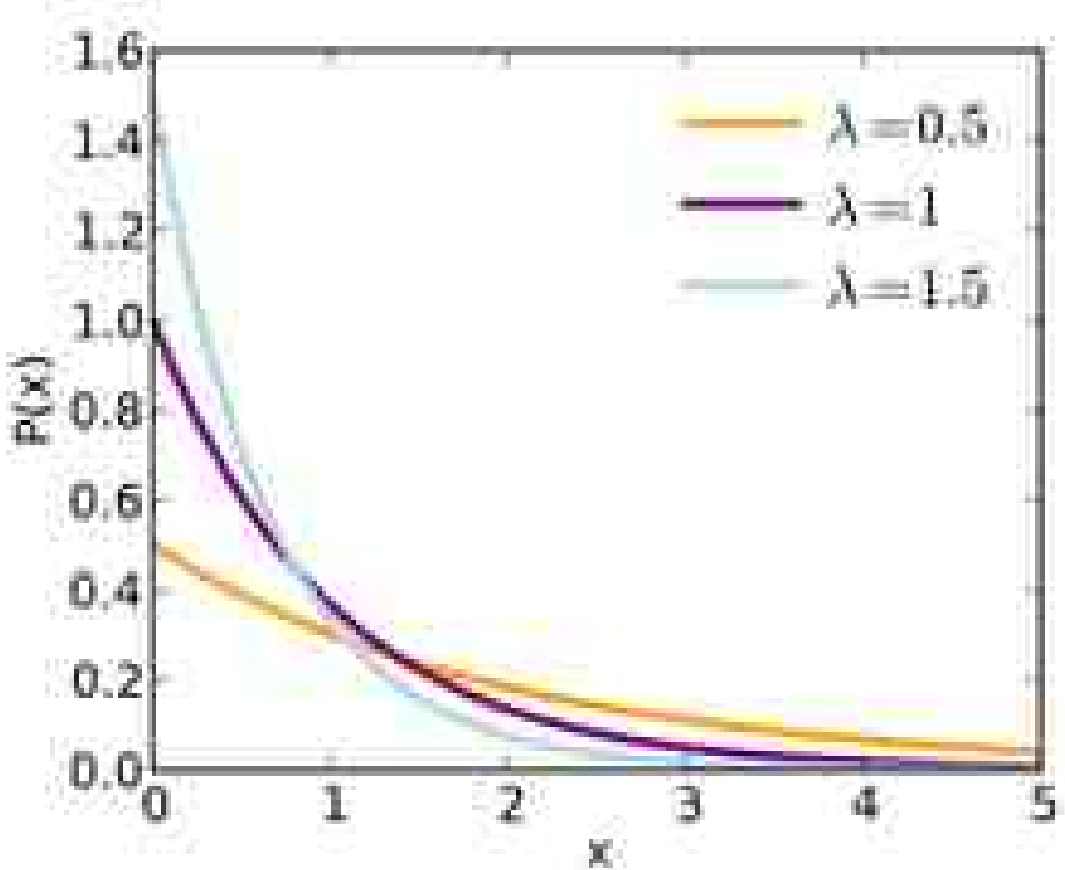
Example of the CDF of a Continuous Random Variable

For a continuous random variable X ,the CDF is as follows:

- $F(x) = 0$ for all x values which are less than zero
- $F(x) = x$ for all those x values that lies between 0 and 1.
- $F(x) = 1$ for all those x values are greater than 1.

PDF of the **Exponential Distribution** with mean=0.5, 1, 1.5

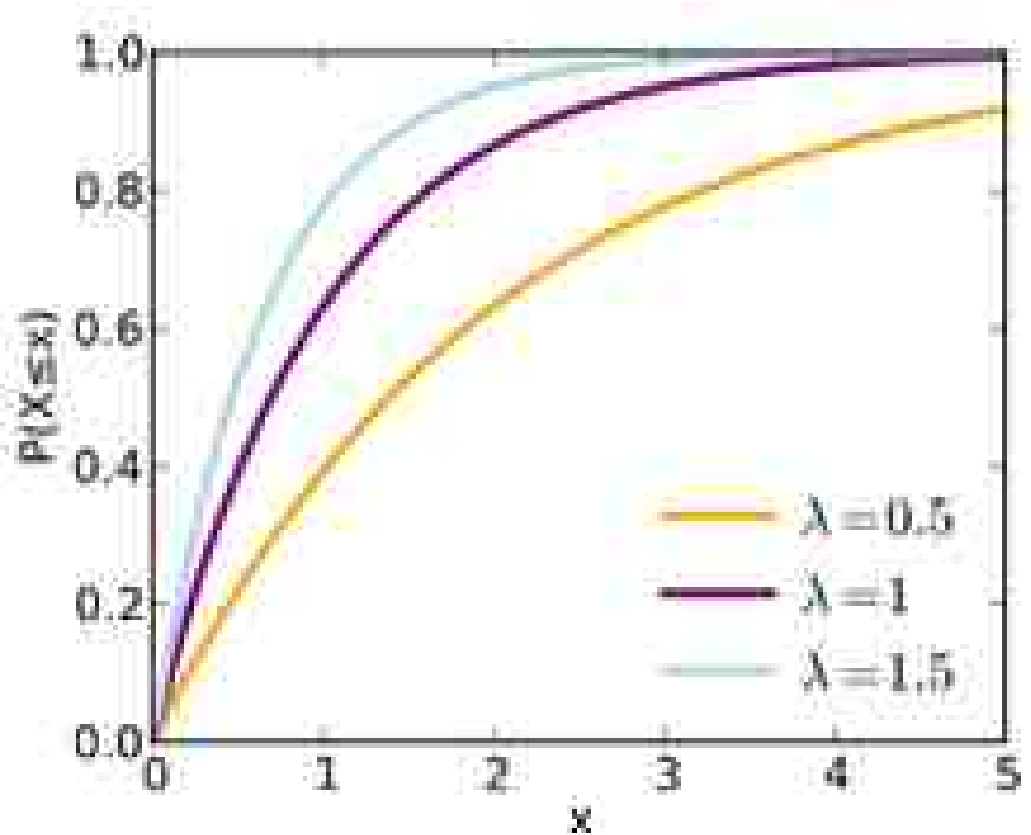
- $f(x) = \left(\frac{1}{0.5}\right) e^{-\left(\frac{x}{0.5}\right)}, \quad 0 < x < \infty$ → orange
- $f(x) = e^{-(x)}, \quad 0 < x < \infty$ → purple
- $f(x) = \left(\frac{1}{1.5}\right) e^{-\left(\frac{x}{1.5}\right)}, \quad 0 < x < \infty$ → skyblue


1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
P(x)
0
1
2
3
4
5
x
λ=0.5
λ=1
λ=1.5

CDF of the **Exponential Distribution** with mean=0.5, 1, 1.5

$$F(x) = 1 - e^{-x}, \quad 0 < x < \infty$$

- $F(x) = 1 - e^{-\frac{x}{0.5}}, \quad 0 < x < \infty$
- $F(x) = 1 - e^{-x}, \quad 0 < x < \infty$
- $F(x) = 1 - e^{-\frac{x}{1.5}}, \quad 0 < x < \infty$



Virtual University of Pakistan
Probability Distributions

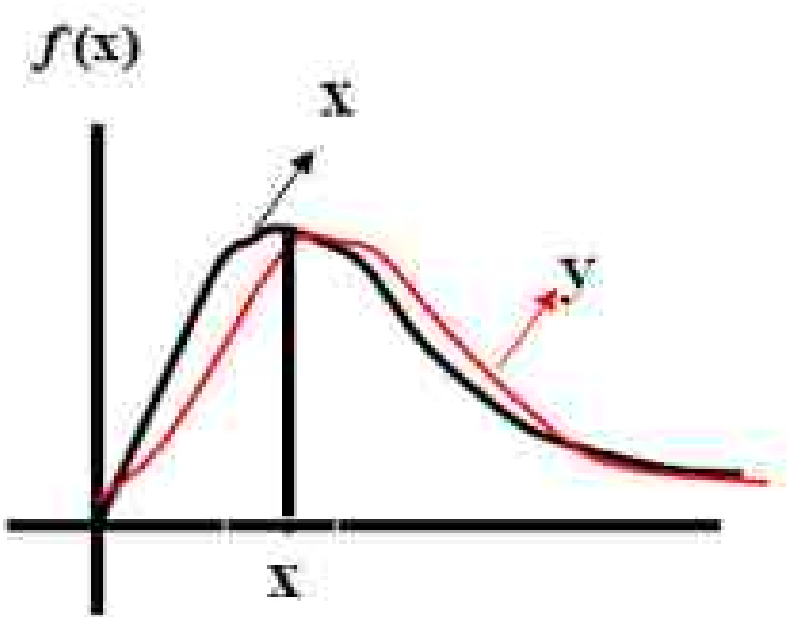
by

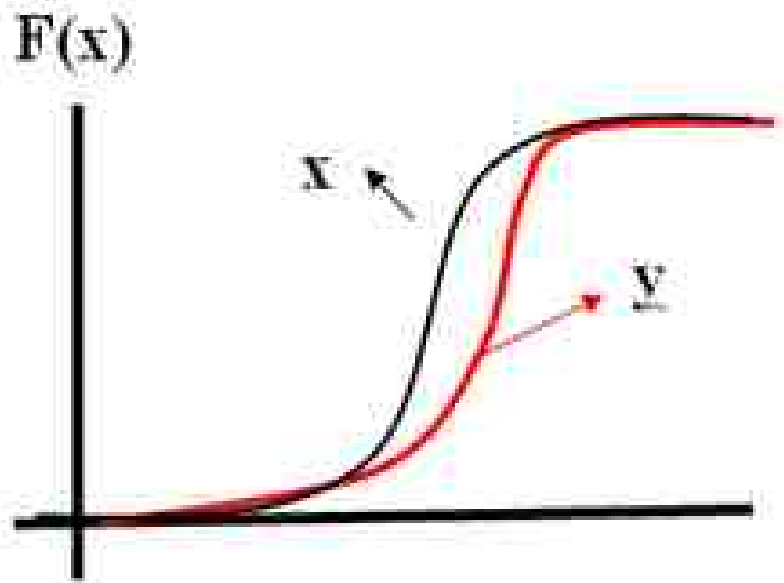
Dr. Saleha Naghmi Habibullah

Topic No. 17

Concept of two Random Variables being Equal in Distribution

Let us first consider that situation when two random variables X and Y are not equal in distribution





$$F_X(x) \neq F_Y(x)$$

We say that X and Y are equal in distribution $X \overset{D}{=} Y$ if and only if

$$F_X(x) = F_Y(x), \text{ for all } x \in R.$$

Mathematically we express it as $X \overset{D}{=} Y$

It is important to note while X and Y may be equal in distribution, they may be quite different.

Example

Recall that X denotes a real number chosen at random between 0 and 1. we now obtain the cdf of X. First, if x<0, the P(X≤ *x*) = 0.

Next, if x ≥1, then P(X≤ *x*) = 1. Finally, 0<x<1, it follows from expression (5.3) that P(X≤ *x*)= P(0<X≤ *x*)= x-0=x.

Hence the cdf of X is

$$F_X(x) = \begin{cases} 0 & if\ x < 0 \\ x & if\ 0 \le x < 1 \\ 1 & if\ x \ge 1 \end{cases} \qquad (5.6)$$

For instance, in the above example define the random variable Y and transform it as $Y = 1-X$.

Then $Y \neq X$.

For example, let, $X = 0.2$ then, $Y = 1-0.2 = 0.8$.

So, when X is 0.2 then Y is 0.8 which is not the same.

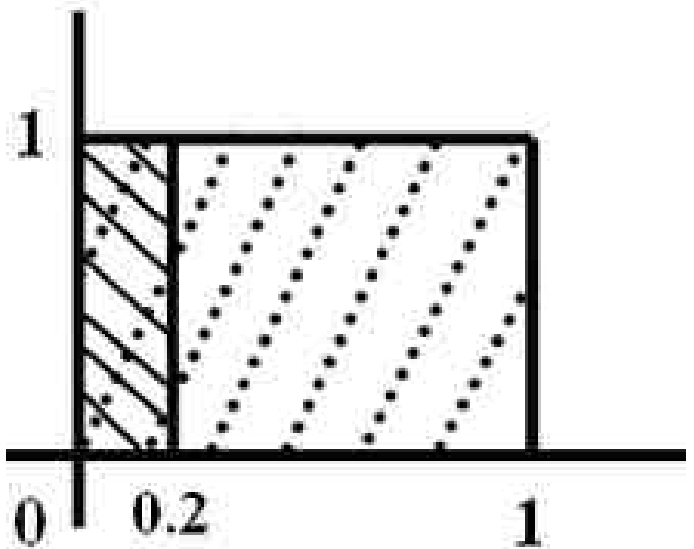
Therefore it is interesting to note that the space of Y is the interval (0,1), the same as X.

X	Y=1-X
0	1
0.2	0.8
0.3	0.7
0.5	0.5
0.6	0.6
0.8	0.2
1	0
Y=X	

$F(0.8)=1-P(X<1-0.8)=1-P(X<1-0.2)=1-0.2=0.8$

$F(0.6)=1-P(X<1-0.6)=1-P(X<1-0.4)=1-0.4=0.6$

But we know that for the Uniform distribution defined on (0,1), P(X<0.2)=0.2



Therefore, P(X<1-y) = 1-y

Further, the cdf of Y is 0 for $y < 0$; 1 for $y \geq 1$; and for $0 \leq y < 1$, it is

$$F_Y(y) = P(Y \leq y) = P(1 - X \leq y)$$
$$= P(X \geq 1 - y) = 1 - P(X < 1 - y)$$
$$= 1 - (1 - y) = y.$$

Hence, $F_Y(y)=y$, $0 \leq y \leq 1$ which is of exactly the same form as the CDF of X. Therefore, it is clarified that two random variables may not be the same yet, sometimes they maybe equal in distribution.

$$i.e., Y \overset{D}{=} X, \quad but\, Y \neq X.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 18

First property

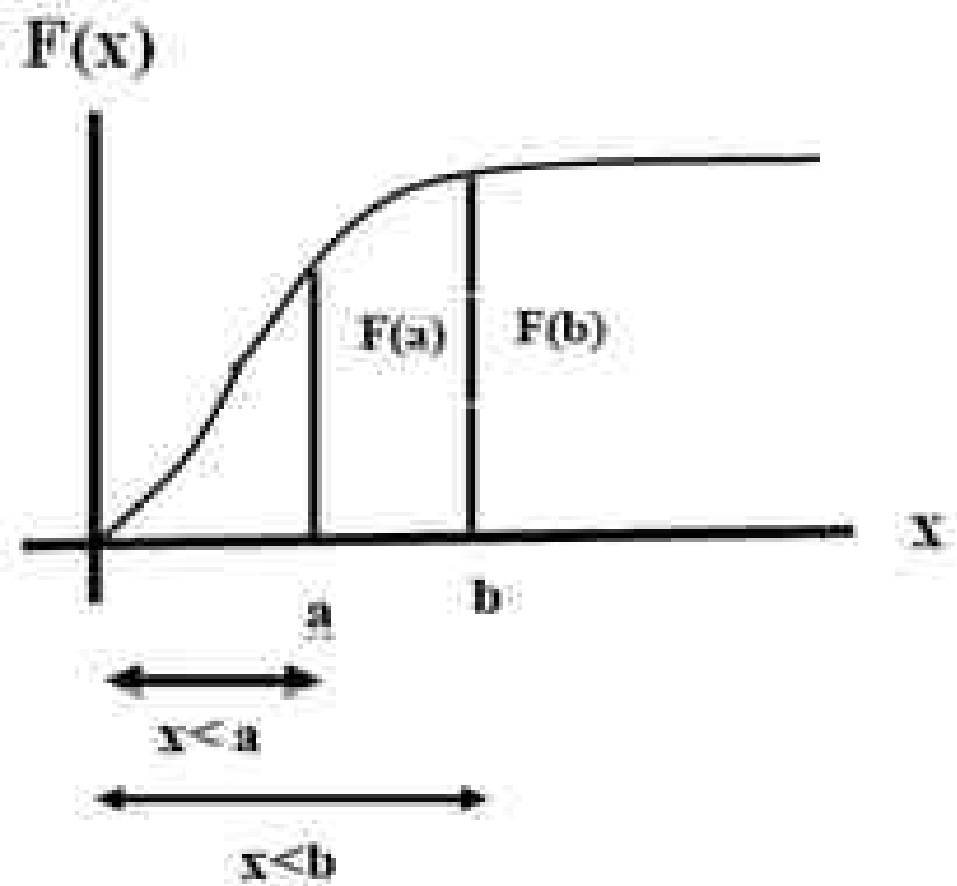
of Cumulative Distribution Function and its Proof

Theorem 1:

Let X be a random variable with cumulative distribution function $F(x)$.

Then

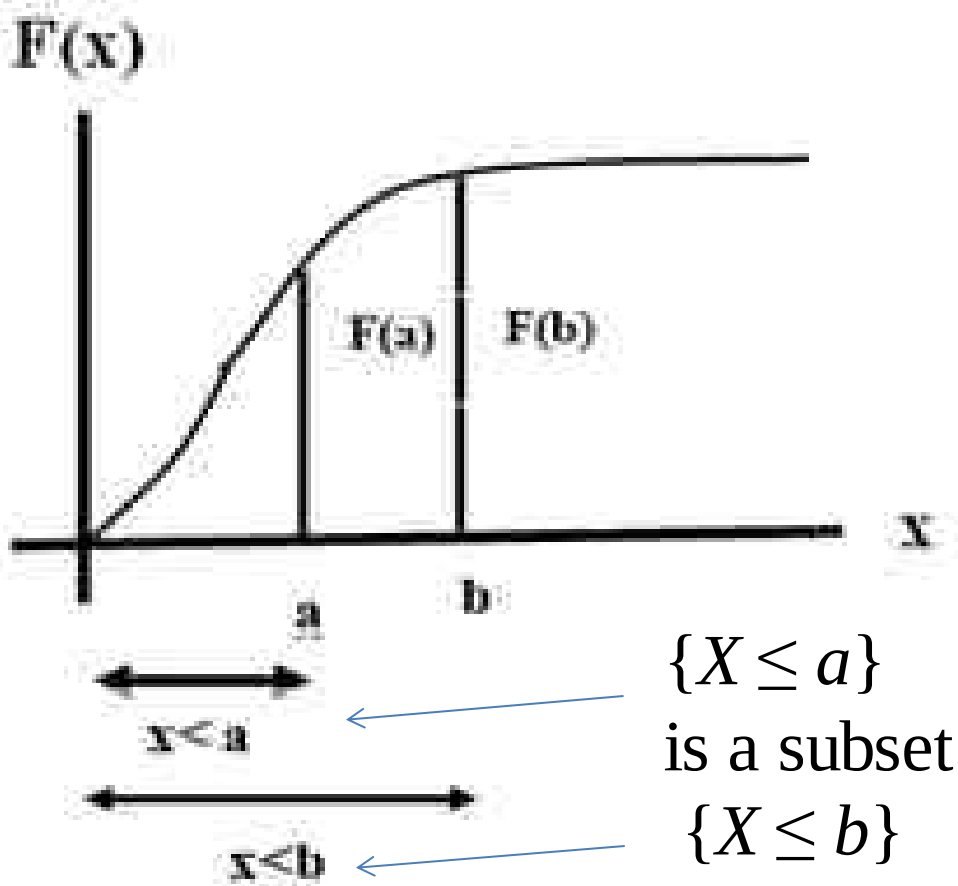
(a) For all a and b, if $a < b$, the $F(a) \leq F(b)$ (F is nondecreasing).


F(x)
F(a)
F(b)
x
a
b
x<a
x<b

Proof:

Part (a):

Because of the fact that $a < b$, we can see that the interval being $\{X \leq a\} \subset \{X \leq b\}$



$\{X \leq a\}$
is a subset of
$\{X \leq b\}$

Theorem:

If C_1 *and* C_2 are events such that $C_1 \subset C_2$, then , $P(C_1) \leq P(C_2)$.

So, applying this theorem we say that since the interval $X \leq a$ is contained in the interval $X \leq b$, therefore the probability of $X \leq a$ is less than or equal to the probability $X \leq b$., i.e. $F(a) \leq P(b)$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 19

Second property

of Cumulative Distribution Function and its Proof

Theorem 1:

Let X be a random variable with cumulative distribution function $F(x)$.

Then,

(b) $\lim_{x \to -\infty} F(x) = 0$ (the lower limit of F is 0).

We know that

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

so obviously

$$F(-\infty) = P(X \leq -\infty)$$

which is an impossible event.

Therefore, $F(-\infty) = 0$

Mathematically,

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

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 20

Third property

of Cumulative Distribution Function and its Proof

Theorem 1:

Let X be a random variable with cumulative distribution function $F(x)$.

Then,

(c) $\lim_{x \to \infty}$, $F(x) = 1$ (the upper limit of F is 1).

Proof:

We know that

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

so obviously

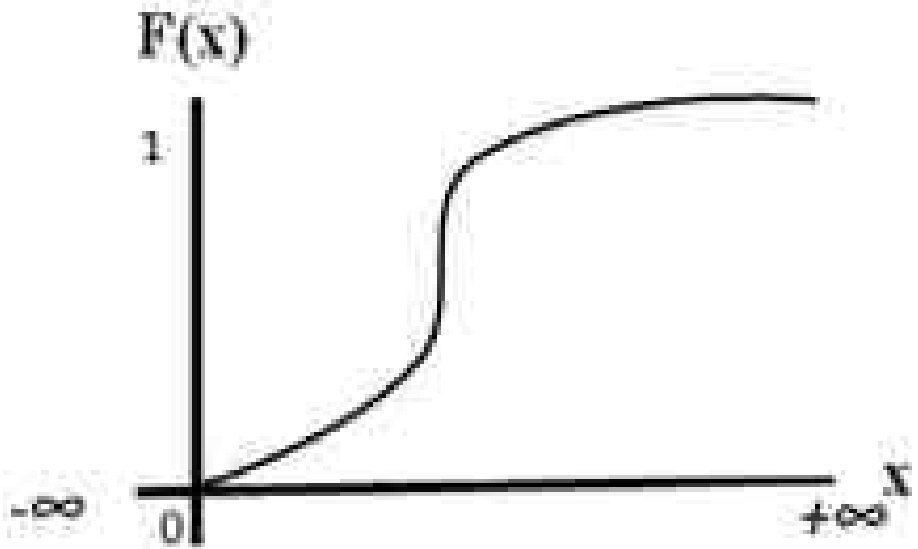
$$F(\infty) = P(X \leq \infty) \quad \text{is a sure event.}$$

Therefore, $F(\infty) = 1$

$$\lim_{x \to \infty} F(x) = 1.$$

Graphically, the $F(x)$ of any random variable will always starts from level zero and rise upto level 1.

Subsequent to that point 1 it goes upto to infinity .



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 21

Fourth property

of Cumulative Distribution Function and its Proof

Theorem 1:

Let X be a random variable with cumulative distribution function $F(x)$.

Then

(d) $\lim_{x \downarrow x0}$, $F(x) = F(x_0)$ (F is right continuous).

Explanation:

In case of a discrete distribution, we take an example of a Discrete Uniform distribution.

- Toss a fair die (1,2,3,4,5,6).

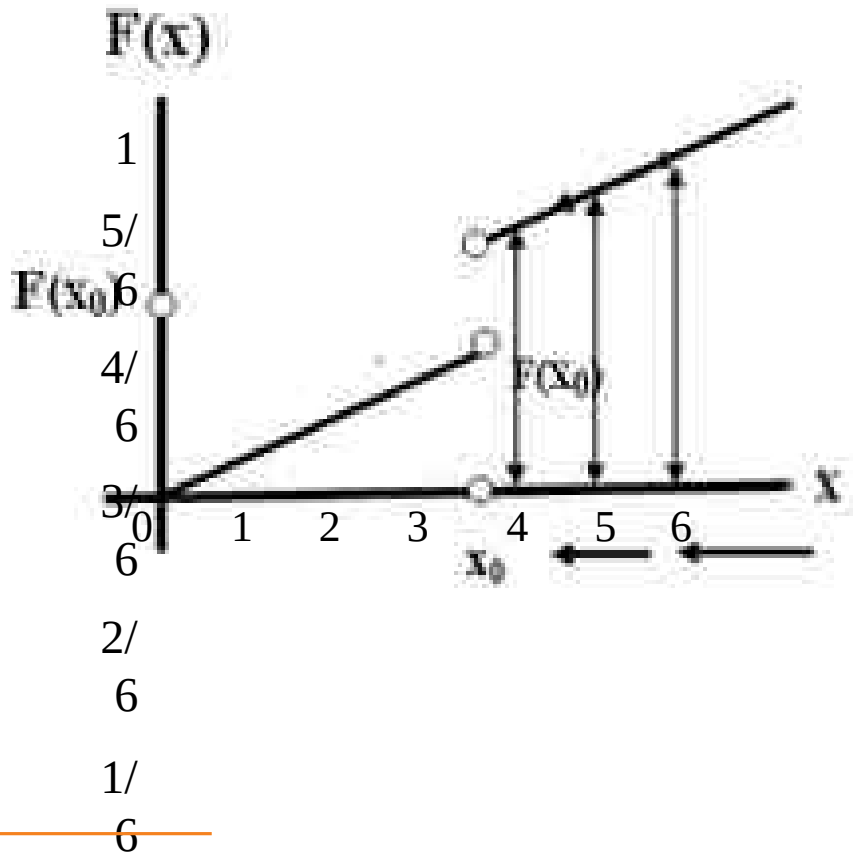
-The probability of getting a 1 or 2 or 3 or 4 or 5 or 6 are equally likely to occur.

x	$P(x)$	$F(x)$
1	1/6	1/6
2	1/6	1/6+1/6=2/6
3	1/6	2/6+1/6=3/6
4	1/6	3/6+1/6=4/6
5	1/6	4/6+1/6=5/6
6	1/6	5/6+1/6=1

Consider a CDF that is continuous everywhere other than at the point x_0

(ref. CDF's that have a few points of discontinuity)

As x tends to x_0 from the R.H.S, the ordinate $F(x)$ tends to the ordinate $F(x_0)$.



Part (d):

Let $\{x_n\}$ be any sequence of real numbers such that $x_n \downarrow x_0$. Let $C_n = \{X \leq x_n\}$. Then the sequence of sets $\{C_n\}$ is decreasing and

$$\bigcap_{n=1}^{\infty} C_n = \{X \leq x_0\}$$

Hence, by Theorem 3.6,

$$\lim_{n\to\infty} F(x_n) = P\left(\bigcap_{n=1}^{\infty} C_n\right) = F(x_0),$$

which is the desired result.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 22

Evaluating

Probabilities using CDF

Theorem 1:

Let X be a random variable with the CDF F_x.

Then,
for $a < b$,

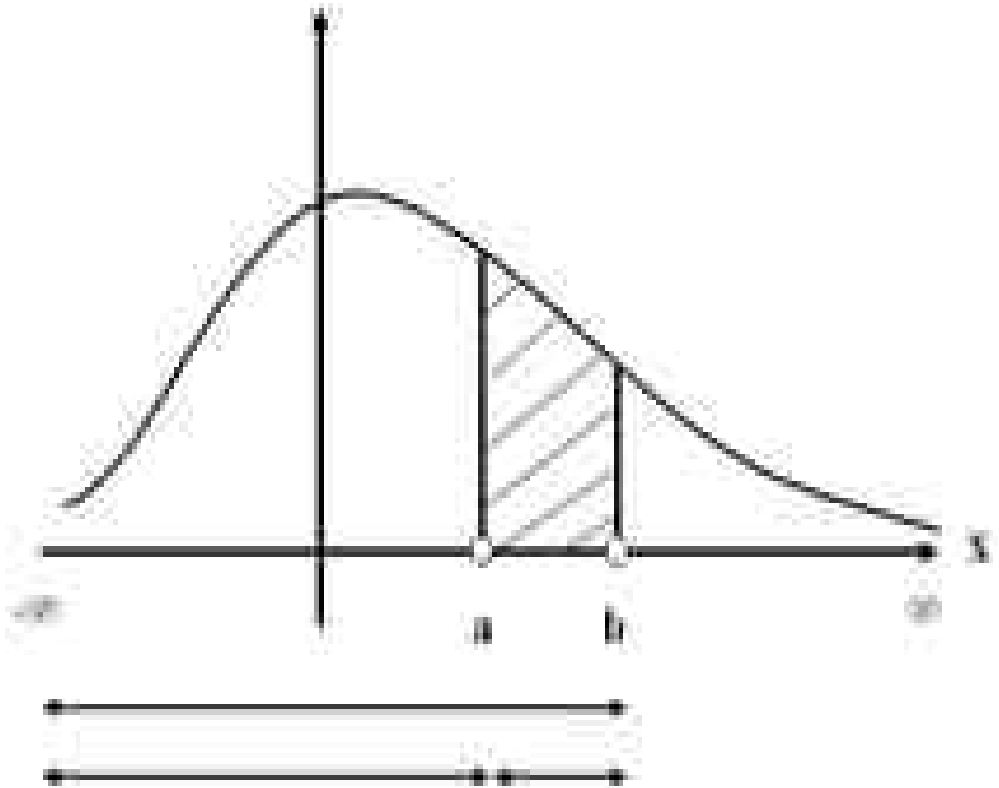
$$P[a < X \leq b] = F_x(b) - F_x(a).$$

Proof:

Since the interval

$$\{-\infty < X \leq b\} = \{-\infty < X \leq a\} \cup \{a < X \leq b\}.$$

Since these two interval are mutually exclusive (not overlapping).



So, therefore,

we can apply the addition theorem of probability and can say that the probability of the union is equal to the sum of the probabilities.

$$P\{-\infty < X \leq b\} = P\{-\infty < X \leq a\} + P\{a < X \leq b\}.$$

or $$P\{a < X \leq b\} = P\{-\infty < X \leq b\} - P\{-\infty < X \leq a\}$$

Example:

Now, find $P(1< X < 2), \theta=2$

Then required probability

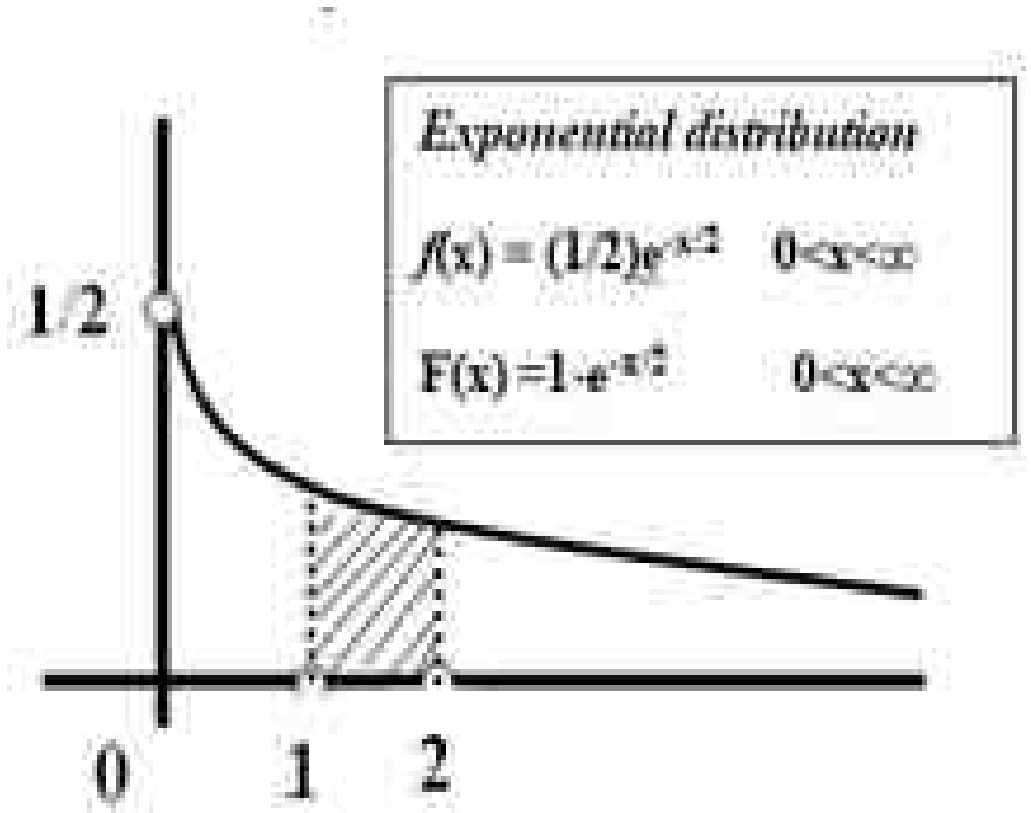
$=F(2)-F(1)=1-1-e^{-1}+e^{-1/2}$

$= e^{-1/2}-e^{-1}=(1/\sqrt{e})-(1/e)$

$=(1/1.64872)-(1/2.71828)$

$=0.6065-0.3679 =0.2386$

$=23.86\%$



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 23

Derivative

Of the CDF is the PDF

The function $f_X(x)$ is called a probability density function (pdf) of X. If $f_X(x)$ is also continuous, the Fundamental Theorem of Calculus implies that

$$\frac{d}{dx} F_X(x) = f_X(x) \qquad (1)$$

Example:

Consider the CDF of the exponential distribution with mean = 1

i.e. $F(x)=1 - e^{-x}$

Then, the PDF is given by

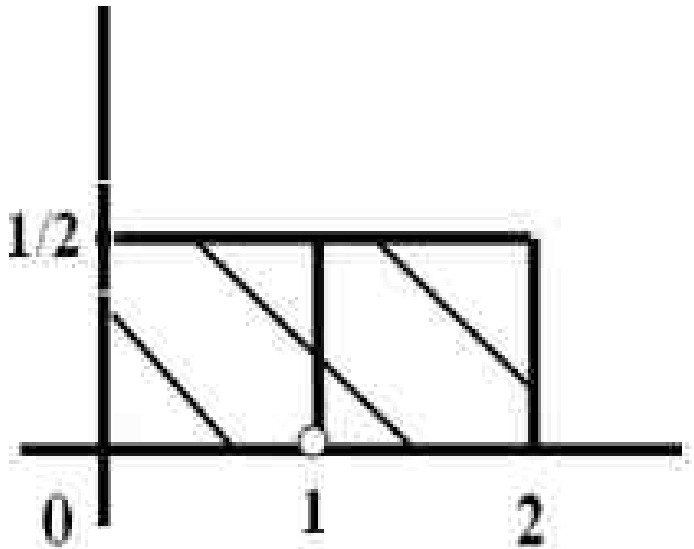
$$\frac{d}{dx}F_X(x) = \frac{d}{dx}\left(1-e^{-x}\right) = 0-(-1)e^{-x} = e^{-x}, \quad x \geq 0$$

2: Another Example

If $F_X(x) = x/2, \quad 0 < x < 2$

Then pdf is

$$\frac{d}{dx} F_X(x) = \frac{d}{dx}\left(\frac{x}{2}\right) = \frac{1}{2}\frac{d}{dx}(x) = \frac{1}{2}, \quad 0 < x < 2$$



i.*e*. Uniform distribution.

Virtual University of Pakistan
Probability Distributions

by

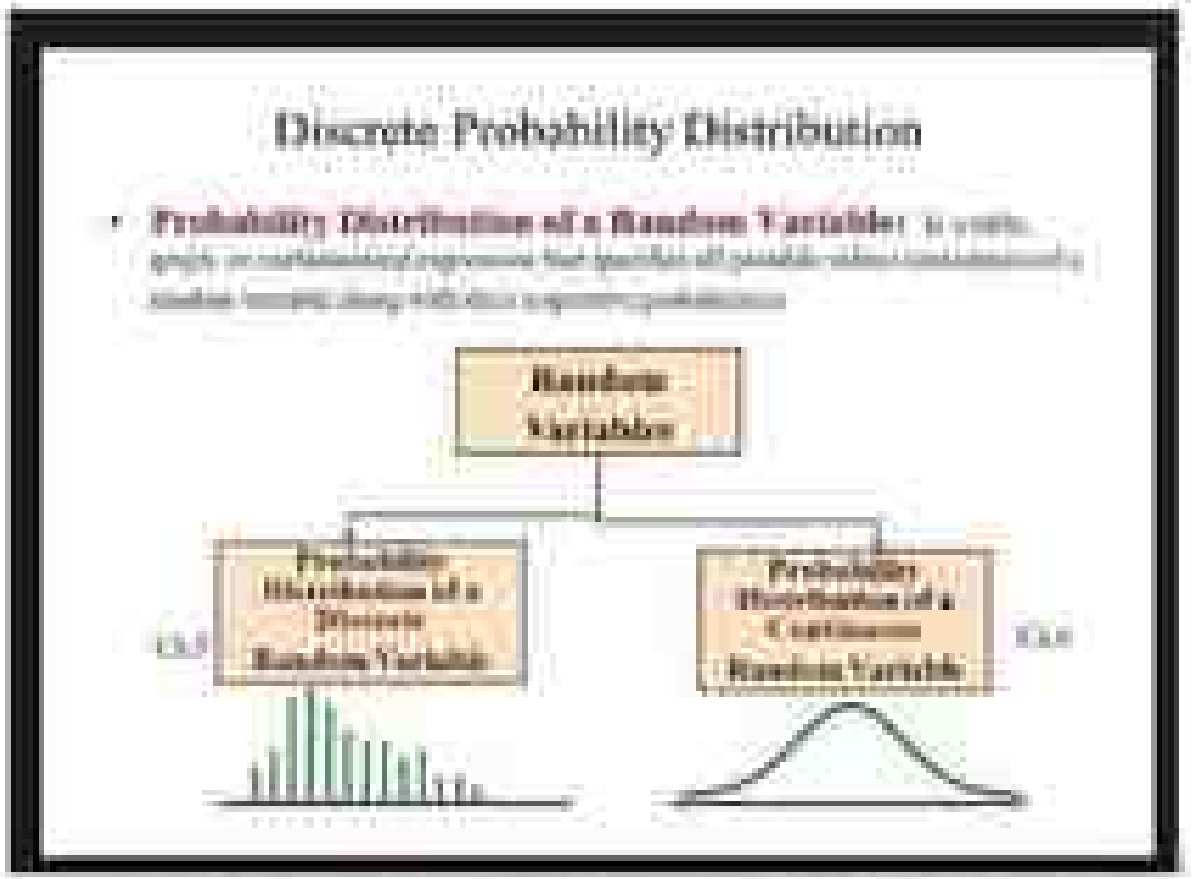
Dr. Saleha Naghmi Habibullah

Topic No. 24

In the continuous case,

probability does not Exist at any particular point

In the discrete case probability is given by the length of the ordinate erected on top of any particular point.


Discrete Probability Distribution
Probability Distribution of a Random Variable:
Random Variable
Probability Distribution of a Discrete Random Variable
Probability Distribution of a Continuous Random Variable

- On the other hand, in the continuous case, probability is <u>not</u> given by the ordinate erected on top of any particular point (on the contrary, it is given by the <u>area</u> <u>between</u> any two ordinates).

- If $f(x)$ is the pdf of x, then the probability x belongs to A, where A is some interval, is given by the integral of $f(x)$ over that interval, that is:

$$P(X \epsilon A) = \int_A f(x)\, dx$$

- If $f(x)$ is the pdf of x, then the probability x belongs to A, where A is some interval, is given by the integral of $f(x)$ over that interval, that is:

$$P(X \epsilon \mathrm{A}) = \int_{\mathrm{A}} f(x)\, dx$$

WHY do we say that probability does not exist at any particular point ?

i.e.

$$P(X = x_0) = 0$$

We say this because it is not possible to have a measurement exactly equal to X_0 whatever we are saying is an approximation to the truth.

What is P(*X*=1/2)?

Solution:

It is a straightforward integration to see that the probability is 0:

In fact, in general, if X is continuous, the probability that X takes on any specific value x is 0.
That is, when X is continuous , $P(X = x) = 0$ for all x in the support.

$$\int_{1/2}^{1/2} 3x^2 dx = \left[x^3 \right]_{x=1/2}^{x=1/2} = \frac{1}{8} - \frac{1}{8} = 0$$

An implication of the fact that $P(X=x) = 0$ for all x when X is continuous is that you can be careless about the endpoints of intervals when finding probabilities of continuous random variables.

That is:

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b)$$

for any constant a and b.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 25

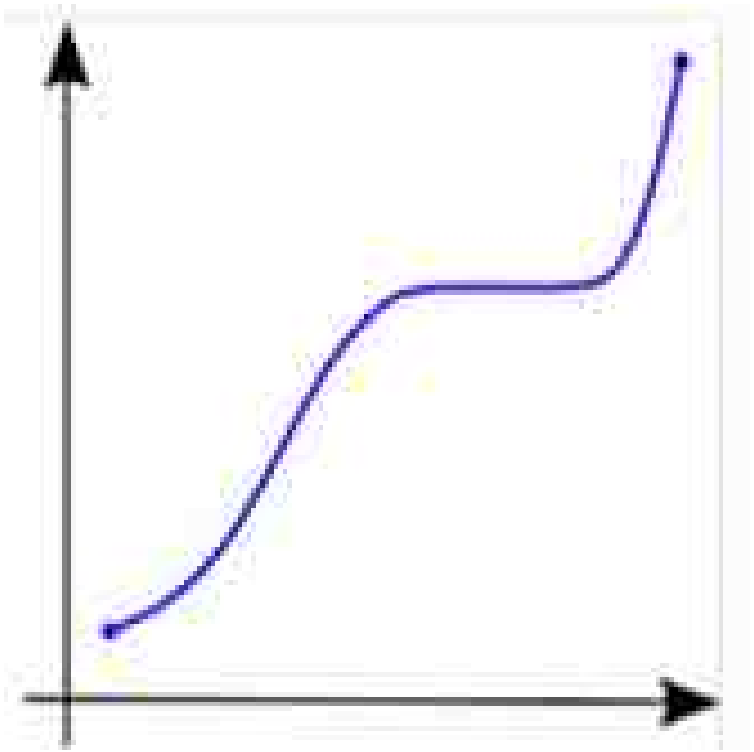
The concept of

Monotonicity

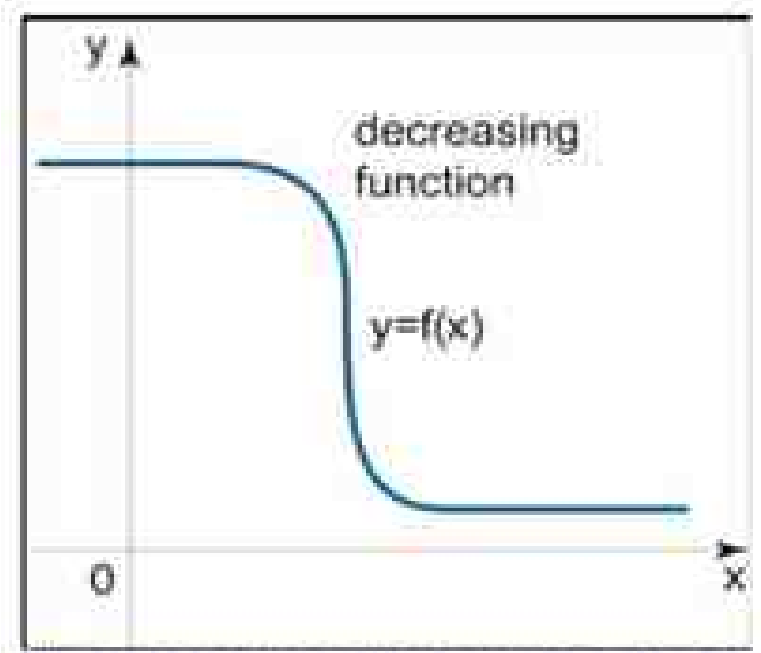
Monotonically Increasing Function

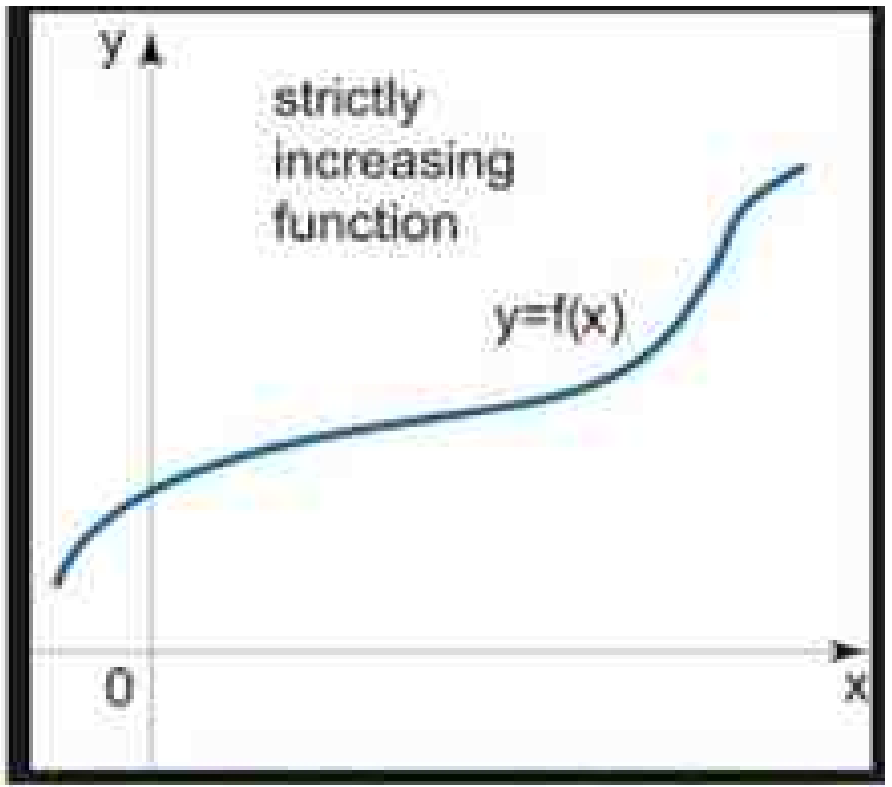
In calculus, a function defined on a subset of the real numbers with real values is called **monotonic** if and only if it is either entirely non-**increasing**, or entirely non-decreasing.

That is, as per Fig., a **function** that increases **monotonically** does not exclusively have to **increase** , it simply must not decrease.



Monotonically Decreasing Function



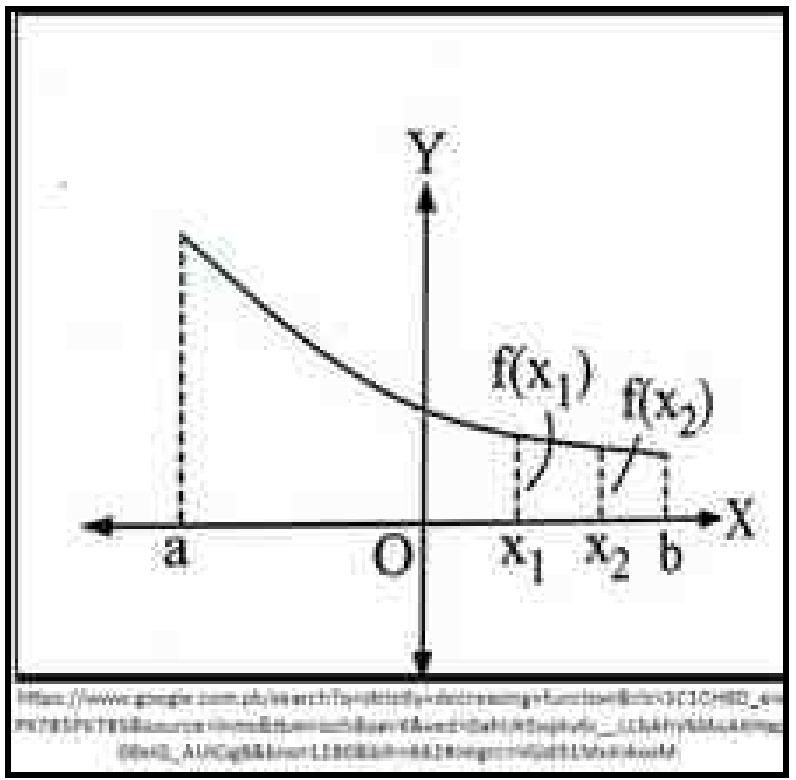


Strictly Increasing Function

A function $f(x)$ is said to be strictly increasing on an interval I if $f(b) > f(a)$ for all $b>a$, where $a,b \in I$.

On the other hand, if $f(b) \geq f(a)$ for all $b > a$, the function is said to be (nonstrictly) increasing.

Strictly decreasing Function



Focus on the word

“Mono-tone”

The concept of the CDF, will always be monotonically increasing function.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 26

Total probability is 1

Total probability is 1

The total probability associated with a random variable X of the discrete type with pmf $p_X(x)$

$$\Sigma_{x\epsilon D}\, p_X(x)=1$$

or

the continuous type with pdf $f_X(x)$ is 1, i.e.

$$\int_D f_X(x)dx=1,$$

where D is the space of X.

The definition for the pdf of a continuous random variable differs from the definition for the pmf of a discrete random variable by simply changing the summations that appeared in the discrete case to integrals in the continuous case.

It is very well-known that pmfs satisfy the two properties

(i) $p_X(x) \geq 0$ and (ii) $\Sigma_{x \epsilon D} p_X(x)=1$

Therefore, the minimum possible value of any probability is 0 and the maximum possible value of any probability is equal to 1.

It is very well-known that pdfs satisfy the two properties

(i) $f_X(x) \geq 0$ and (ii) $\int_{-\infty}^{\infty} f_X(t)\, dt=1$

(ii) The second property, of course, follows from $F_X(\infty)=1$

- In the continuous case the *f*(x) represents the ordinate against the value *x*, and the ordinate can be greater than 1 (although area under the curve can never be greater than 1).

Example:

Let X be a continuous random variable whose probability density function is:

$$f(x) = 3x^2$$

for $0<x<1$. First, note that $f(x) \neq P(X=x)$.

For example, $f(0.9)=3(0.9)^2=2.43$, which is clearly not a probability!

In the continuous case, $f(x)$ is instead the height of the curve at $X=x$, so that the total area under the curve is 1.

In the continuous case, it is areas under the curve that define the probabilities.

Now, let's first start by verifying that $f(x)$ is a valid probability density function.

- The function $f(x) = 3x^2$ is not a Probability density function that is why we are getting $f(x)$ greater than 1. So, let me show you that it is a PDF.

$$\int_0^1 3x^2 dx = 3\frac{x^3}{3}\Big|_0^1 = x^3\Big|_0^1 = 1^3 - 0^3 = 1$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 27

Example showing the determination of an unknown constant c Such that $f(x)$ or p(x) is a the PMF

Example:

Find the constant c do that p(x) satisfies the condition of being a pmf of one random variable X.

$$p(x)=\begin{cases} c\left(\frac{2}{3}\right)^{x}, & x=1,2,3,\ldots \\ 0 & \textit{elsewhere} \end{cases}$$

- If $p(x)$ to be a probability mass function, then:

$$1=\sum_{x=1}^{\infty} p(x)=\sum_{x=1}^{\infty} c\left(\frac{2}{3}\right)^{x}=c\left(\frac{2}{3}\right)^{1}+c\left(\frac{2}{3}\right)^{2}+c\left(\frac{2}{3}\right)^{3}+\ldots$$

$$=c\left(\left(\frac{2}{3}\right)^{1}+\left(\frac{2}{3}\right)^{2}+\left(\frac{2}{3}\right)^{3}+\ldots\right)$$

What we have inside the bracket is an Infinite Geometric series,

Therefore, applying the formula $S = \frac{a}{1-r}$,we obtain

$$1 = c\left(\frac{\left(\frac{2}{3}\right)}{1-\left(\frac{2}{3}\right)}\right) = c\left(\frac{\left(\frac{2}{3}\right)}{\frac{1}{3}}\right) = c(2) \qquad \textit{and hence} \qquad c = \frac{1}{2}.$$

As an illustration, suppose we want to find a probability of x=4.

Then putting $c = \frac{1}{2}$, we have $p(x) = \frac{1}{2}\left(\frac{2}{3}\right)^x$.

now putting x=4,

we obtain,

$$p(4) = \frac{1}{2}\left(\frac{2}{3}\right)^4 = \frac{1}{2}\left(\frac{16}{81}\right) = \frac{8}{81} = 0.099$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 32

Example

showing the determination of an unknown constant c Such that f(x) is a the PDF

Example

Suppose X has the pdf

$$f_X(x) = \begin{cases} cx^3 & 0 < x < 2 \\ 0 & elsewhere \end{cases}$$

for a constant c

Solution:

If $f(x)$ is to be a pdf, then,

$$1 = \int_0^2 cx^3 dx = c\left[\frac{x^4}{4}\right]_0^2 = c\left[\frac{2^4}{4} - \frac{0}{4}\right] = c\left[\frac{16}{4}\right] = c[4] = 4c,$$

and hence, c=1/4.

So, now the pdf will be as follows

$$f_X(x) = \begin{cases} \dfrac{x^3}{4} & 0 < x < 2 \\ 0 & \textit{elsewhere} \end{cases}$$

For illustration of the computation of a probability involving X, we have

$$P\left(\frac{1}{4} < X < 1\right) = \int_{1/4}^{1} \frac{x^3}{4} dx = \frac{255}{4096} = 0.0623.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 29

Concept of Transformation

of a Discrete variable

(when the transformation is one-to-one)

We have a discrete random variable X and we know its distribution.

We are interested, though, in a random variable Y which is some transformation of X, say, $Y = g(X)$.

In particular, we want to determine the distribution of Y.

Assume X is discrete with space D_x. We consider two cases.

Then the space of Y is

$$D_y=\{g(x):x \epsilon D_x\}.$$

Let us assume that g is one-to-one.

Then, clearly, the pmf of Y is obtained as

$$P[Y = y] = P[g(X) = y] = P[X = i^{-1}(y)].$$

Example

Let X have the pmf

$$P(x) - (1/2)^{|x|}, \qquad x = -1,-2,-3,\ldots$$

Find the pmf of $Y = X^4$

Solution:

The transformation,

$$y = g(x) = x^4$$

$$maps\ D_x = \{x : x = -1, -2, -3, \ldots\}\ \ onto$$

$$D_y = \{y : y = 1, 16,\ 81, \ldots\}$$

We have

$$y = g(x) = x^4 \Rightarrow x = g^{-1}(y) = \sqrt[4]{y}$$

So, utilizing the equation

$$P[Y = y] = P[g(X) = y] = P[X = g^{-1}(y)].$$

we get the single-valued inverse function as follows:

$$\therefore \; P[Y = y] = P\left(X = \sqrt[4]{y}\right) = \left(\frac{1}{2}\right)^{\left|\sqrt[4]{y}\right|}, \quad y = 1,\ 16,\ 81,\ \ldots$$

So, this is the PMF of $Y=X^4$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 30

Concept of Transformation of a Discrete Variable (when the transformation is not one-to-one)

Consider a sequence of independent flips of a **fair** coin, each resulting in a head H or a tail T.

Let the random variable X denote the no. of flips needed to obtain the first Head.

Then, X is a **geometric** random variable.

If $X=x$, where x=1,2,3,4,…, there must be a string of x-1 tails followed by a head:

that is, TT…TH, where there are x-1 tails in TT…T.

Thus, from independence, we have a geometric sequence of probabilities, namely,

$$P(X=x)=(1/2)^{x-1}(1/2) = (1/2)^{x} \quad , x = 1,2,3,\ldots, \qquad (1)$$

Suppose we are playing a game in which if the first head appears on an odd number of flips, we lose one dollar whereas, if the first head appears on an even number of flips, we win one dollar.

Let Y denote our net gain.

Then the space of Y is **{-1,1}.**

An interesting event is that the first head appears on an odd number of flips;

i.e., $X \epsilon \{1,3,5,\ldots\}$.

The probability of this event is

$$P[X \text{ ò} \{1 \text{ or } 3 \text{ or } 5 \text{ or} \ldots\}] = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^{2x-1}$$

Next,

$$= (1/2)^1 + (1/2)^3 + (1/2)^5 + \ldots$$

And infinite geometric series, $\frac{a}{1-r} = a = (1/2), r = (1/2)^2 = 1/4$

$$\text{Required probability } \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^{2x-1} = \frac{\frac{1}{2}}{1-\left(\frac{1}{4}\right)} = \frac{\frac{1}{2}}{\left(\frac{3}{4}\right)} = \frac{2}{3}.$$

We have already determined that the probability that X is odd is 2/3.

Hence, the distribution of Y is given by

$$p_Y(-1)=2/3 \text{ and } p_Y(1)=1/3.$$

So, the probability that player A will win 1$ is 1/3 and the probability that he will lose 1 $ is 2/3.

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

Topic No. 31

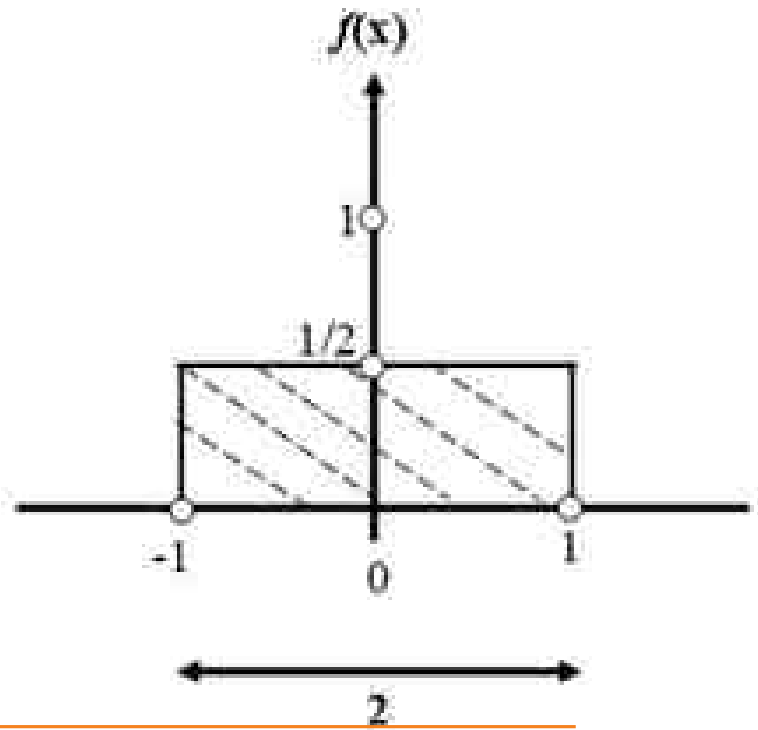
Concept of Transformation of a Continuous Variable using the CDF technique

Example:

Let $f_X(x)=1/2,\ -1 < x < 1,$

zero elsewhere,

be the pdf of a random variable X.



From the graph it is obvious that we are dealing with continuous uniform distribution defined on the interval (-1,1).

Now, let us suppose that we are interested in determining the PDF of X^2.

Let the random variable Y be $Y=X^2$.Now to find the pdf of Y.

The point to be understood is that, if $y \geq 0$, then
$X^2 \leq y \Rightarrow -\sqrt{y} \leq X \leq \sqrt{y}$
e.g. $0 \leq X^2 \leq 0.5 \Rightarrow -0.25 \leq X \leq 0.25$
(because square can never be negative)

From the table it is obvious that the range of the variable Y is from 0 to 1.

$X^2=Y$	X
0	0
0.04	± 0.2
0.25	± 0.5
0.49	± 0.7
1.00	± 1.0

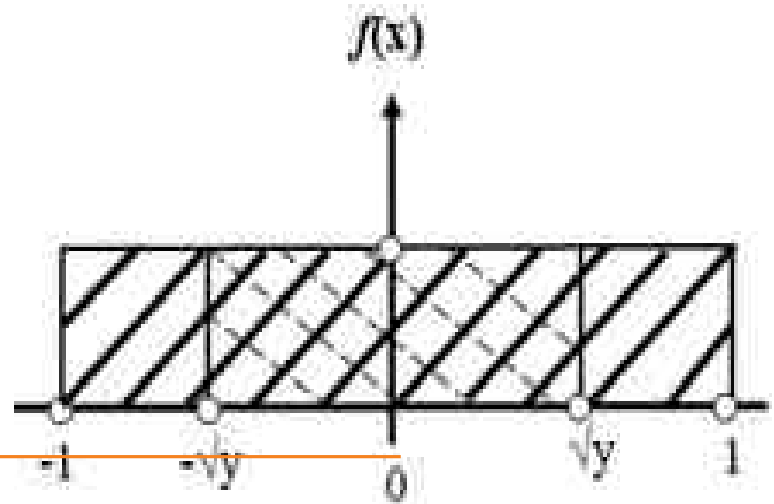
Now finding the pdf of *Y*.

The cdf of Y, $F_Y(y) = P(Y \leq y)$, is given by the cdf of Y,

$F_Y(y) = P(Y \leq y)$, is given by

$$P(X^2 < y) = P(-\sqrt{y} < X < \sqrt{y})$$

$$= \int_{-\sqrt{y}}^{\sqrt{y}} f(x)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\,dx$$



For the range $0 \le y \le 1$, we have

$$F(y) = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\,dx = \frac{1}{2}x\Big|_{-\sqrt{y}}^{\sqrt{y}} = \frac{1}{2}\left[\sqrt{y} - \left(-\sqrt{y}\right)\right] = \frac{2\sqrt{y}}{2} = \sqrt{y}$$

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \displaystyle\int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\,dx = \sqrt{y} & 0 \le y < 1 \\ 1 & 1 \le y \end{cases}$$

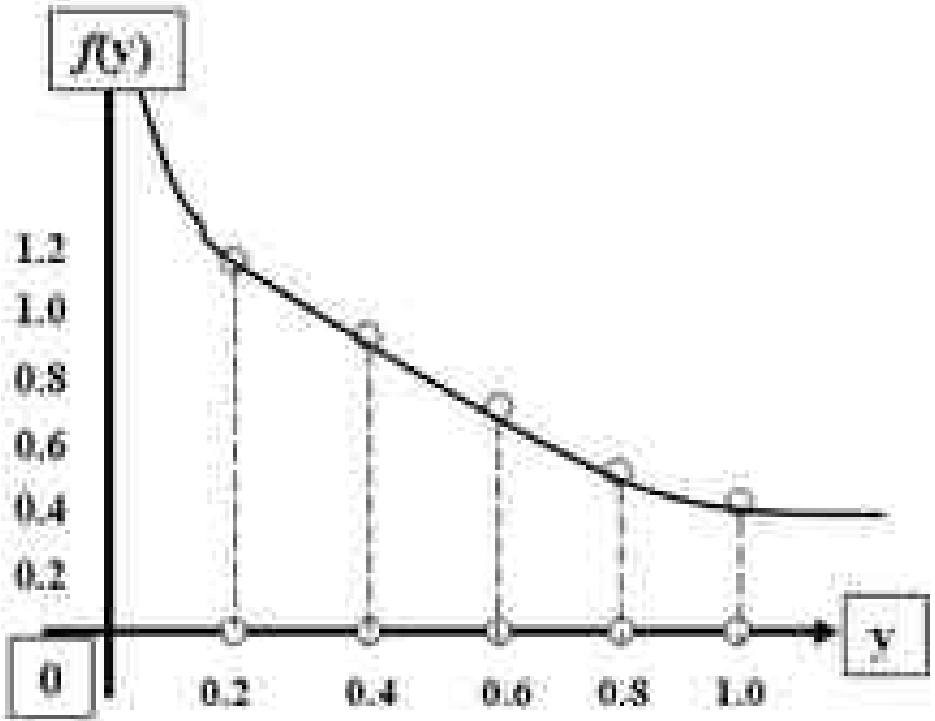
Hence, the pdf of Y is given by

$$f_Y(y) = \begin{cases} \dfrac{1}{2\sqrt{y}} & 0 < y < 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

$$f(y) = \frac{d}{dy} F(y) = \begin{cases} \dfrac{d}{dy}(0) = 0, & y < 0 \\ \dfrac{d}{dy}\sqrt{y} = \dfrac{d}{dy}\sqrt{y} = \dfrac{1}{2} y^{\frac{1}{2}-1} = \dfrac{1}{2} y^{-\frac{1}{2}} = \dfrac{1}{2\sqrt{y}}, & 0 \le y \le 1 \\ \dfrac{d}{dy}(1) = 0, & y \ge 1 \end{cases}$$

$$\textit{or} \quad f(y) = \begin{cases} \dfrac{1}{2\sqrt{y}}, & 0 \le y \le 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

y	$1/2\sqrt{y}$
0	∞
0.2	1.12
0.4	0.79
0.6	0.65
0.8	0.56
1	0.50



Hence we can see that the shape of the PDF of the transformed variable is very different from the shape of the PDF of the original variable.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 32

Transformation of a Continuous Variable using the Jacobian of transformation

Theorem 1:

Let X be a continuous random variable with pdf $f_X(x)$ and **support** S_X.

Let $Y=g(X)$, where g(x) is **one-to-one** **differentiable function**, on the support of X, S_X.

Denote the inverse of g by $x=g^{-1}(y)$ and
let $dx/dy=d[g^{-1}(y)]/dy$.

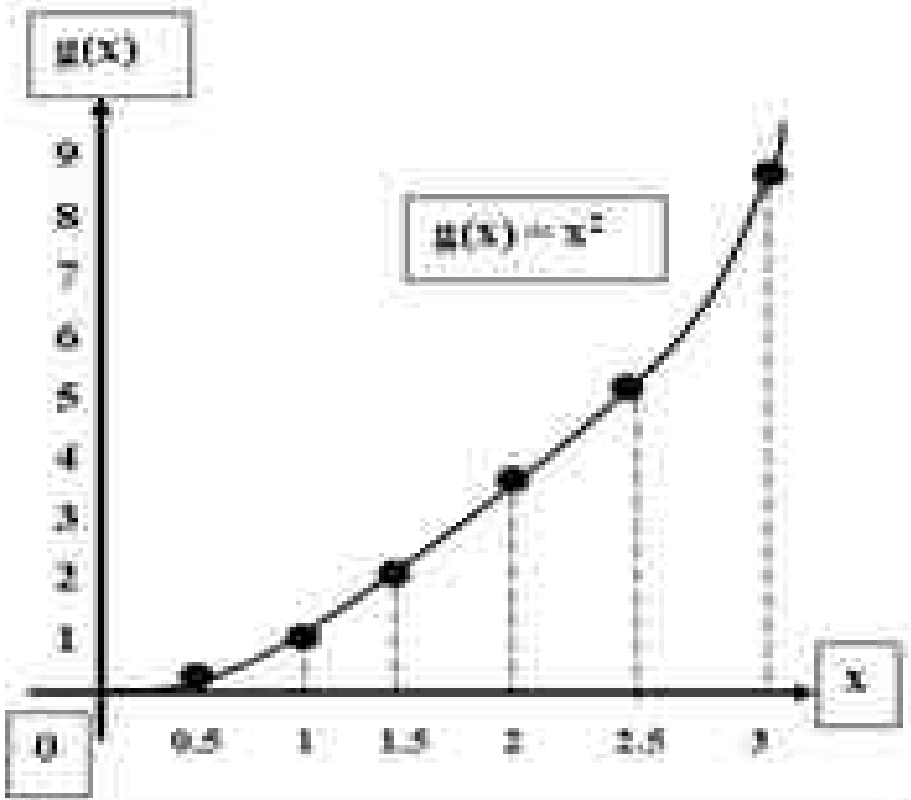
Then the PDF of Y is given by

$$f_Y(y) = f_X(g^{-1}(y)) \, |dx / dy|, \quad \text{for } y \in S_Y, \qquad (1)$$

where the support of Y is the set $S_Y = \{y=g(x):x \in S_X\}$.

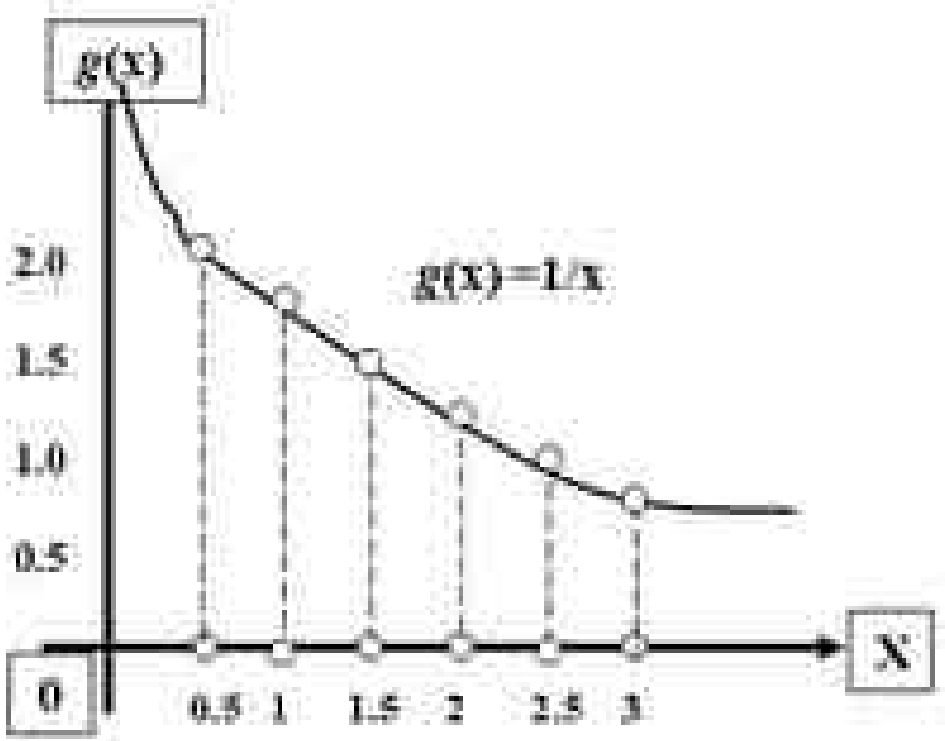
Proof:

If $g(x)$ is one-to-one and continuous, it is either **strictly** increasing or **strictly** decreasing.



Here g(x) is one-to-one function and we can see that it is strictly increasing.

This can be regarded as a case of Direct relationship between *X* and g(*x*).



Here again g(x) is one-to-one and we can see that it is strictly decreasing.

This can be regarded as a case of Inverse relationship between *X* and g(*x*).

Case 1:

Let us assume that it is strictly increasing,.

Then the CDF of Y is given by

$$F_Y(y) = P[Y < y] = P[g(x) < y]$$
$$= P[X < g^{-1}(y)] = F_X(g^{-1}(y)). \qquad (2)$$

Hence, the pdf of Y is

$$f_Y(y) = \frac{d}{dy}(F_Y(y)) = \frac{d}{dy}(F_X(g^{-1}(y))) = f_X(g^{-1}(y))\left(\frac{dx}{dy}\right). \qquad (3)$$

where dx/dy is the derivative of the function $x = g^{-1}(y)$.

In this case, because g is increasing, dx/dy > 0.

Hence, we can write dx/dy =|dx/dy|.

Case 2:

Suppose $g(x)$ is strictly decreasing.

Then

$$F_Y(y) = P[Y < y] = P[g(x) < y] = P[X < g^{-1}(y)] = F_X(g^{-1}(y)). \qquad (2)$$

becomes $F_Y(y) = 1 - F_X(g^{-1}(y))$.

Hence, the pdf of Y is

$$f_Y(y) = 0 - f_X(g^{-1}(y))\left(\frac{dx}{dy}\right) = f_X(g^{-1}(y))\left(-\frac{dx}{dy}\right).$$

But since g is decreasing, dx/dy < 0 and,

hence, $-dx/dy = |dx/dy|$.

Thus eq. 1 is true in both cases.

- Henceforth, we refer to $dx/dy=(d/dy)g^{-1}(y)$ as Jacobian (denoted by J) of the transformation.

- In most mathematical areas, J = dx/dy is referred to as the Jacobian of the inverse transformation $x = g^{-1}(y)$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

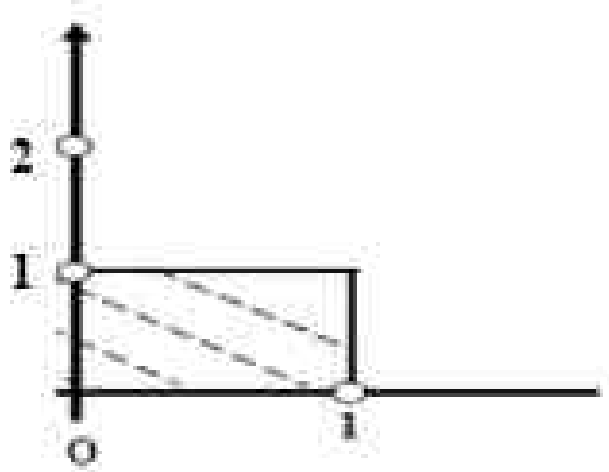
Topic No. 33

Example

of Transformation of a Continuous Variable (using the Jacobian of transformation)

Example : Let X have the pdf

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



Consider the random variable $Y = -2logX$.

The support sets of X and Y are given by $(0,1)$ and $(0,\infty)$, respectively.

The transformation $y = g(x) = -2\log x$ is one-to-one between these sets.

The inverse of the transformation is $x = g^{-1}(y) = e^{-y/2}$

The Jacobian of the Transformation is

$$J = \frac{de^{-y/2}}{dy} = -\frac{1}{2}e^{-y/2}$$

Accordingly, the pdf of $Y = -2 \log X$ is

$$f_Y(y) = \begin{cases} f_Y\left(e^{-y/2}\right)|J| = 1.\frac{1}{2}e^{-y/2} = \frac{1}{2}e^{-y/2} & 0 < y < \infty \\ 0 & \textit{elsewhere} \end{cases}$$

This is the Exponential Distribution with Mean = 2

The above is an example of

- Transformation of Continuous Uniform Random Variable to an Exponentially distributed random variable

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 34

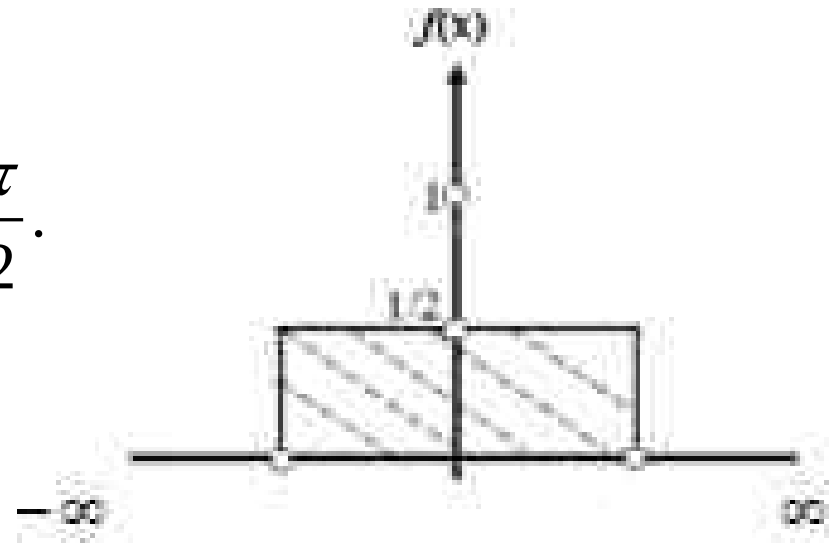
Another example of Transformation of a Continuous variable (using the Jacobian of transformation)

Example:

Let X have the uniform pdf

$$f_X(x) = \frac{1}{\pi}, \qquad for \quad -\frac{\pi}{2} < x < \frac{\pi}{2}.$$

Find the pdf of $Y = \tan X$.



Let y=tanx

When, $x \rightarrow -\pi/2, y = -\infty$, $x \rightarrow \pi/2, y = \infty$

$y = tanx \Rightarrow x = tan^{-1}y \Rightarrow dx/dy = 1/(1+y^2)$

$|J| = |dx/dy| = |1/(1+y^2)| = 1/(1+y^2)$

Accordingly, the pdf of Y=tanX is

$$f_Y(y) = \begin{cases} f_X\left(\tan^{-1}(y)\right)|J| = \dfrac{1}{\pi\left(1+y^2\right)} & -\infty < y < \infty \\ 0 & \textit{elsewhere} \end{cases}$$

This is the pdf of a Cauchy distribution which is one of the well-known distributions.

In other words, the above is an example of

Transformation of Continuous Uniform Random variable to a Cauchy random variable

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 35

Mode

of Discrete Random Variable

Definition

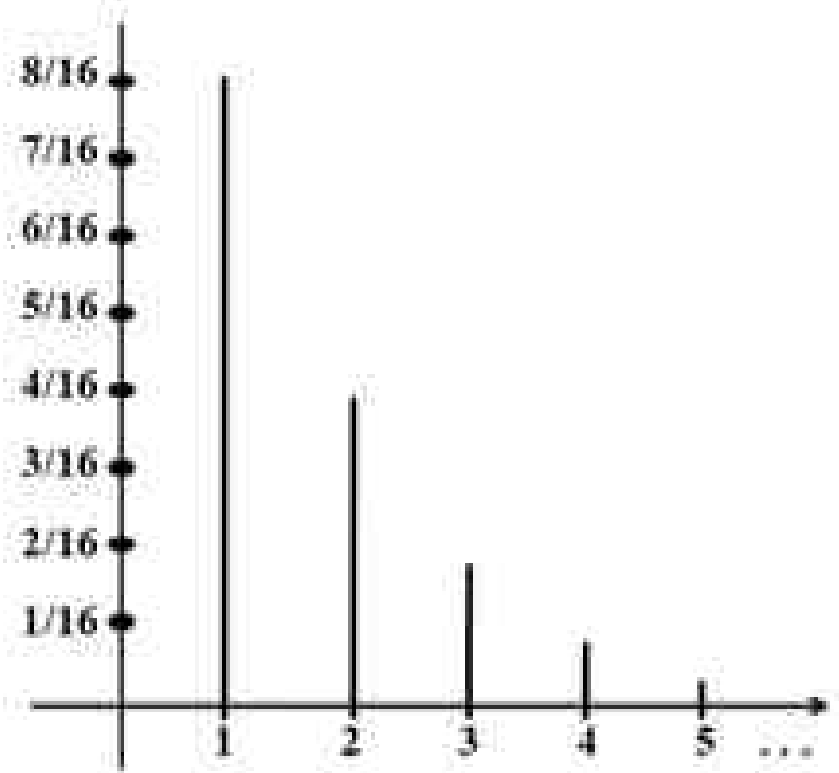
A mode of the distribution of a random variable X is a value of x that **maximizes** the pdf or pmf.

If there is only **one** such x, it is called the **mode** of the distribution.

Find the mode of the following distribution.

a) $p(x) = (1/2)^x$, $x = 1,2,3,\ldots$, zero elsewhere.

X	P(x)
1	$(1/2)^1$
2	$(1/2)^2$
3	$(1/2)^3$
4	$(1/2)^4$
-	-
-	-



Therefore, simply by inspection we say that the mode = 1.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 36

Mode of a Continuous Random Variable

Definition

A mode of the distribution of a random variable X is a value of x that **maximizes** the pdf.

If there is only one such x, it is called the mode of the distribution.

Procedure:

We will be following the procedure that is adopted when determining the maxima and minima of a function.

1. First derivative to be equated to zero and the equation thus obtained to be solved for *x*.
2. The value of the second derivative to be evaluated at that value of *x* which was obtained in step 1.
3. In case the second derivative turns out to be less than zero, that particular value of 'x' will be regarded as the mode of the distribution.

Find the mode of the following distribution.

$$c)\quad f(x)=\begin{cases}\frac{1}{2}x^2e^{-x}, & 0<x<\infty,\\ \textit{zero elsewhere.}\end{cases}$$

Solution:

Taking the first derivative w.r.t. 'x'

$$f'(x)=\frac{d}{dx}\left[\frac{1}{2}x^2e^{-x}\right]$$

$$=\frac{1}{2}\frac{d}{dx}\left[x^2e^{-x}\right]$$

$$f'(x) = \frac{1}{2}\left[x^2 e^{-x}(-1) + e^{-x} 2x\right]$$

$$= -\frac{1}{2} x^2 e^{-x} + \frac{1}{2} 2xe^{-x}$$

$$f'(x) = xe^{-x} - \frac{1}{2} x^2 e^{-x}$$

Now, equating it to zero, we get

$$\Rightarrow xe^{-x} - \frac{1}{2} x^2 e^{-x} = 0$$

$$\Rightarrow xe^{-x} = \frac{1}{2} x^2 e^{-x} \Rightarrow 2 = \frac{1}{xe^{-x}} x^2 e^{-x}$$

$$\therefore x = 2$$

$$f'(x) = xe^{-x} - \frac{1}{2}x^2e^{-x}$$

Now, taking the second derivative

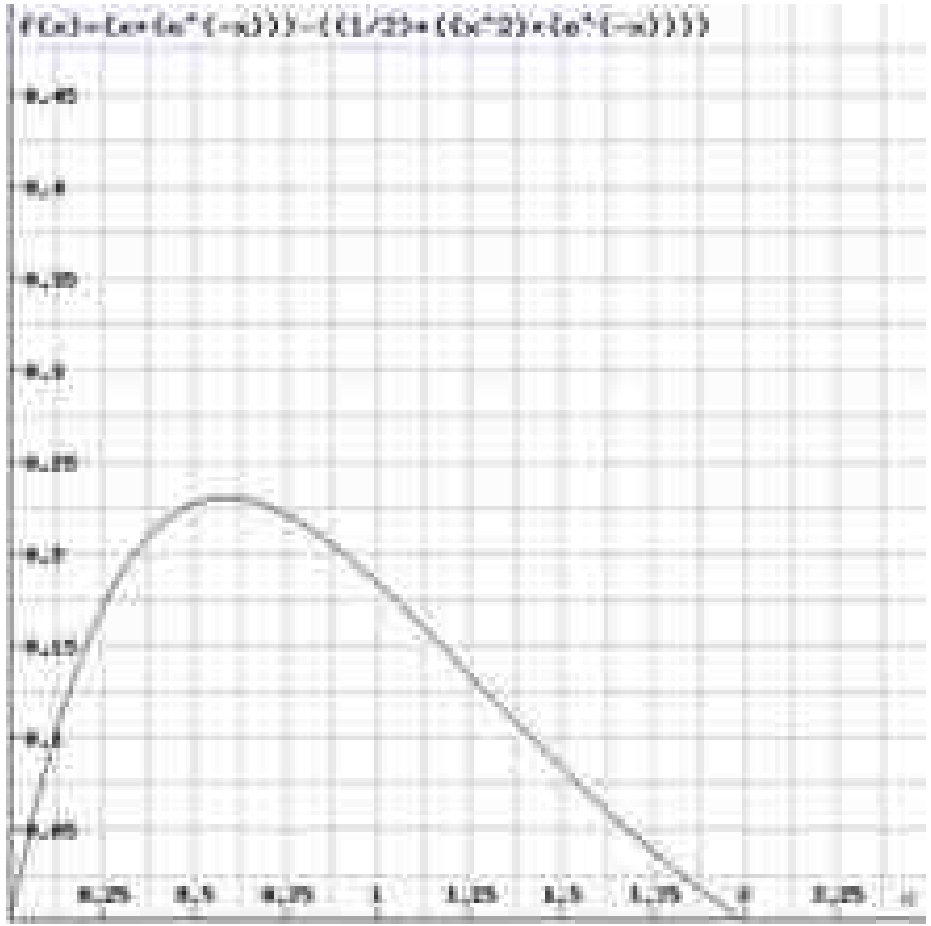
$$f''(x) = \left(x.e^{-x}(-1) + e^{-x}(1)\right) - \frac{1}{2}\left(x^2e^{-x}(-1) + e^{-x}2x\right) = -xe^{-x} + e^{-x} + \frac{1}{2}x^2e^{-x} - xe^{-x}$$

$$f''(x) = e^{-x} - 2xe^{-x} + \frac{1}{2}x^2e^{-x}$$

putting the value of x in the second derivative, we get

$$f''(2) = e^{-2} - 2(2)e^{-2} + \frac{1}{2}(2)^2e^{-2} = e^{-2} - 4e^{-2} + 2e^{-2} = -e^{-2} = -0.135$$ which is less than zero

Therefore, the **MODE** of the distribution is equal to **2**.



Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 41

Median
Of a
Discrete Random Variable

Definition

The median of a distribution of a random variable X of the discrete or continuous type is a value of x such that

If there is only one such x, it is called the **median** of the distribution, $P(X < x) \leq \frac{1}{2}$ *and* $P(X \leq x) \geq \frac{1}{2}$

Example: Find the median of the following distribution:

$$a)\ p(x)=\left(\frac{4!}{x!(4-x)!}\right)\left(\frac{1}{4}\right)^{x}\left(\frac{3}{4}\right)^{4-x}, \qquad x=0,1,2,3,4 \quad zero\ elsewhere$$

x	$P(x)$
0	$\frac{4!}{0!(4-0)!}\left(\frac{1}{4}\right)^{0}\left(\frac{3}{4}\right)^{4-0}=1\times 1\times\left(\frac{3}{4}\right)^{4}=\frac{81}{256}$
1	$\frac{4!}{1!(4-1)!}\left(\frac{1}{4}\right)^{1}\left(\frac{3}{4}\right)^{4-1}=\frac{4!}{3!}\times\frac{1}{4}\times\left(\frac{3}{4}\right)^{3}=\frac{108}{256}$
2	$\frac{4!}{2!(4-2)!}\left(\frac{1}{4}\right)^{2}\left(\frac{3}{4}\right)^{4-2}=\frac{4!}{2!2!}\times\frac{1}{16}\times\left(\frac{3}{4}\right)^{2}=\frac{54}{256}$
3	$\frac{4!}{3!(4-3)!}\left(\frac{1}{4}\right)^{3}\left(\frac{3}{4}\right)^{4-3}=\frac{4!}{3!1!}\times\frac{1}{64}\times\left(\frac{3}{4}\right)^{1}=\frac{12}{256}$
4	$\frac{4!}{4!(4-4)!}\left(\frac{1}{4}\right)^{4}\left(\frac{3}{4}\right)^{4-4}=\frac{4!}{4!0!}\times\frac{1}{256}\times\left(\frac{3}{4}\right)^{0}=\frac{1}{256}$

The median of a distribution of a random variable X of the discrete or continuous type is a value of x such that

$$P(X < x) \leq \frac{1}{2} \text{ and } P(X \leq x) \geq \frac{1}{2}$$

Trial and Error

$$P(X < 2) = \frac{81}{256} + \frac{108}{256} = \frac{189}{256} = 0.74 > 0.5$$

$\therefore$ *2 cannot be the median*

Try

$$P(X < 1) = \frac{81}{256} = 0.32 < 0.5$$

So,

$$P(X \leq 1) = \frac{81}{256} + \frac{108}{256} = \frac{189}{256} = 0.74 > 0.5$$

Both requirements fulfilled $\Rightarrow \tilde{x} = 1$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 38

Median of a Continuous Random Variable

Definition

The median of a distribution of a random variable X of the discrete or continuous type is a value of x such that $P(X<x) \leq \frac{1}{2}$ and $P(X \leq x) \geq \frac{1}{2}$.

If there is only one such x, it is called the **median** of the distribution.

Procedure:

In the continuous case, the median is obtained by integrating the pdf from –inf to M and equating it to 1/2;

by solving this equation for M, we obtain the median.

Find the median of the following distribution:

c) $f(x)= 2/\pi(1+x^2), \quad 0 < x < \infty$

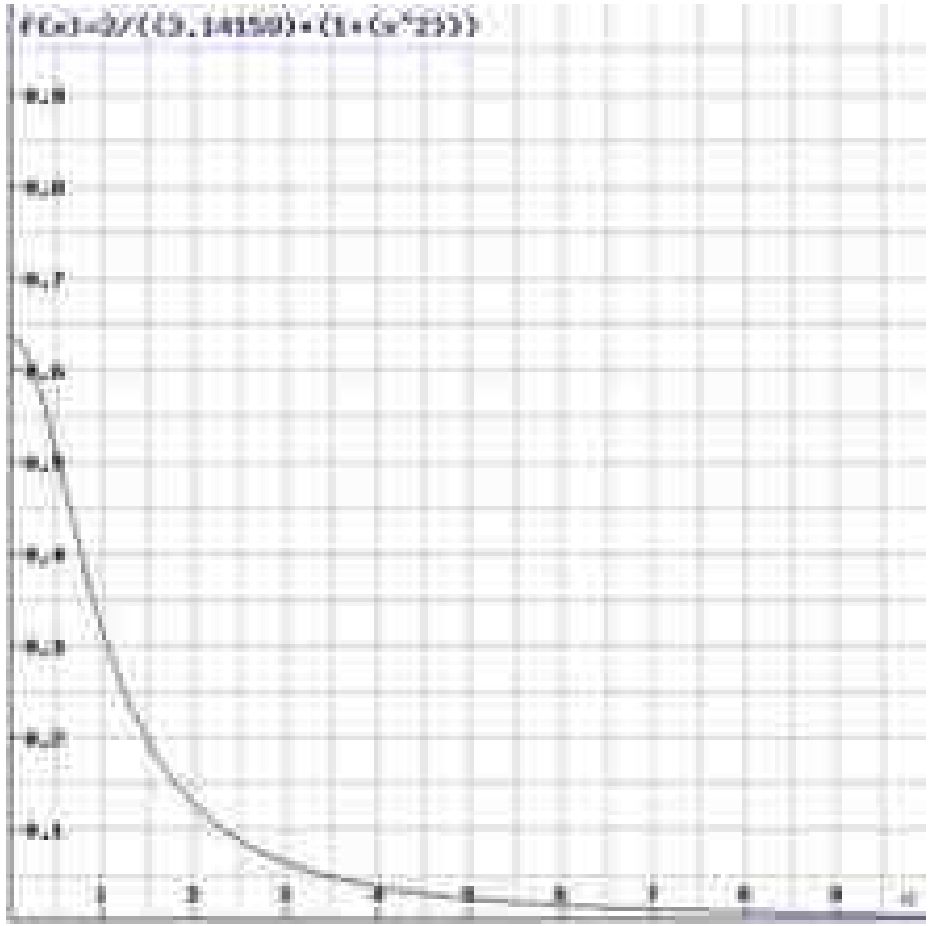
Solution:

$$f(x)=\int_0^M \frac{2}{\pi\left(1+x^2\right)}dx=\frac{1}{2}\Rightarrow\frac{2}{\pi}\int_0^M\frac{1}{\left(1+x^2\right)}dx=\frac{1}{2}\Rightarrow\frac{2}{\pi}\tan^{-1}(x)\Big|_0^M=\frac{1}{2}$$

$$\Rightarrow\frac{2}{\pi}\left[\tan^{-1}(M)-\tan^{-1}(0)\right]_0^M=\frac{1}{2}$$

$$\Rightarrow\left[\tan^{-1}(M)-0\right]=\frac{\pi}{4}\Rightarrow\tan^{-1}(M)=\frac{\pi}{4}$$

$$\Rightarrow M=\tan\left(\frac{\pi}{4}\right)=1\leftarrow \textit{Median}$$



Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 39

Concept of $(100p)^{th}$ percentile (quantile of order p) of a Continuous Random Variable

Definition

Let $0 < p < 1$. A quantile of order p of the distribution of a random variable X is a value ZAAI **p** ξ_p such that $P(X < \xi_p) \leq p$ and $P(X \leq \xi_p) \geq p$.

Note:

A quantile of order p can also be regarded as the $(100 >< p)^{th}$ percentile.

Explanation:

By a quantile of order p, we mean that point on the x-axis to the left of which the area under the curve of the probability density function is equal to p or in other words is equal to $100 >< p$ %.

Now, we know that, the point to the left of which the area is, say, 20% is known as the 20th percentile, the point to the left of which the area is, say, 35% is known as the 35th percentile and so on.

Therefore, that point to the left of which the area under the curve is equal to 100 >< p % will be known as the $(100 >< p)^{th}$ percentile.

Explanation of the way in which these two equations are fulfilled in the case of a continuous variable.

$$P(X < \xi p) \leq p \qquad (1)$$

$$P(X \leq \xi p) \geq p \qquad (2)$$

(1) & (2) need to be fulfilled simultaneously. But the LHS's of both (1) and (2) are the same

$$P(X < c) = P(X \leq c)$$

Hence it is obviously that both (1) & (2) can hold simultaneously of and only of the two inequalities are replaced by the equal sign

$$P(X < \xi p) = P(X \leq \xi p) = p$$

$$or \qquad \int_{-\infty}^{\xi p} f(x)\,dx = p$$

Find the 20th percentile of the distribution that has pdf $f(x)=4x^3$, $0< x< 1$, zero elsewhere.

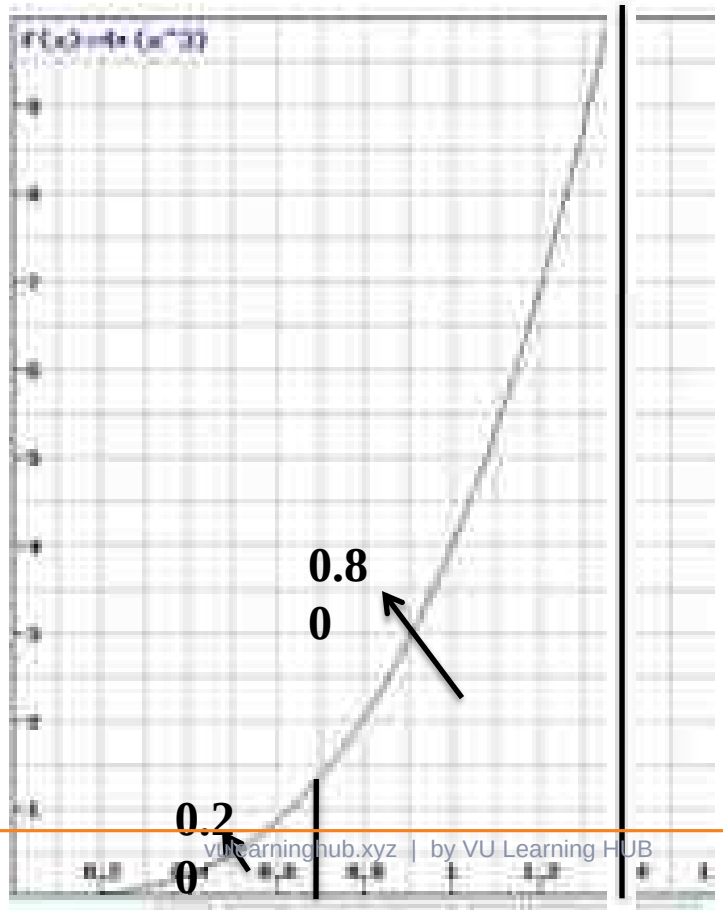
Hint: With a continuous-type random variable X, $P(X< \xi_p) = P(X\leq \xi_p)$ and hence that common value must equal p.

$\xi_{0.20}$:

$$\int_0^{\xi_{0.20}} 4x^3 dx = 0.20$$

$$\Rightarrow 4\frac{x^4}{4}\Big|_0^{\xi_{0.20}} = 0.20 \Rightarrow \xi_{0.20}^4 - 0^4 = 0.20 \Rightarrow \xi_{0.20}^4 = 0.20 \Rightarrow \xi_{0.20} = (0.20)^{\frac{1}{4}}$$

$$\Rightarrow \xi_{0.20} = 0.6687 \cong 0.67$$



vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 40

Real-Life Example

of the computation of the $(100p)^{th}$ percentile (quantile of order p) of a Continuous Random Variable

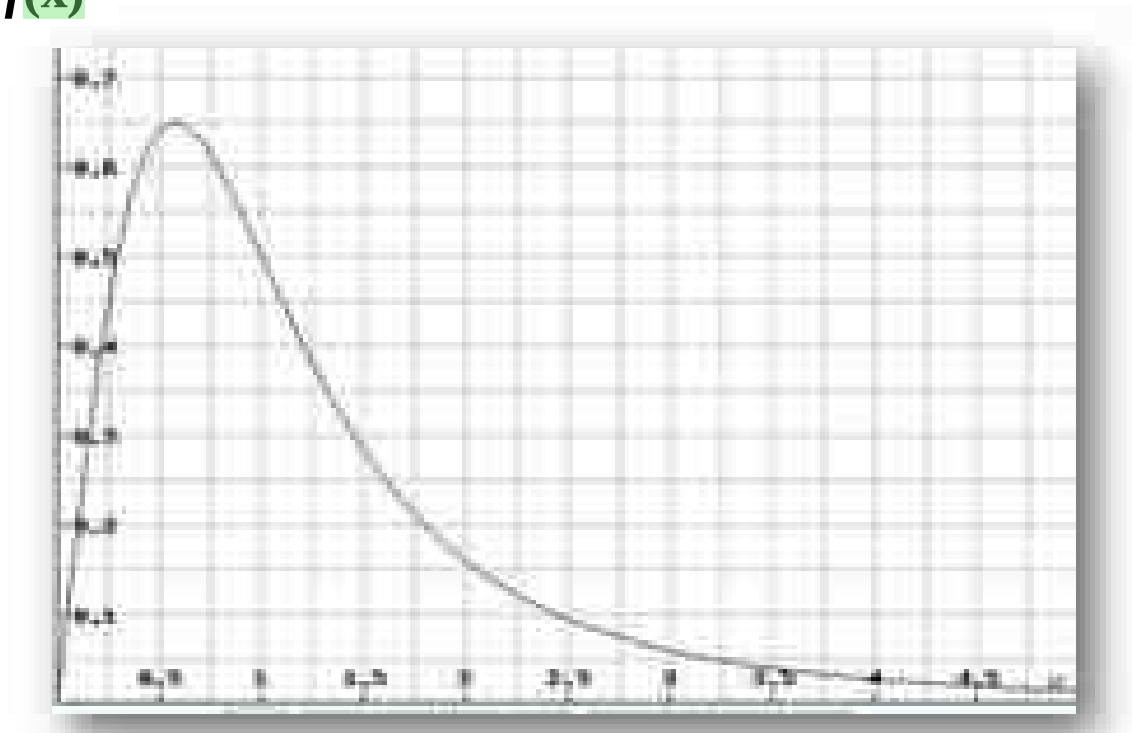
Let X be the number of gallons of ice cream that is requested at a certain store on a hot summer day.

Assume that X follows the log-logistic distribution with $\alpha=1$ and $\beta=2$. In other words, the pdf of X is given by

$$f(x)=\frac{2x}{\left(1+x^2\right)^2},\ 0<x<\infty$$

How many gallons of ice cream should the store have on hand each of these days, so that the probability of exhausting its supply on a particular day is 0.10?

Solution: $f(x)$



where X represents the number of gallons requested i.e., the demand

The store will be exhausting its supply only if the demand **exceeds** the amount in hand.

We want the probability of this circumstance to be only 0.10.

Hence, it is obvious that we are talking about 90th percentile of the distribution of the random variable X.

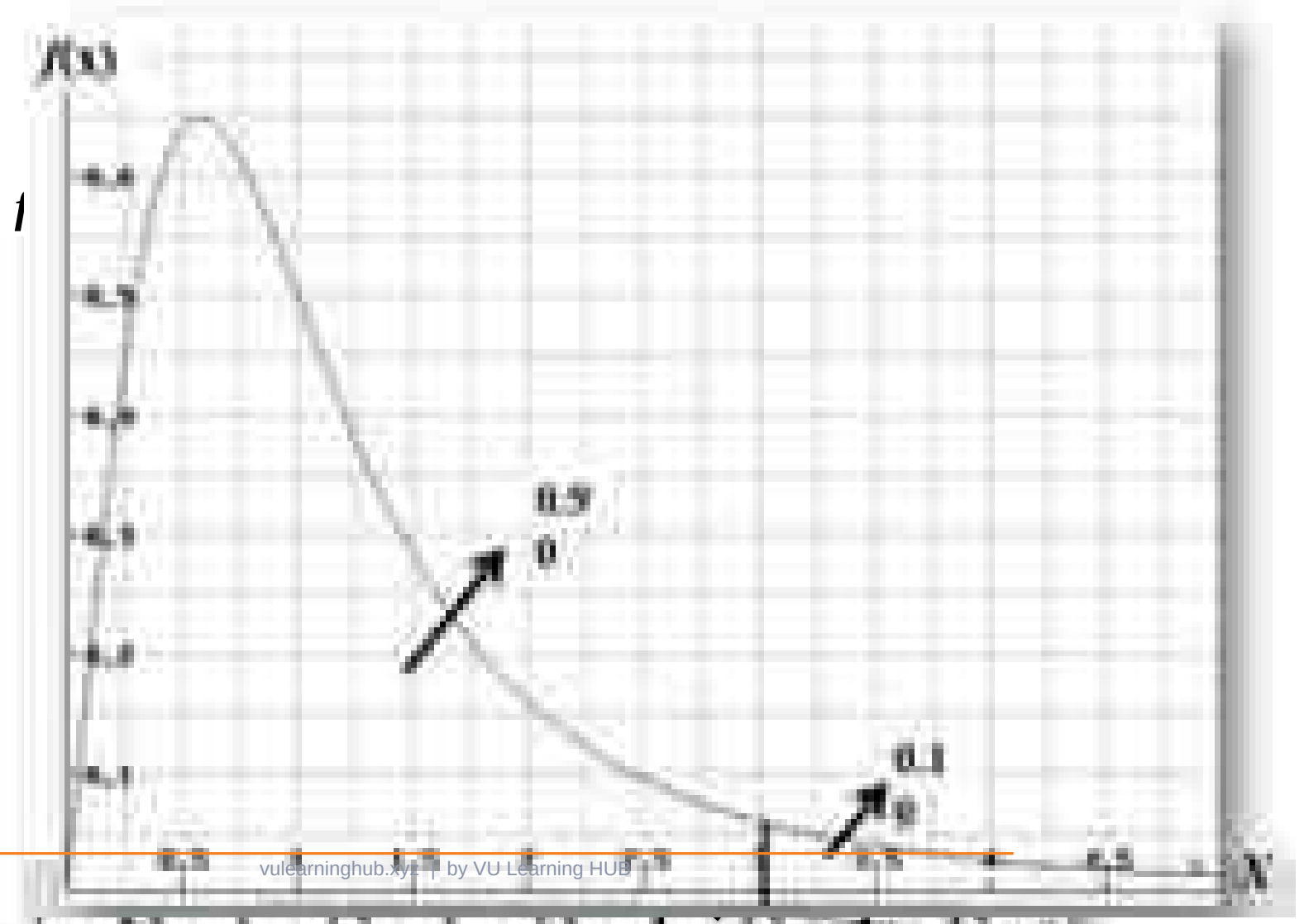
The pdf of X is given by $f(x)=\dfrac{2x}{\left(1+x^2\right)^2},\ 0<x<\infty,$ *when* α=1and β=2

$$f(x)=\int_0^{\xi_{0.90}}\frac{2x}{\left(1+x^2\right)^2}dx=0.90\Rightarrow-\frac{1}{1+x^2}\Big|_0^{\xi_{0.90}}=0.90$$

$$\Rightarrow\left(\frac{1}{1+\xi_{0.90}^2}-\frac{1}{1+0}\right)=-0.90\Rightarrow\frac{1}{1+\xi_{0.90}^2}-1=-0.90$$

$$\Rightarrow\frac{1}{1+\xi_{0.90}^2}=1-0.90\Rightarrow\frac{1}{0.10}=1+\xi_{0.90}^2\Rightarrow\frac{1}{0.10}-1=\xi_{0.90}^2\Rightarrow9=\xi_{0.90}^2\Rightarrow\sqrt{\xi_{0.90}^2}=\sqrt{9}$$

$$\therefore\xi_{0.90}=3$$



vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 41

Inverse CDF

or

Quantile function

- The **quantile function** is also called the **inverse cumulative distribution function**.

For a given probability distribution,
the **quantile function** specifies the quantile of order p for all values of p lying between 0 and 1.

In other words, the **quantile function** specifies the **value** on the x-axis for which the probability of the random variable X being less than or equal to this value is p (for all values of p lying between 0 and 1).

Example

Suppose we have the Exponential distribution

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

Expected value or (mean) $1/\lambda$

The **cumulative distribution function** of Exponential Distribution

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x < 0 \end{cases}$$

The quantile function for Exponential(λ) is derived by finding the value of x for which

$$\Rightarrow p = 1 - e^{-\lambda x} \Rightarrow e^{-\lambda x} = 1 - p$$

Taking log on both sides

$$\Rightarrow \ln e^{-\lambda x} = \ln(1-p) = -\lambda x \ln e = \ln(1-p)$$

$$\Rightarrow -\lambda x = \ln(1-p) \qquad \because \ln e = 1$$

$$\therefore x = -\frac{\ln(1-p)}{\lambda}$$

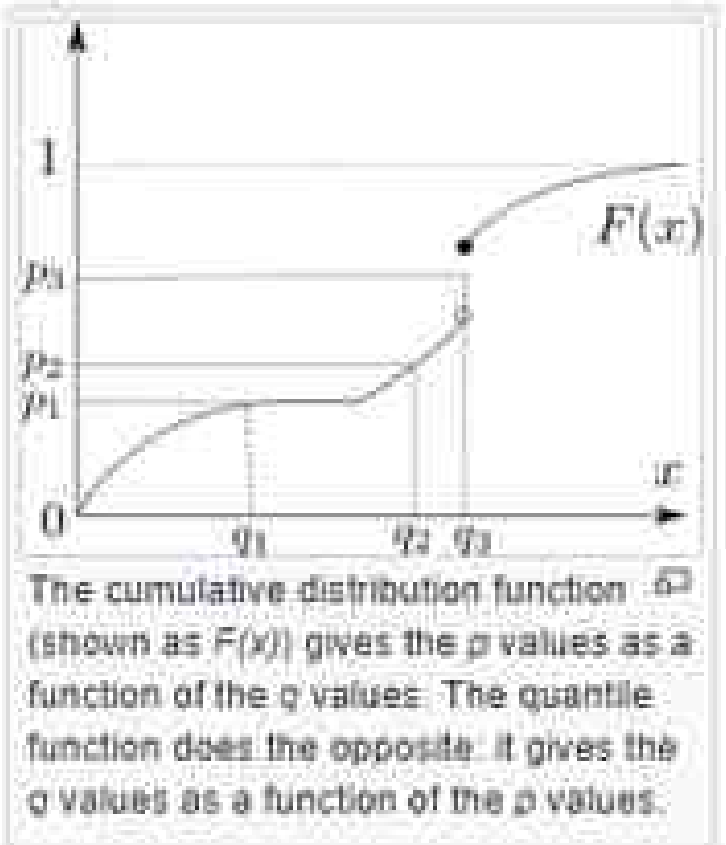
$$Q(p;\lambda) = -\frac{\ln(1-p)}{\lambda}, \text{ for } 0 < p < 1.$$

The quartiles are therefore,

first quartile (p = 1/4)= ln (4/3)/ λ,

median (p=2/4)= ln (2) / λ

and **third quartile (p=3/4)** = ln (4)/ λ



The cumulative distribution function (shown as $F(x)$) gives the p values as a function of the q values. The quantile function does the opposite: it gives the q values as a function of the p values.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 42

Example of a Random Variable being Stochastically Larger than another random variable

Definition

A random variable X is said to be stochastically larger than another random variable Y if

$$P(X > z) \geq P(Y > z),$$

for all real values of z, with strict inequality holding for at least one z value.

This requires that the CDFs enjoy the following property:

for all real z, with strict inequality holding for at least one z value.

$$F_X(z) \leq F_Y(z),$$

This requires that the CDFs enjoy the following property:

$$F_Y(z) \le F_X(z),$$

Now, $F_X(z) = P(X \le z)$

whereas, $F_Y(z) = P(Y \le z) = P(X + \Delta \le z) = P(X \le z - \Delta)$

$$P(X \le z - \Delta) < P(X \le z) \qquad (1)$$

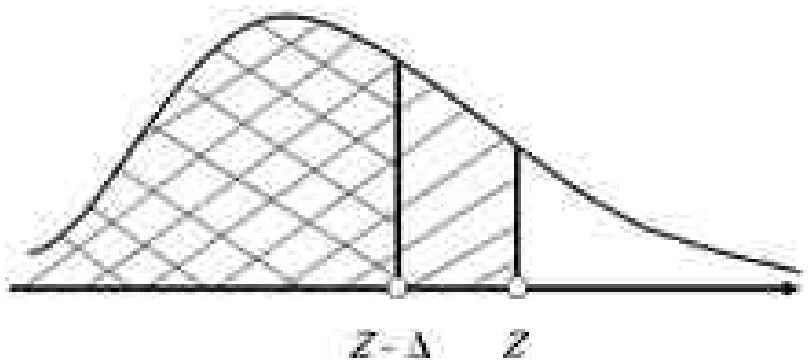
eq. (1) can be re-write as

$$P(X + \Delta \le z) < P(X \le z)$$

or

$$P(Y \le z) < P(X \le z)$$

or $F_Y(z) \le F_X(z).$



Example:

Let X be a continuous random variable with support $(-\infty, \infty)$. Consider the random variable $Y = X + \Delta$, where $\Delta > 0$. Suppose that we want to show that Y is stochastically larger than X.

Solution:

First and foremost, let us re-state the above definition interchanging the roles of X and Y.

This requires that the CDFs enjoy the following property:

$$F_Y(z) \le F_X(z),$$

Now, $F_X(z) = P(X \le z)$

whereas, $F_Y(z) = P(Y \le z) = P(X + \Delta \le z) = P(X \le z - \Delta)$

$$P(X \le z - \Delta) < P(X \le z) \qquad (1)$$

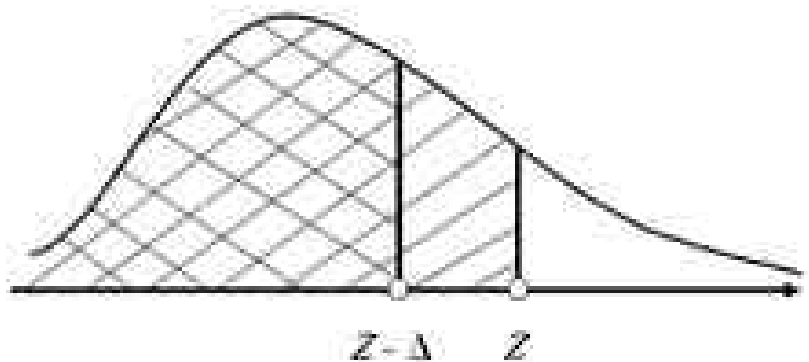
eq. (1) can be re-write as

$$P(X + \Delta \le z) < P(X \le z)$$

or

$$P(Y \le z) < P(X \le z)$$

or $F_Y(z) \le F_X(z).$



Therefore the random variable $Y = X + \Delta,$ where $\Delta > 0.$

has showed that Y **is stochastically larger than** X**.**

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 43

Concept of Mathematical Expectation

for Discrete and Continuous Random Variables

Definition 8.1. (Expectation). Let X be a random variable. If X is a continuous random variable with pdf $f(x)$ and

$$\int_{-\infty}^{\infty} |x| \, f(x)\, dx < \infty,$$

then the expectation of X is

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

vulearninghub.xyz | by VU Learning HUB

If X is a discrete random variable with pmf p(x) and

$$\sum_x |x|\, p(x)\, dx < \infty$$

Then the expectation of X is

$$E(X) = \sum_x x\, p(x).$$

Sometimes the expectation E(X) is called the mathematical expectation of X, the expected value of X, or the mean of X. When the mean designation is used, we often denote the $E(X)$ by μ; i.e., $\mu = E(X)$.

Example

Suppose that we have a discrete random variable X given by -1, 0, and 1. and the probabilities of P(-1) = ¼, P(0)= ½, P(1) = ¼.

So,

$$P(-1)+P(0)+P(1)=\frac{1}{4}+\frac{1}{2}+\frac{1}{4}=\frac{1+2+1}{4}=\frac{4}{4}=1.$$

Now, to find the Mean

So,

$$Mean = -\frac{1}{4} + 0 + \frac{1}{4} = 0.$$

x	$p(x)$	$xp(x)$
-1	1/4	-1/4
0	1/2	0
1	1/4	1/4

Now to fullfil the condition to get a finite number

So,

$$\sum_x |x|\, p(x)\, dx < \infty$$

$$= \frac{1}{4} + 0 + \frac{1}{4} = \frac{2}{4} = \frac{1}{2} = 0.5.$$

x	$p(x)$	$xp(x)$	$\lvert x\rvert$	$\lvert x\rvert p(x)$
-1	1/4	-1/4	1	1/4
0	1/2	0	0	0
1	1/4	1/4	1	1/4

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 44

Concept of Mathematical Expectation

of a Function of a Random Variable X

(for discrete and continuous random variables)

Theorem :

Let X be a random variable and let $Y = g(x)$ for some function g.

a) Suppose X is continuous with pdf $f_X(x)$.

If $\int_{-\infty}^{\infty} |g(x)| f_X(x)\,dx < \infty,$ then the expectation of Y exists and it is given by

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)\,dx. \qquad (1)$$

b) Suppose X is discrete with pmf $p_X(x)$. Suppose the support of X is denoted by S_X . If $\sum_{x \in S_X} |g(x)| p_X(x) < \infty,$ then the expectation of Y exists and it is given by

$$E(Y) = \sum_{x \in S_X} g(x) p_X(x). \qquad (2)$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 45

Concept of Linear combination of the Expected Values of two different functions of a Random Variable ‘X’

(for discrete and continuous random variables)

Theorem 1: Let $g_1(X)$ and $g_2(X)$ be functions of a random variable X. Suppose the expectations of $g_1(X)$ and $g_2(X)$ exists. Then for any constants k_1 and k_2, the expectation of $k_1 g_1(X) + k_2 g_2(X)$ exists and it is given by

$$E\left[k_1 g_1(X) + k_2 g_2(X)\right] = k_1 E\left[g_1(X)\right] + k_2 E\left[g_2(X)\right]$$

Suppose that we have a discrete random variable X that have the values -1, 0, 1 and the probabilities of three values are P(-1) = 1/4 , P(0)= 1/2 and P(1)=1/4.

Now let us consider two different functions of the r.v. 'X'

Let $g_1(X)$ b given by X+2 and $g_2(X)$ be given by X+5

And k_1 =2 and k_2 = 5,

So, let us construct a table:

x	$p(x)$	$g_1(X)$ $=X+2$	$k_1g_1(X)$	$g_2(X)$ $=X+5$	$k_2g_2(X)$	$k_1g_1(X)$ $+ k_2g_2(X)$	$p(x)(k_1g_1(X) + k_2g_2(X))$
-1	-¼	1	2	4	20	22	-22/4
0	½	2	4	5	25	29	29/2
1	¼	3	6	6	30	36	36/4
							72/4=**18**

In the same way, you can find the R.H.S of the formula and will find

$$E\left[k_1 g_1(X)+k_2 g_2(X)\right]=k_1 E\left[g_1(X)\right]+k_2 E\left[g_2(X)\right]$$

That both sides are equal as it was required.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 46

Example of computing mathematical Expectation of a Function of a Discrete Random Variable ‘X’

Example: Let the pmf p(x) be positive at x = -1, 0, 1 and zero elsewhere.

a) If p(0) = 1/4 , find E(X^2).

x	x^2
-1	1
0	0
1	1

Solution:

First we find the pmf for X^2 .The random variable X^2 takes on the values 0 and 1.

Now,

$$P\left(X^2=0\right)=P\left(X=0\right)=\frac{1}{4},$$

and

$$P\left(X^2=1\right)=P\left[\left(X=-1\right)\cup P\left(X=1\right)\right]$$

$$=p\left(-1\right)+p\left(1\right)$$

$$=1-p\left(0\right)$$

$$=\frac{3}{4}$$

It then follows that

$$\mathrm{E}\left(X^2\right)=0.P\left(X^2=0\right)+1.P\left(X^2=1\right)=\frac{3}{4}.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 47

Another Example of computing mathematical Expectation of a Function of a Discrete Random Variable '*X*'

Example: Let the pmf p(x) be positive at x = -1, 0, 1 and zero elsewhere.

b) If p(0) = 1/4 and if E(X) = 1/4, determine p(-1) and p(1).

Solution:

From

$$E(X) = (-1)\,p(-1) + 0\,p(0) + 1\,p(1)$$

$$= (-1)\,p(-1) + 1\,p(1) = \frac{1}{4},$$

and

$$p(-1) + p(0) + p(1) = 1$$

we obtain the equations

$$\begin{cases} p(-1) - p(1) = -1/4, & (1) \\ p(-1) + p(1) = 3/4 & (2) \end{cases}$$

Solving the equations simultaneously yields

$$p(-1) = \frac{1}{4},$$

and $$p(1) = \frac{1}{2}.$$

Therefore, the sum of the probabilities are

$$p(-1)+p(0)+p(1)=\frac{1}{4}+\frac{1}{4}+\frac{1}{2}$$

$$=\frac{1+1+2}{4}=\frac{4}{4}=1.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 48

Proof of the fact that $E(X^2)$ is greater than or equal to $[E(X)]^2$

Theorem

If the variance of the random variable X exists, then

$$E\left(X^2\right) \geq \left[E\left(X\right)\right]^2$$

Now

$$\text{Var}(X) = E\left[X - E(X)\right]^2$$

$$= E\left[X^2 - 2XE(X) + \left[E(X)\right]^2\right]$$

$$= E\left[X^2\right] - 2E(X)E(X) + \left[E(X)\right]^2$$

$$= E\left[X^2\right] - 2\left[E(X)\right]^2 + \left[E(X)\right]^2$$

$$= E\left[X^2\right] - \left[E(X)\right]^2$$

Hence, it is established that

$$E\left(X^2\right)-\left[E\left(X\right)\right]^2=E\left[X-E\left(X\right)\right]^2$$

But $$\left[X-E\left(X\right)\right]^2\geq 0$$

as the square of any quantity has to be nonnegative.

Therefore, $$E\left(X-E\left(X\right)\right)^2\geq 0$$

$$E\left(X^2\right)\geq\left[E\left(X\right)\right]^2\Rightarrow E\left(X^2\right)-\left[E\left(X\right)\right]^2\geq 0$$

Hence proved.

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 49

Mean, Variance and Standard deviation of a Random Variable

Mean

First, let X be a random variable of the discrete type with pmf $p(x)$.

Then

$$E(X) = \sum_{x} xp(x).$$

If the support of X is $\{a_1, a_2, a_3, \ldots\}$, it follows that'

$$E(X) = a_1 p(a_1) + a_2 p(a_2) + a_3 p(a_3) + \ldots$$

This sum of products is seen to be a "weighted average" of the values of $a_1, a_2, a_3, \ldots$ the "weight" associated with each a_i being $p(a_i)$.

$$Let \;\; a_1, a_2, a_3, ..., a_n$$

$$= \frac{a_1 + a_2 + a_3 + ... + a_n}{n} = \frac{a_1}{n} + \frac{a_2}{n} + \frac{a_3}{n} + \cdots + \frac{a_n}{n}$$

$$= \left(a_1 \frac{1}{n} + a_2 \frac{1}{n} + a_3 \frac{1}{n} + \cdots + a_n \frac{1}{n} \right)$$

This suggests that we call E(X) the arithmetic mean of the values of X, or, more simply,

the mean value of X (or the mean value of the distribution).

vulearninghub.xyz | by VU Learning HUB

Variance

Let X be a discrete random variable with support $\{a_1, a_2, a_3, \ldots\}$, and with pmf $p(x)$, then

$$E\left[\left(X-\mu\right)^2\right]=\sum_x\left(x-\mu\right)^2\ p\left(x\right)=\left(a_1-\mu\right)^2\ p\left(a_1\right)+\left(a_2-\mu\right)^2\ p\left(a_2\right)+\ldots$$

This sum of products may be interpreted as a "weighted average" of the squares of the deviations of the numbers $a_1, a_2, a_3, \ldots$ from the mean value μ,

where the "weight" associated with each $(a_i - \mu)^2$ *is* $p(a_i)$.

Standard deviation

The positive square root of the variance is called Standard deviation.

$$\sigma = \sqrt{E\left[\left(X - \mu\right)^2\right]}$$

Short-cut formula

$$\sigma^2 = E\left[\left(X-\mu\right)^2\right] = E\left(X^2 - 2\mu X + \mu^2\right);$$

and since E is a linear operator,

$$\sigma^2 = E\left(X^2\right) - 2\mu E\left(X\right) + \mu^2$$

$$= E\left(X^2\right) - 2\mu^2 + \mu^2$$

$$= E\left(X^2\right) - \mu^2,$$

This frequently affords an easier way of computing the variance of X.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 50

The Concept of Degenerate Random Variable and one of its basic properties

Definition

A degenerate distribution is a probability distribution with support only on **a single point**.

Examples

- A coin both sides of which are showing a head
- A die all six sides of which are showing the same number.

This distribution satisfies the definition of **“random variable”** even though it does not appear **random** in the everyday sense of the word; hence it is considered **degenerate**.

Mean and Variance of a Degenerate Random Variable

If the space contains only one point k for which

$$p(k) > 0,$$

Then

$$p(k) = 1,\ \mu = k,\ \text{and}\ \sigma^2 = 0.$$

Explanation:

Let us consider another example, suppose that a teacher administers a test out of 10 in a class and k= 20.

If every single student get 7 out of 10 then, the mean mark will be 7.

In this case, the mean of the random variable X is also equal to k and the variance of the random variable X is equal to zero.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 51

Proof of the fact that the Mean of a symmetric distribution lies at the point of symmetry

Example

Let a random variable X of the continuous type have a pdf $f(x)$ whose graph is symmetric with respect to $x = c$.

If the mean value of X exists, show that

$$E(X) = c.$$

Hint: Show that $E(X - c)$ equals zero

by writing $E(X - c)$ as the sum of two integrals:

Where one from $-\infty$ to c and the other from c to ∞.

In the first, let $y = c - x$; and, in the second, $z = x - c$.

Finally, use the symmetry condition $f(c - y) = f(c + y)$ in the first.

Solution:

Given that $f(c\text{-}x) = f(c+x)$, we will show that $E(X - c) = E(X) - c = 0$.

$$E[X-c] = \int_{-\infty}^{\infty} (x-c)\, f(x)\, dx$$

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

vulearninghub.xyz | by VU Learning HUB

In the first integral,

let $y = c - x \Rightarrow dy = -dx$

and $x \to -\infty \Rightarrow y \to \infty$, $x \to c \Rightarrow y \to 0$

and in the second integral,

let $z = x - c \Rightarrow dz = dx$

and $x \to c \Rightarrow z \to 0$, $x \to \infty \Rightarrow z \to \infty$.

We have

$$E(X-c)=\int_{-\infty}^{c}(x-c)\,f(x)\,dx+\int_{c}^{\infty}(x-c)\,f(x)\,dx.$$

Applying the above transfornmations, we obtain

$$E(X-c)=-\int_{\infty}^{0}(-y)\,f(c-y)\,dy+\int_{0}^{\infty}zf(c+z)\,dz$$

$$=-\int_{0}^{\infty}yf(c-y)\,dy+\int_{0}^{\infty}zf(c+z)\,dz$$

Now, let us focus on the first integral:

Since $f(.)$ is symmetric about c,

therefore $f(c-y) = f(c+y)$

$$\Rightarrow -\int_0^{\infty} yf(c-y)\,dy = -\int_0^{\infty} yf(c+y)\,dy$$

$$= -\int_0^{\infty} zf(c+z)\,dz$$

using the concept of dummy variable.

Hence

$$E(X-c) = -\int_0^{\infty} zf(c+z)\,dz + \int_0^{\infty} zf(c+z)\,dz = 0.$$

Hence, we conclude that if the density function for a random variable X is symmetric about c, then $\mu = E[X] = c$.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 52

Proof of the fact that the Mean of a Standardized variable is zero and the variance is 1.

Proof:

Let the random variable X have mean μ and standard deviation σ,

$-h < t < h$. Show that

$$E\left(\frac{X-\mu}{\sigma}\right)=0,\ E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]=1.$$

Using the linear properties of expected value and the definition of $\mu = E[X]$, we calculate

$$E\left[\frac{X-\mu}{\sigma}\right] = \frac{1}{\sigma}E\left[X-\mu\right] = \frac{1}{\sigma}\left(E\left[X\right]-\mu\right) = \frac{\mu-\mu}{\sigma} = 0,$$

which verifies the first equation.

Using the linear properties of expected value again and the definition of $\sigma^2 = E[(X - \mu)^2]$, we calculate

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = E\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{E\left[(X-\mu)^2\right]}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1,$$

which verifies the second equation.

Note:

$$E\left(\frac{X-\mu}{\sigma}\right)=0,\ E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right]=1$$

implies that

$$\text{Var}\left(\frac{X-\mu}{\sigma}\right)=1.$$

vulearninghub.xyz by VU Learning HUB

Proof:

$$\text{Var}(w) = E\left[(w)^2\right] - \left[E(w)\right]^2, \text{we have}$$

$$\text{Var}\left(\frac{X-\mu}{\sigma}\right) = E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] - \left[E\left(\frac{X-\mu}{\sigma}\right)\right]^2$$

$$= 1 - 0^2 = 1 - 0 = 1.$$

Hence, the mean and standard deviation of a standardized variable are 0 and 1 respectively.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

Topic No. 53

The Concept of

Moments

m^{th} Moment about Arbitrary origin

Consider the expression $E(X-a)^m$

where m is a positive integer.

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

or

$$E(X-a)^m = \sum_x (x-a)^m p(x)$$

depending on whether X is a continuous or a discrete random variable

and it is known as the mth moment about an Arbitrary origin.

Notation

for m^{th} Moment about Arbitrary origin:

$$\mu_m^{/} = E\left(X - a\right)^m$$

m^{th} Moment about the Mean

In the expression $E\left(X-a\right)^m$

if we put $a = \mu$, we obtain

$$E\left(X-\mu\right)^m = \int_{-\infty}^{\infty}\left(x-\mu\right)^m f\left(x\right)dx$$

or

$$E\left(X-\mu\right)^m = \sum_{x}\left(x-\mu\right)^m p\left(x\right)$$

depending on whether X is a continuous or a discrete random variable

and it is known as the mth moment about the Mean.

vulearninghub.xyz | by VU Learning HUB

Notation for m^{th} Moment about The Mean:

$$\mu_m = E\left(X - \mu\right)^m$$

Moments about the Mean are also known as **Central Moments.**

m^{th} Moment about the Origin

In the expression $E\left(X-a\right)^m$

if we put $a=0$, we obtain

$$E\left(X-0\right)^m = E\left(X^m\right) = \int_{-\infty}^{\infty} x^m f\left(x\right) dx$$

or

$$E\left(X-0\right)^m = E\left(X^m\right) = \sum_x x^m p\left(x\right)$$

depending on whether X is a continuous or a discrete random variable

and it is known as the mth moment about the Origin.

vulearninghub.xyz | by VU Learning HUB

Notation

for m^{th} Moment about The Origin:

$$\mu_m^{/} = E\left(X^m\right)$$

A Special Case:

In the mth moment about the Origin, if we put $m = 1$, we obtain

$$\mu_1' = E\left(X^1\right) = E\left(X\right) = \int_{-\infty}^{\infty} xf\left(x\right)dx$$

or

$$\mu_1' = E\left(X^1\right) = E\left(X\right) = \sum_{x} xp\left(x\right)$$

implying that the 1st moment about the Origin is the Mean of the distribution.

Another Special Case:

In the mth moment about the Mean, if we put m = 2, we obtain

$$\mu_2 = E\left(X-\mu\right)^2 = \int_{-\infty}^{\infty}\left(x-\mu\right)^2 f\left(x\right)dx$$

or

$$\mu_2 = E\left(X-\mu\right)^2 = \sum_{x}\left(x-\mu\right)^2 p\left(x\right)$$

implying that the 2nd moment about the mean is the Variance of the distribution.

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 54

Moment Ratios depicting

Skewness (β_1) and Kurtosis (β_2)

Measures of Skewness

Let X be a random variable with mean μ and variance σ^2 (i.e. standard deviation σ) such that the third moment about μ i.e. $\mu_3 = E\left[\left(X-\mu\right)^3\right]$ exists.

The value of the ratio

$$\frac{E\left[\left(X-\mu\right)^3\right]}{\sigma^3} = \frac{\mu_3}{\sigma^3}$$

is often used as a measure of skewness.

It is important to note that this measure of skewness is

- negative for distributions that are skewed to the left,
- zero for distributions that are not skewed (symmetric),

and

- Positive for distributions that are skewed to the right.

The square of this ratio i.e.

$$\frac{\mu_3^2}{\left(\sigma^3\right)^2} = \frac{\mu_3^2}{\left(\sigma^2\right)^3} = \frac{\mu_3^2}{\left(\mu_2\right)^3}$$

is known as the 'first' moment-ratio and is denoted by β_1 i.e.

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

Measure of Kurtosis

Let X be a random variable with mean μ and variance σ^2 (i.e. standard deviation σ) such that the fourth moment about μ i.e. $\mu_3 = E\left[\left(X-\mu\right)^4\right]$ exists.

The value of the ratio

$$\frac{E\left[\left(X-\mu\right)^4\right]}{\sigma^4} = \frac{\mu_4}{\left(\sigma^2\right)^2} = \frac{\mu_4}{\left(\mu_2\right)^2}$$

is often used as a measure of kurtosis.

yulearninghub.xyz | by VU Learning HUB

This ratio is known as the 'second' moment-ratio

and is denoted by β_2 i.e.

$$\beta_2 = \frac{\mu_4}{{\mu_2}^2}$$

It is important to note that the numerical value of this measure of kurtosis is

- greater than 3 for distributions that are leptokurtic,
- equal to 3 for distributions that are mesokurtic, and
- less than 3 for distributions that are platykurtic.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 55

The Concept of Moment Generating Function

Definition:

Let X be a random variable such that for some $h > 0$, the expectation of e^{tX} exists for

$$-h < t < h.$$

The **moment generating function** of X is defined to be the function

$$M(t) = E\left(e^{tX}\right) \qquad -h < t < h.$$

We use the abbreviation **mgf** to denote the moment generating function of a random variable.

vulearninghub.xyz | by VU Learning HUB

Role of mgf

In general, if m is a positive integer and if $M^{(m)}(t)$ means the mth derivative of $M(t)$, we have, by repeated differentiation with respect to t,

$$M^{(m)}(0) = E\left(X^m\right).$$

Since $M(t)$ generates the values of $E(X^m)$; $m = 1,2,3,\ldots$, it is called the **moment-generating function** (mgf).

vulearninghub.xyz | By VU Learning Hub

In the same way, if we take the second derivative of the first derivative, and putting $t = 0$, we will get the second moment about the origin.

Example

Consider the exponential distribution given by

$$f\left(x;\lambda\right)=\lambda e^{-\lambda x},\ 0<x<\infty,\ \lambda>0$$

Find the mgf of X and use it to find the mean and variance of the exponential distribution.

Solution

By definition, we have

$$M(t) = E\left(e^{tX}\right) = \lambda \int_0^{\infty} e^{tx} e^{-\lambda x} dx$$

$$= \lambda \int_0^{\infty} e^{x(t-\lambda)} dx = \lambda \left(\frac{e^{x(t-\lambda)}}{(t-\lambda)} \right)_0^{\infty}$$

$$= \frac{\lambda}{(t-\lambda)} \left(e^{x(t-\lambda)} \right)_0^{\infty}$$

Now, if $\lambda > t$, then $t - \lambda < 0$

and hence

$$M(t) = \frac{\lambda}{(t-\lambda)}\left(e^{-\infty} - e^{0}\right)$$

$$= \frac{\lambda}{(t-\lambda)}(0-1) = \frac{\lambda}{(\lambda - t)}$$

for $t < \lambda$.

Now

$$M^{(1)}(t) = M'(t) = \frac{d}{dt}\left(M(t)\right)$$

$$= \frac{d}{dt}\left(\frac{\lambda}{(\lambda - t)}\right) = \lambda \frac{d}{dt}\left((\lambda - t)^{-1}\right)$$

$$= -\lambda\left((\lambda - t)^{-2}(0 - 1)\right)$$

$$= \lambda\left((\lambda - t)^{-2}\right) = \frac{\lambda}{(\lambda - t)^2}$$

for $t < \lambda$.

vulearninghub.xyz by VU Learning HUB

Evaluating the derivative at $t=0$, we obtain

$$M^{(1)}(0) = \frac{\lambda}{(\lambda - 0)^2} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

which is the first moment about zero

i.e. $E(X)$

i.e. the well-known mean of the distribution.

Next

$$M^{(2)}(t) = M''(t) = \frac{d}{dt}\left(\frac{d}{dt}\left(M(t)\right)\right)$$

$$= \frac{d}{dt}\left(\frac{\lambda}{(\lambda - t)^2}\right) = \lambda \frac{d}{dt}\left((\lambda - t)^{-2}\right)$$

$$= (-2)\lambda\left((\lambda - t)^{-3}(0 - 1)\right)$$

$$= 2\lambda\left((\lambda - t)^{-3}\right) = \frac{2\lambda}{(\lambda - t)^3}$$

for $t < \lambda$.

Evaluating the derivative at $t=0$, we obtain

$$M^{(2)}(0) = M''(0) = \frac{2\lambda}{(\lambda - 0)^3} = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

which is the second moment about zero

i.e. $E\left(X^2\right)$.

Using the short-cut formula for the variance, we obtain

$$Var(X) = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2$$

$$= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

which is well-known.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 56

Algebraic expressions of Some well-known MGFs

Discrete Distributions

Distribution	Moment Generating Function (MGF)
Bernoulli $P(X=1) = p$	$1-p+pe^t$
Geometric $(1-p)^{k-1}p$	$\frac{pe^t}{1-(1-p)e^t} \quad \forall t < -\ln(1-p)$
Binomial $B(n, p)$	$\left(1-p+pe^t\right)^n$
Poisson λ	$e^{\lambda\left(e^t-1\right)}$
Negative binomial NB(r, p)	$\frac{(1-p)^r}{\left(1-pe^t\right)^r}$
Uniform (discrete) $U(a, b)$	$\frac{e^{at}-e^{(b+1)t}}{(b-a+1)(1-e^t)}$

vulearninghub.xyz | by VU Learning HUB

Continuous Distributions

Distribution	Moment Generating Function (MGF)
Uniform (continuous) $U(a\ ,b)$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
Normal $N(\mu\text{-}\sigma^2)$	$e^{t\mu+\frac{1}{2}\sigma^2t^2}$
Chi-Squared $\chi^2{}_k$	$(1-2t)^{-\frac{k}{2}}$
Gamma $\Gamma\ (k,\theta)$	$(1-t\theta)^{-k}\ ;\forall t<\frac{1}{\theta}$
Exponential λ	$\left(1-t\lambda^{-1}\right)^{1},(t<\lambda)$

Distribution	Moment Generating Function (MGF)
Laplace $L(\mu\ ,b)$	$\frac{e^{t\mu}}{1-b^2t^2}$
Cauchy $N(\mu\ -\ \theta)$	Do no exist

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 57

Explanation of the fact that in general the expected value of a product is not equal to the product of the expected values

Suppose that we have two random variables X and Y.

Then, in general

$$E(XY) \neq E(X)E(Y).$$

Reason:

X times Y is a non-linear function

and the expectation operator doesn't go inside non-linear functions,

it only goes inside linear functions such as $X + Y$.

Example:

Let us divide, at random, a horizontal line segment of length 5 into two parts.

If X is the length of the left-hand part, then $5 - X$ will be the length of the right-hand part.

Since we are diving the line segment into two parts at random, therefore it is reasonable to assume that the length X is uniformly distributed i.e.

$$f(x) = \begin{cases} \frac{1}{5}, 0 < x < 5 \\ 0 \text{ elsewhere} \end{cases}$$

Now, the expected length of the left-hand part is given by

$$E(X) = \int_0^5 xf(x)dx = \int_0^5 x\left(\frac{1}{5}\right)dx$$

$$= \int_0^5 \left(\frac{x}{5}\right)dx = \frac{x^2}{10}\Big|_0^5 = \frac{25}{10} - \frac{0}{10} = \frac{25}{10} = \frac{5}{2}$$

Similarly, the expected length of the right-hand part is given by

$$E(5-X)=\int_0^5 (5-x)\, f(x)dx=\int_0^5 (5-x)\left(\frac{1}{5}\right)dx$$

$$=\int_0^5 \left(\frac{5-x}{5}\right)dx=\int_0^5 \left(1-\frac{x}{5}\right)dx$$

$$=\left[x-\frac{x^2}{10}\right]_0^5=5-\frac{25}{10}=5-\frac{5}{2}=\frac{5}{2}$$

vulearninghub.xyz | by VU Learning HUB

On the other hand, the expected length of the product is calculated as follows:

$$E\left[X\left(5-X\right)\right]=E\left(5X-X^2\right)=\int_0^5\left(5x-x^2\right)f\left(x\right)dx$$

$$=\int_0^5\left(5x-x^2\right)\left(\frac{1}{5}\right)dx=\int_0^5\left(\frac{5x-x^2}{5}\right)dx=\int_0^5\left(x-\frac{x^2}{5}\right)dx$$

$$=\left[\frac{x^2}{2}-\frac{x^3}{15}\right]_0^5=\frac{25}{2}-\frac{125}{15}-0+0=\frac{375-250}{30}=\frac{125}{30}=\frac{25}{6}$$

Now, $E(X)E(5-X)=\frac{5}{2}\times\frac{5}{2}=\frac{25}{4}$

whereas

$$E\left[X(5-X)\right]=\frac{25}{6}$$

so that

$$E\left[X(5-X)\right]\neq E(X)E(5-X).$$

Hence, it is easy to see that, in general, the expected value of a product is not equal to the product of the expected values.

However, we have the following important result:

If X and Y are independent random variables, then

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

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 58

Yet another example of computing Mathematical Expectation of a Function of a Continuous Random Variable X

Example:

Let X be a non-negative random variable of the continuous type with pdf $f(x), x > 0.$

Show that

$$E(X) = \int_0^\infty \left(1 - F(x)\right) dx$$

where $F(x)$ is the cdf of X.

vulearninghub.xyz | Powered by VU Learning HUB

Solution:

$$1-F(x)=1-P(X\le x)=P(X>x)=\int_{x}^{\infty} f_X(y)\,dy$$

So

$$\int_{0}^{\infty}\left(1-F(x)\right)dx=\int_{0}^{\infty}\int_{x}^{\infty} f_X(y)\,dy\,dx$$

Now,

$$1-F(x)=1-P(X\le x)=P(X>x)=\int_{x}^{\infty} f_X(y)\,dy$$

$$\text{So } \int_{0}^{\infty}\left(1-F(x)\right)dx=\int_{0}^{\infty}\int_{x}^{\infty} f_X(y)\,dy\,dx$$

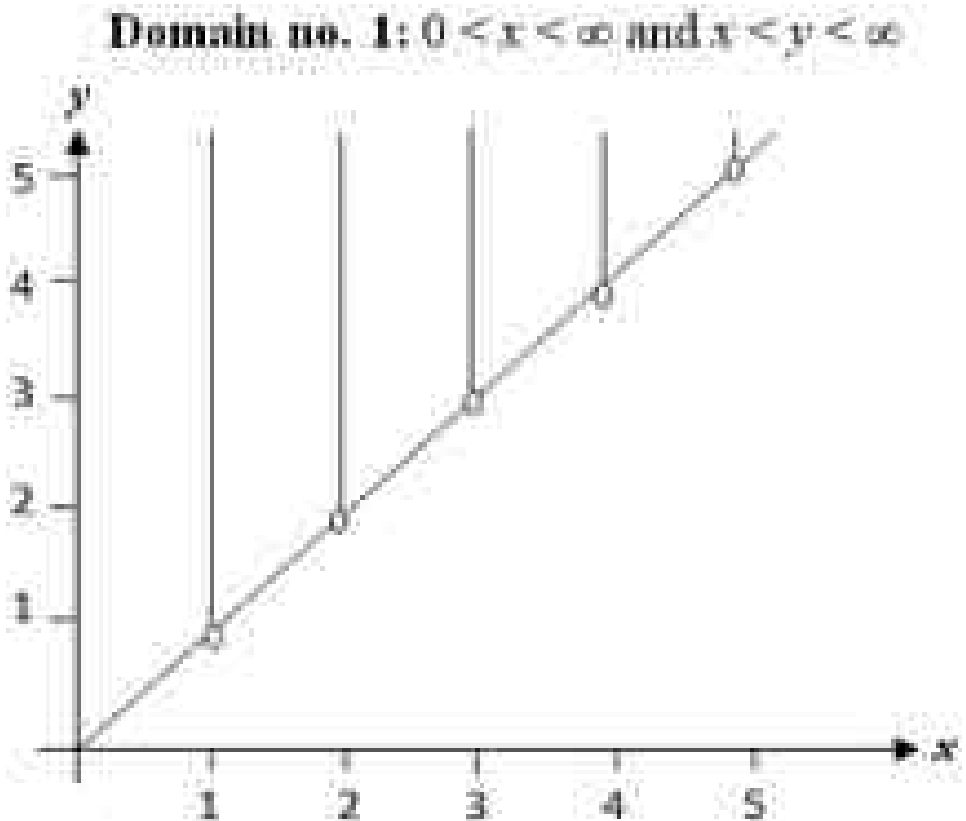
Note that in the above, we are integrating over the domain

$0<x<\infty$ and $x<y<\infty$.

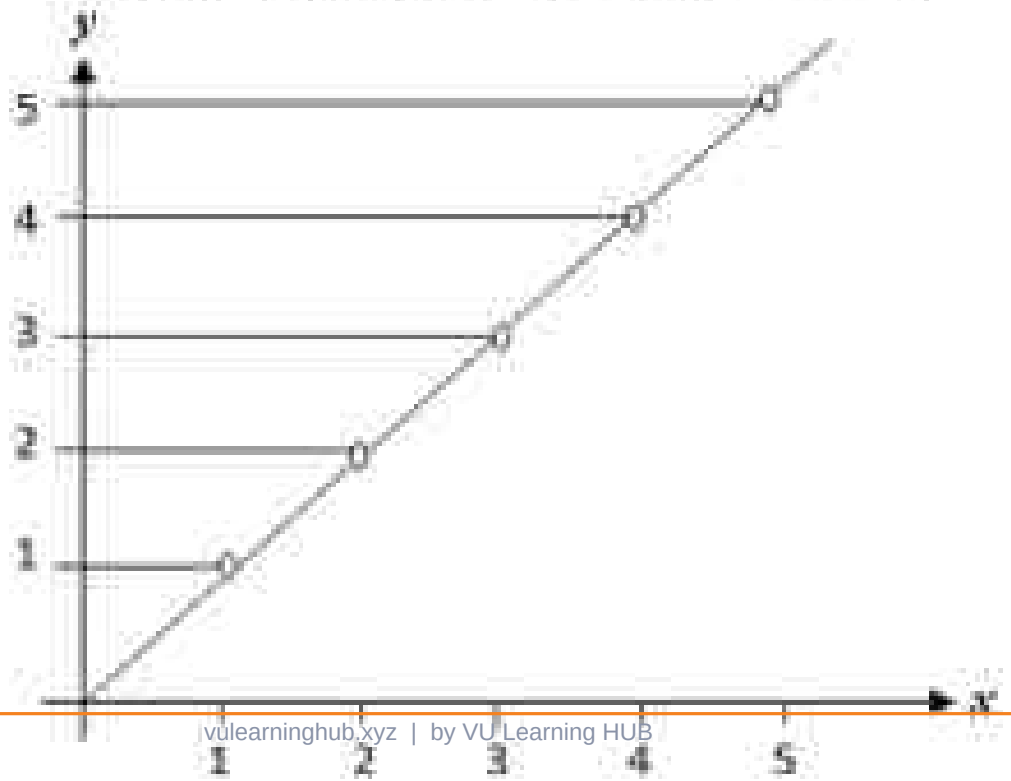
Now, integrating over the domain $0 < x < \infty$ and $x < y < \infty$

is the same as integrating over the domain $0 < y < \infty$ and $0 < x < y$.

Explanation:



Domain no. 2: $0 < y < \infty$ and $0 < x < y$



vulearninghub.xyz | by VU Learning HUB

So, integrating over the domain $0 < x < \infty$ and $x < y < \infty$

is the same as integrating over the domain $0 < y < \infty$ and $0 < x < y$.

$$\text{So } \int_0^\infty \left(1 - F(x)\right) dx = \int_0^\infty \int_x^\infty f_X(y)\, dy dx$$

$$= \int_0^\infty \int_0^y f_X(y) dx dy = \int_0^\infty \left(\int_0^y f_X(y) dx \right) dy$$

Now,

$$\int_0^y f_X(y)\,dx = f_X(y)\int_0^y 1.dx = f_X(y)\left[x\right]_0^y = f_X(y)(y-0) = yf_X(y)$$

Therefore

$$\int_0^\infty \left(1-F(x)\right)dx = \int_0^\infty \left(yf_X(y)\right)dy = \int_0^\infty yf_X(y)\,dy = E(X).$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 59

Derivation

of

Mean and Variance of a distribution

in terms of its MGF

(by repeated differentiation of the mgf)

- Since a distribution that has an mgf M(t) is **completely determined** by M(t), it would not be surprising if we could obtain some **properties** of this distribution **directly** from M(t).

- For example, the existence of M(t) for $-h < t < h$ implies that **derivatives** of M(t) **of all orders exist** at $t = 0$.

So here we will apply the method of **successive derivatives** of the MGF in order to prove that , for any pdf the MGF of which exists,

$$\mu = M'(0)$$

and

$$\sigma^2 = M''(0) - \left[M'(0)\right]^2$$

If X is a continuous random variable, then

$$M'(t) = \frac{dM(t)}{dt} = \frac{d}{dt}\int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_{-\infty}^{\infty}\left(\frac{d}{dt}e^{tx}\right) f(x)\,dx = \int_{-\infty}^{\infty} xe^{tx} f(x)\,dx$$

Now, $\frac{d}{dt}e^{mx} = me^{mx}$

But here x is acting as the "constant" and t is the variable.

$$\therefore \frac{d}{dt}e^{tx} = \frac{d}{dt}e^{xt} = xe^{xt} = xe^{tx}$$

Hence, $M'(t) = \frac{dM(t)}{dt} = \int_{-\infty}^{\infty} xe^{tx} f(x)\,dx$

Upon setting $t = 0$, we have

$$M'(0) = \left| \frac{dM(t)}{dt} \right|_{t=0} = \int_{-\infty}^{\infty} xe^{0x} f(x)\,dx = \int_{-\infty}^{\infty} xe^{0} f(x)\,dx$$

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

Now, the second derivative:

$$M''(t) = \frac{dM'(t)}{dt} = \frac{d}{dt}\int_{-\infty}^{\infty} xe^{tx} f(x)\,dx = \int_{-\infty}^{\infty}\left(\frac{d}{dt}xe^{tx}\right) f(x)\,dx = \int_{-\infty}^{\infty} x f(x)\left(\frac{d}{dt}e^{tx}\right)dx$$

Now, $\frac{d}{dt}e^{mx} = me^{mx}$

But here x is acting as the "constant" and t is the variable

Therefore $\frac{d}{dt}e^{tx} = xe^{tx}$

& hence $M''(t) = \int_{-\infty}^{\infty} x f(x) xe^{tx} dx = \int_{-\infty}^{\infty} x^2 f(x) e^{tx} dx$

Upon setting $t = 0$, we have

$$M''(0) = \int_{-\infty}^{\infty} x^2 f(x) e^{0x} dx = \int_{-\infty}^{\infty} x^2 f(x) e^{0} dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx = E\left(X^2\right)$$

According to the short-cut formula,

$$\sigma^2 = \text{Var}(X) = E\left(X^2\right) - \mu^2$$

But since $M'(0) = \mu$ and $M''(0) = E\left(X^2\right)$ (as proved above)

Therefore

$$\sigma^2 = M''(0) - \left[M'(0)\right]^2$$

The point to be remembered is that sometimes one way is **easier** than the other.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 60

Derivation

of

m^{th} moment about the Origin of a distribution from its MGF (by repeated differentiation of the MGF)

Theorem

If m is a positive integer and if $M^{(m)}(t)$ means the mth derivative of $M(t)$, we have, by repeated differentiation with respect to t,

$$M^{(m)}(0) = E\left(X^m\right).$$

and

$$E\left(X^m\right) = \int_{-\infty}^{\infty} x^m f(x)\,dx$$

where *or*

$$\sum_x x^m p(0),$$

vulearninghub.xyz | by VU Learning HUB

Proof:

$$M'(t) = \int_{-\infty}^{\infty} xe^{tx} f(x)\,dx,$$

$$M''(t) = \int_{-\infty}^{\infty} x^2 e^{tx} f(x)\,dx,$$

$$M'''(t) = \int_{-\infty}^{\infty} x^3 e^{tx} f(x)\,dx\ ,$$

$$M^{(m)}(t) = \int_{-\infty}^{\infty} x^m e^{tx} f(x)\,dx$$

$$\Rightarrow M^{(m)}(0) = \int_{-\infty}^{\infty} x^m e^{0x} f(x)\,dx$$

$$= \int_{-\infty}^{\infty} x^m f(x)\,dx = E\left(X^m\right) = m^{th}\,moment\,about$$

$$about\,the\,origin.$$

- Since $M(t)$ generates the values of $E(X^m)$, $m = 1, 2, 3, \ldots$, it is called the moment generating function (mgf).

- m^{th} moment about the Origin of a distribution from its MGF is the same as the mth derivative evaluated at $t = 0$.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 61

Derivation

of

m^{th} moment about an Arbitrary origin of a distribution

from its MGF

(by repeated differentiation of the mgf)

Theorem

Let X be a random variable such that

$$R(t) = E(e^{t(X-a)})$$

exists for t such that $-h < t < h$.

If m is a positive integer, $R^{(m)}(0)$ is equal to the mth moment of the distribution about the point a.

Vilearningbub.xyz | by VU Learning HUB

Proof:

$$R(t) = E\left[e^{t(x-a)}\right] = \int_{-\infty}^{\infty} e^{t(x-a)} f(x)\,dx, = \int_{-\infty}^{\infty} e^{t(x-a)} f(x)\,dx$$

$$R'(t) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{t(x-a)} f(x)\,dx, \quad = \int_{-\infty}^{\infty} \left(\frac{d}{dt} e^{t(x-a)}\right) f(x)\,dx$$

Now,

$$\frac{d}{dt}e^{t(x-a)} = (x-a)e^{t(x-a)}$$

$$\therefore R'(t) = \int_{-\infty}^{\infty} (x-a)e^{t(x-a)} f(x)\,dx$$

$$\Rightarrow R'(0) = \int_{-\infty}^{\infty} (x-a)e^{0(x-a)} f(x)\,dx = \int_{-\infty}^{\infty} (x-a)e^{0} f(x)\,dx$$

or

$$R'(0) = \int_{-\infty}^{\infty} (x-a)\, f(x)\, dx$$

$$= E(x-a) = 1^{st}\ moment\ about\ a$$

or

$$E(x-a)^1 == 1^{st}\ moment\ about\ a.$$

Next,

$$R''(0) = \frac{d}{dt} R'(t) = \frac{d}{dt}\left[\int_{-\infty}^{\infty} (x-a) e^{t(x-a)} f(x) \right]$$

$$= \int_{-\infty}^{\infty} (x-a) f(x) \left(\frac{d}{dt} e^{t(x-a)} \right) dx = \int_{-\infty}^{\infty} (x-a)(x-a) e^{t(x-a)} f(x) dx$$

$$= \int_{-\infty}^{\infty} (x-a)^2 e^{t(x-a)} f(x) dx$$

$$\Rightarrow R''(0) = \int_{-\infty}^{\infty} (x-a)^2 e^{0(x-a)} f(x) dx = \int_{-\infty}^{\infty} (x-a)^2 f(x) dx$$

$$= E(X-a)^2 = 2^{nd} \text{ moment about } a.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 62

Alternative method for finding the moments from the MGF using Maclaurin's series expansion of the mgf (through an Example)

We can differentiate M(t) any number of times to find the moments of X.

However, it is instructive to consider the following **alternative** method --- one that involves the Maclaurin Series expansion of M(t).

Maclaurin Series

By definition, for any function *f(x)*, a Maclaurin's series is given by:

$$f(x) = f(0) + \frac{f'(0)}{1!}x + \frac{f''(0)}{2!}x^2 + \frac{f^{(3)}(0)}{3!}x^3 + \ldots + \frac{f^{(n)}(0)}{n!}x^n + \ldots$$

If our function is M(t), the Maclaurin's series will be given by

$$M(t) = M(0) + \frac{M'(0)}{1!}t + \frac{M''(0)}{2!}t^2 + \ldots + \frac{M^{(m)}(0)}{m!}t^m + \ldots$$

But we know that

$$M^{(m)}(0) = E\left(X^m\right)$$

Also, we know that

$$M(t) = E(e^{tX}) \Rightarrow M(0) = E(e^{0X})$$

$$\Rightarrow M(0) = E(e^0) \Rightarrow M(0) = E(1) = 1$$

vulearninghub.xyz | by VU Learning HUB

Therefore

$$M(t) = 1 + \frac{E(X)}{1!}t + \frac{E(X^2)}{2!}t^2 + \ldots + \frac{E(X^m)}{m!}t^m + \ldots$$

or

$$M(t) = 1 + E(X)\frac{t}{1!} + E(X^2)\frac{t^2}{2!}t + \ldots + E(X^m)\frac{t^m}{m!} + \ldots$$

Thus the **coefficient** of $(t^m/m!)$ in the Maclaurin's series representation of M(t) is $E(X^m)$ i.e. the ***m*th moment about zero.**

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 63

Derivation of the

Relationship

between the MGF of the Standardized Variable

and

the MGF of the Original Random Variable

Theorem:

Let the random variable X have mean μ, standard deviation σ, and mgf $M_X(t), -h < t < h$

Then $$M_Z(t) = e^{-\frac{\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right), \quad -h\sigma < t < h\sigma.$$

where $$Z = \frac{X - \mu}{\sigma}.$$

vulearninghub.xyz | by VU Learning HUB

Proof:

The first point to be noted is that

$$\text{if } -h\sigma < t < h\sigma, \text{ then } -h < \frac{t}{\sigma} < h,$$

which shows that $\frac{t}{\sigma}$ is <u>in the domain</u> of $M_X(t)$.

Now, using the definition of $M_X(t) = E[e^{tX}]$ and the linear properties of the expected value, we start from the RHS of the equation:

$$RHS = e^{-\frac{\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right) = e^{-\frac{\mu t}{\sigma}} E\left[e^{\frac{t}{\sigma}X}\right] = E\left[e^{\frac{tX}{\sigma}-\frac{\mu t}{\sigma}}\right]$$

$$= E\left[e^{t\frac{(X-\mu)}{\sigma}}\right] = E\left[e^{tZ}\right] = M_Z(t) = LHS$$

which verifies the equation.

Next:

Three general results pertaining to the moment generating function:

Three general results pertaining to the moment generating function:

1. $M_{X+a}(t) = E\left[e^{(X+a)t}\right] = e^{at}.M_X(t);$
2. $M_{bX}(t) = E\left[e^{tbX}\right] = E\left[e^{(bt)X}\right] = M_X(bt);$
3. $M_{\frac{X+a}{b}}(t) = E\left[e^{t\left(\frac{X+a}{b}\right)}\right] = E\left[e^{t\left(\frac{a}{b}\right)+t\left(\frac{X}{b}\right)}\right] = e^{t\left(\frac{a}{b}\right)}E\left[e^{\left(\frac{t}{b}\right)X}\right] = e^{\frac{a}{b}t}.M_X\left(\frac{t}{b}\right).$

vulearninghub.xyz | by VU Learning HUB

Note that the third result i.e.

$$M_{\frac{X+a}{b}}(t) = e^{\frac{a}{b}t} M_X\left(\frac{t}{b}\right)$$

is of special importance when $a = -\mu$ and $b = \sigma$,

in which case , we have

$$M_{\frac{X-\mu}{\sigma}}(t) = e^{\frac{-\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right)$$

or, in other words,

$$M_Z(t) = e^{\frac{-\mu t}{\sigma}} M_X\left(\frac{t}{\sigma}\right)$$

exactly the same as what we proved.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 64

Proof of the fact that for a PDF that is Symmetric about 0 and for which the MGF exists, $M(-t)=M(t)$

Theorem:

Let X be a random variable with a pdf $f(x)$ and mgf $M(t)$.

Suppose f is symmetric about 0 i.e. $f(-x) = f(x)$.

Then $M(-t)=M(t)$.

Proof:

By definition

$$M(-t) = E\left(e^{(-t)x}\right) = \int_{-\infty}^{\infty} e^{(-t)x} f(x)\,dx$$

$$\because \begin{pmatrix} \text{if } f(x) \text{ is symmetric about 0,} \\ \text{then } f(x) \text{ itself is } f(\text{-}x) \end{pmatrix}$$

$$= \int_{-\infty}^{\infty} e^{t(-x)} f(x)\,dx = \int_{-\infty}^{\infty} e^{t(-x)} f(-x)\,dx$$

Applying the transformation

$$u = -x \Rightarrow du = \quad -dx$$

&

$$x \to -\infty \Rightarrow u \to \infty,\ x \to \infty \Rightarrow u \to -\infty$$

So

$$M(-t) = \int_{-\infty}^{\infty} e^{t(-x)} f(-x)\,dx = -\int_{\infty}^{-\infty} e^{tu} f(u)\,du = \int_{-\infty}^{\infty} e^{tu} f(u)\,du$$

$$= \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = E\left(e^{tX}\right) = M(t).$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 65

Cumulant Generating Function (CGF) and its role in finding the mean and variance of a probability distribution

We begin with the Concept of

Cumulants

The First Three Cumulants

- The **first cumulant** is the **first moment about zero** i.e. the **mean** of the distribution.

- The **second cumulant** is the **second central moment** i.e. the **variance.**

- The **third cumulant** is the **third central moment.**

Higher Order Cumulants

The higher cumulants are neither moments about zero nor central moments, but rather **more complicated polynomial functions of moments**.

In fact, it can be shown that the kth cumulant $\kappa_n(X)$ of a random variable X is the value of **a certain polynomial in the first k moments of *X*** i.e. a polynomial in $E(X^{\ell})$, $\ell=1,\ldots,k$

Next,
we consider
the concept of the

Cumulant Generating Function

Definition of Cumulant Generating Function:

The natural logarithm of the moment generating function is called the cumulant generating function.

It is denoted by $\Psi(t)$.

In other words:

$$\Psi(t) = \log M(t)$$

vulearninghub.xyz by VU Learning HUB

Next,
let us consider
the Role of the Cumulant Generating Function in finding the mean and variance of a distribution.

Two simple results:

$$\Psi'(0) = \mu$$

and

$$\Psi''(0) = \sigma^2$$

Proof:

i) Mean:

$$\Psi'(t) = \frac{d}{dt}\left(\Psi(t)\right) = \frac{d}{dt}\left(\log M(t)\right) = \frac{1}{M(t)}\left(M'(t)\right) = \frac{M'(t)}{M(t)}$$

$$\text{Hence } \Psi'(0) = \left|\frac{M'(t)}{M(t)}\right|_{t=0} = \frac{M'(0)}{M(0)} = \frac{\mu}{M(0)}$$

However

$$\because M(t) = E\left(e^{tX}\right) \therefore M(0) = E\left(e^{0X}\right) = E\left(e^{0}\right) = E(1) = 1$$

So that $\Psi'(0) = \frac{\mu}{1} = \mu.$ Hence proved.

Proof:

ii) Variance:

We have $\Psi'(t) = \dfrac{M'(t)}{M(t)}$

Hence $\Psi''(t) = \dfrac{d}{dt}\left(\dfrac{M'(t)}{M(t)}\right) = \dfrac{M(t)M''(t) - \left(M'(t)\right)^2}{\left(M(t)\right)^2}$

so that $\Psi''(0) = \dfrac{M(0)M''(0) - \left(M'(0)\right)^2}{\left(M(0)\right)^2} = \dfrac{(1)M''(0) - \left(\mu\right)^2}{\left(1\right)^2}$

$= M''(0) - \left(\mu\right)^2$

However, we know that $M''(0) = E\left(X^2\right)$

$\therefore \Psi''(0) = E\left(X^2\right) - \left(\mu\right)^2 = \text{Var}(X) = \sigma^2$. Hence proved.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 66

This is known as the

Additivity Property

of the CGF and the Cumulants

The CGF of the sum
of two independent random variables
equals the sum of their CGFs

so that

each cumulant of a sum
of independent random variables
is the sum of
the corresponding cumulants of the addends

1. Additivity Property of the CGF:

If X and Y are independent random variables,

then,

the cumulant generating function of $X+Y$ is related to the cumulant generating function of X and the cumulant generating function of Y by the relation

$$K_{X+Y}(t) = K_X(t) + K_Y(t).$$

Proof:

If X and Y are independent random variables, then

$$\begin{aligned} K_{X+Y}(t) &= \log \mathrm{E}\left[e^{t(X+Y)}\right] \\ &= \log \mathrm{E}\left[e^{t(X)}e^{t(Y)}\right] \\ &= \log\left(\mathrm{E}\left[e^{tX}\right]\mathrm{E}\left[e^{tY}\right]\right) \\ &= \log \mathrm{E}\left[e^{tX}\right] + \log \mathrm{E}\left[e^{tY}\right] \\ &= K_X(t) + K_Y(t). \end{aligned}$$

2. Additivity Property of the Cumulants:

07:05

If X and Y are independent random variables,

then,

the nth cumulant of $X+Y$ is related to the nth cumulant of X and the nth cumulant of Y by the relation

$$\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y).$$

The proof of this result is somewhat advanced.

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 67

Relation between Central MGF and CGF

AND

Relation between Moments about the Mean and Cumulants

- The central moment generating function is given denoted by $C(t)$

- **Definition**

It is shown as the

$$C(t) = E\left(e^{t(x-\mu)}\right)$$

When we consider the mgf for moments about ‘0’

also

we say $E\left(e^{tx}\right)$ can also be written as $E\left(e^{t(x-0)}\right)$

The central moment generating function is given

$$C(t) = E\left[e^{t(x-\mu)}\right] = E\left[e^{tx-\mu t}\right] = E\left[e^{tx}e^{-\mu t}\right]$$

$$= e^{-\mu t}E\left[e^{tx}\right] = e^{-\mu t}M(t)$$

Now, by definition, we have

$$K(t) = \log M(t)$$

$$\therefore e^{K(t)} = e^{\log M(t)} \Rightarrow e^{K(t)} = M(t)$$

$$\therefore C(t) = e^{K(t)} e^{-\mu t} = e^{\left(K(t) - \mu t\right)}.$$

- To express the *central* moments as functions of the cumulants, just drop from these polynomials all terms in which κ_1 appears as a factor:

$$\mu_1 = 0$$

$$\mu_2 = \kappa_2$$

$$\mu_3 = \kappa_3$$

$$\mu_4 = \kappa_4 + 3\kappa_2^2$$

$$\mu_5 = \kappa_5 + 10\kappa_3\kappa_2$$

$$\mu_6 = \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3.$$

To express the cumulants κ_n for $n > 1$ as functions of the central moments, drop from these polynomials all terms in which μ'_1 appears as a factor:

$$\kappa_2 = \mu_2$$

$$\kappa_3 = \mu_3$$

$$\kappa_4 = \mu_4 - 3\mu_2^2$$

$$\kappa_5 = \mu_5 - 10\mu_3\mu_2$$

$$\kappa_6 = \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 68

Chevbyshev's Inequality

and

its proof

&

an alternative form of the inequality

- **Chebyshev's inequality** guarantees that, for a wide class of probability distributions, **no more than a certain fraction of values can be more than a certain distance from the mean.**

According to the Chebychev Theorem, no more than $1/k^2$ of the distribution's values can be more than k standard deviations away from the mean

where k is any positive real number

or equivalently,

at least $1-1/k^2$ of the distribution's **values lie within** k **standard deviations** of the mean.

When a random variable is not normally distributed, then we can make use of **Chebyshev's inequality** in order to find out the **minimum** amount of data that is **within** **k** **standard deviations** of the mean --- in percentage form.

In reference to normal distribution, we know that

68.26% are under the curve lies between $\mu \pm \sigma$
95.45% are under the curve lies between $\mu \pm 2\sigma$
and
99.73% are under the curve lies between $\mu \pm 3\sigma$

If our data is **not** normally distributed, then either we will go to the exact calculations through some amount of labor or we can make use of **Chevbyshev's Inequality.**

Theorem: (Chevbyshev's Inequality)

Let the random variable X have a distribution probability about which we assume only that there is a finite variance σ^2. Then for every $k > 0$,

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

or, equivalently

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

Note

If the variance of a distribution exists, then its mean necessarily exist.

Short-cut formula

Again see the statement as

Let the random variable X have a distribution probability about which we assume only that there is a finite variance σ^2.

Example:

Let k= 2

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

$$\Rightarrow P\left(|X-\mu|<2\sigma\right)\geq 1-\frac{1}{4}\Rightarrow P\left(|X-\mu|<2\sigma\right)\geq \frac{4-1}{4}$$

$$\Rightarrow P\left(|X-\mu|<2\sigma\right)\geq \frac{3}{4}$$

Area that is within μ-2σ and μ+2σ, this area is greater than or equal to ¾ or 75%

Vulearninghub.xyz // by VU Learning HUB

Hence, the number $1/k^2$ or Chebychev inequality give us an **upper bound** for the **probability** $P(|X - \mu| \geq k\sigma)$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 69

Application of Chebyshev's Inequality

Example:

Let X have the pdf

$$f(x)=\begin{cases}\dfrac{1}{2\sqrt{3}} & -\sqrt{3}<x<\sqrt{3}\\ 0 & \textit{elsewhere.}\end{cases}$$

Here $\mu=0$ and $\sigma^2=1$.

If $k = 3/2$, we have the exact probability

$$P\left(|X-\mu| \geq k\sigma\right) = P\left(|X-0| \geq \left(\frac{3}{2}\right)(1)\right)$$

$$= P\left(|X| \geq \frac{3}{2}\right)$$

$$= 1 - \int_{-3/2}^{3/2} \frac{1}{2\sqrt{3}}\, dx = 1 - \left(\frac{1}{2\sqrt{3}}(x|_{-3/2}^{3/2}\right)$$

$$P\left(|X| \geq \frac{3}{2}\right) = 1 - \left[\frac{1}{2\sqrt{3}}\left(\frac{3}{2} + \frac{3}{2}\right)\right]$$

$$= 1 - \frac{1}{\sqrt{3}}\left(\frac{3}{2}\right) = 1 - \frac{3^{1}.3^{-\frac{1}{2}}}{2} = 1 - \frac{\sqrt{3}}{2}.$$

- By Chebyshev's inequality, this probability has the upper bound $1/k^2=4/9$.

- Since $1-\sqrt{3}/2= 0.134$, approximately, the exact probability in this case is considerably less than the upper bound 4/9.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 70

Another application

Of

Chebyshev's Inequality

Alternative form

By Chevbyshev's inequality, putting $k\sigma = m$, we have

$$P\left[\left|X-\mu\right| \geq m\right] \leq \frac{\sigma^2}{m^2}, \ for \ all \ m > 0$$

$$or \ P\left[\left|X-\mu\right| \geq m\right] \leq \left(\frac{Var(X)}{m^2}\right)$$

Also, $Var(X) = E(X)^2 - \left[E(X)\right]^2 = 13 - 9 = 4.$

Example:

If X is a random variable such that

$E(X) = 3$ and $E(X^2) = 13$.

Use Chebyshev's inequality to determine a **lower bound** for the probability $P(-2<X<8)$.

Solution

By Chevbyshev's inequality, we have

$$P\left[\left|X-\mu\right| \geq m\right] \leq \frac{Var\left(X\right)}{m^2},\ for\ all\ m>0$$

In order to obtain a lower bound, we flip the inequalities:

$$P\left[\left|X-\mu\right| < m\right] > 1-\left(\frac{Var\left(X\right)}{m^2}\right)$$

$$or\ P\left[-m<\left(X-\mu\right)<m\right]>1-\left(\frac{Var\left(X\right)}{m^2}\right)$$

$$or\ P\left[\mu-m<X<\mu+m\right]>1-\left(\frac{Var\left(X\right)}{m^2}\right)$$

So,

$$Var(X) = E(X)^2 - \left[E(X)\right]^2$$

$$= 13 - 9 = 4.$$

Now, we note that if we put $k = 5$,

then $3 - k = -2 \quad and \quad 3 + k = 8$,

Therefore, the required probability

$$P\left[-2 < X < 8\right] = P\left[3 - k < X < k + 3\right]$$

But, according to Chebychev's Inequality,

$$P\left[3-m<X<m+3\right]>1-\frac{4}{m^2}$$

So

$$P\left[-2<X<8\right]>1-4/25 \; or \; P\left[-2<X<8\right]>\left(25-4\right)/25$$

$$or \; P\left[-2<X<8\right]>21/25 \text{ or } P\left[-2<X<8\right]>0.84$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 71

H.M. ≤ G.M. ≤ A.M.

Example: (Harmonic and Geometric Means)

Let $\{a_1,\ldots,a_n\}$ be a set of positive numbers.

Create a distribution for a random variable X by placing weight $1/n$ on each of the numbers $a_1,\ldots,a_n$.

Then the mean of X is the arithmetic mean, (AM),

$$E(X) = n^{-1} \sum_{i=1}^{n} a_i .$$

Then the mean of X is the arithmetic mean, (AM),

$$E(X) = \sum xp(x) = \frac{a_1}{n} + \frac{a_2}{n} + \ldots + \frac{a_n}{n}$$

$$= \frac{1}{n}\left(a_1 + a_2 + \ldots + a_n\right) = \frac{1}{n}\sum a_i$$

Theorem (Jensen's Inequality)

If φ is convex on an open interval I and X is a random variable whose support I contained in I and has finite expectation, then

$$\phi\left[E\left(X\right)\right] \leq E\left[\phi\left(X\right)\right].$$

If φ is strictly convex, then the inequality is strict unless X is a constant random variable.

Then, since $-\log x$ is a convex function, we have by Jensen's inequality that

$$-\log\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right) \leq E\left(-\log X\right) = -\frac{1}{n}\sum_{i=1}^{n}\log a_i = -\log\left(a_1 a_2 \ldots a_n\right)^{1/n}$$

or, equivalent,

$$\log\left(\frac{1}{n}\sum_{i=1}^{n} a_i\right) \geq \log\left(a_1 a_2 ... a_n\right)^{1/n},$$

and, hence

$$\left(a_1 a_2 ... a_n\right)^{1/n} \leq \frac{1}{n}\sum_{i=1}^{n} a_i \qquad (1)$$

The quantity on the left side of this inequality is called the geometric mean (G.M).

So, equation 1 is equivalent to saying that GM $\leq$ AM for any finite set of positive numbers.

vulearninghub.xyz | by VU Learning HUB

Now in eq. (1) replace a_i by $1/a_i$ (which is positive).

We then obtain $$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i} \geq \left(\frac{1}{a_1}\frac{1}{a_2}\cdots\frac{1}{a_n}\right)^{1/n}$$

Or, equivalently, $$\frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{a_i}} \leq \left(a_1 a_2 \cdots a_n\right)^{1/n} \qquad (2)$$

The left member of this inequality is called harmonic mean HM.

Putting the equations together, we have shown the relationship

$$HM \leq GM \leq AM, \qquad (3)$$

for any finite set of positive numbers.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 72

Concept of a Random Vector (explained through an example)

Let us begin the discussion of a **pair of random variables** with the following example.

A coin is tossed three times and our interests is in the ordered number pair (number of H's on first two tosses, number of H's on all three tosses),

where H and T represent respectively, heads and tails.

Let C={TTT,TTH,THT,HTT,THH, HTH,HHT,HHH} denote the sample space.

Let X_1 denote the number of H's on the first two tosses and X_2 denotes the number of HH's on all three flips.

Then our interest can be represented by the pair of random variables (X_1, X_2).

For example, (X_1(HTH), X_2(HTH)) represents the outcome (1,2).

Continuing in this way, X_1 and X_2 are real-valued functions defined on the sample space C, which take us from the **sample space** to the space of ordered number pairs.

$$D = \{(0,0), (0,1),(1,1), (1,2),(2,2),(2,3)\}.$$

Thus X_1 and X_2 are two random variables defined on the space C,

and,

(in this example), the space of these random variables is the **two-dimensional** set $\mathcal{D}$, which is a subset of the two-dimensional Euclidean Space R^2.

Hence (X_1, X_2) is a vector function from C to $\mathcal{D}$.

In simple words, the **pair** of random variables (X_1, X_2) is a **random vector.**

So

Formal Definition of a Random Vector in the case of Two random variables:

Definition: (Random Vector)

From a random experiment with a sample space C, consider two random variables X_1 and X_2, which assign to each element c of C one and only one ordered pair of numbers

$$X_1(c) = x_1, \quad X_2(c) = x_2.$$

Then we say that (X_1, X_2) is a random vector.

vulearninghub.xyz | by VU Learning HUB

The space of (X_1,X_2) is the set of ordered pairs

$$\mathcal{D} = \{(x_1, x_2) : x_1 = X_1(c), x_2 = X_2(c), c \in C\}$$

We often denote random vector notation

$$X = \left(X_1, X_2\right)',$$

where the ` denotes the transpose of the row vector (X_1,X_2).

vulearnhub.xyz | by VU Learning HUB

Conversely,

If (X_1, X_2) is random vector then both X_1 and X_2 are random variables.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 73

Concept of an event in a case of a two-dimensional space (i.e. a set of ordered pairs)

- Let $\mathcal{D}$ be the space associated with the random vector (X_1, X_2).

- Let A be a subset of $\mathcal{D}$. As in the case of one random variable, we speak of the event A.

- We wish to define the probability of the event A, which we denote by $P_{X_1, X_2}[A]$

- A coin is tossed three times and our interests is in the ordered number pair (number of H's on first two tosses, number of H's on all three tosses), where H and T represent respectively, heads and tails.

Let C={TTT,TTH,THT,HTT,THH, HTH,HHT,HHH} denote the sample space.

Let X_1 denote the number of H's on the first two tosses and X_2 denotes the number of HH's on all three flips.

Then our interest can be represented by the pair of random variables (X_1, X_2).

For example, (X_1(HTH), X_2(HTH)) represents the outcome (1,2).

Continuing in this way, X_1 and X_2 are real-valued functions defined on the sample space C, which take us from the sample space to the space of ordered number pairs.

$$\mathcal{D} = \{(0,0), (0,1),(1,1), (1,2),(2,2),(2,3)\}.$$

Thus X_1 and X_2 are two random variables defined on the space C, and, in this example, the space of these random variables is the two-dimensional set D, which is a subset of two-dimensional Euclidean R^2.

Hence (X_1, X_2) is a vector function from C to D.

Now we can define various events:

e.g.

‘Number of Heads on the first two tosses greater than zero and number of Heads on all three tosses less than 2’

Then Event $A = \{(1,1)\}$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 74

Joint Cumulative Distribution Function

- Given random variables X_1, X_2, ..., that are defined on a probability space, the **joint probability distribution** for X_1, X_2, ... is a probability distribution that gives the probability that **each** of X_1, X_2, ... falls in any particular range or discrete set of values specified for that variable.

- In the case of only two random variables, this is called a **bivariate distribution**, but the concept generalizes to any number of random variables, giving a **multivariate distribution**.

- The joint probability distribution can be expressed either in terms of a **joint cumulative distribution function** or in terms of a joint probability density function (in the case of continuous variables) or joint probability mass function (in the case of discrete variables).

So, let us focus on

the Joint

Cumulative

Distribution Function

We can uniquely define $P_{X1,X2}$ in terms of the **cumulative distribution (cdf)**, which is given by

$$F_{X_1,X_2}(x_1,x_2) = P\left[\{X_1 \le x_1\} \cap \{X_2 \le x_2\}\right], \qquad (1)$$

for all (x_1, x_2)òR^2

because

$X_1 \le x_1$ and $X_2 \le x_2$ are events with reference to the random variables X_1 and X_2 separately --- AND the intersection of these two events is the 'joint event'.

Thus the expression is well defined.

As with random variables, we write

$$P\left[\left\{X_1 \leq x_1\right\} \cap \left\{X_2 \leq x_2\right\}\right] \text{ as } P\left[X_1 \leq x_1, X_2 \leq x_2\right]$$

Also, we can write,

$$P\left[a_1 < X_1 \le b_1, a_2 < X_2 \le b_2\right] = F_{X_1,X_2}\left(b_1,b_2\right) - F_{X_1,X_2}\left(a_1,b_2\right) - F_{X_1,X_2}\left(b_1,a_2\right) + F_{X_1,X_2}\left(a_1,a_2\right) \qquad (2)$$

Hence, all induced probabilities of sets of the form can be formulated in terms of the cdf. $\left(a_1,b_1\right]\times\left(a_2,b_2\right]$

We often call this cdf the **joint cumulative distribution function** of (X_1, X_2).

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 75

Discrete Random Vector and Joint Probability Mass Function

Definition: A random vector (X_1, X_2) is a discrete random vector if its space $\mathcal{D}$ is finite or countable.

Hence, X_1 and X_2 are both discrete also.

The **joint probability mass function (pmf)** of (X_1, X_2) is defined by

$$p_{X_1,X_2}(x_1, x_2) = P[X_1 = x_1, X_2 = x_2],$$

for all $(x_1, x_2) \in \mathcal{D}$

Properties

It also is characterized by the two properties

$$(i)\ 0 \le p_{X_1,X_2}(x_1,x_2) \le 1 \quad and$$

$$(ii) \sum_D \sum p_{X_1,X_2}(x_1,x_2) = 1. \qquad (1)$$

For an event $\mathrm{B} \in \mathcal{D}$, we have

$$P\left[\left(X_1, X_2\right) \in \mathrm{B}\right] = \sum \sum_{\mathrm{B}} p_{X_1, X_2}\left(x_1, x_2\right).$$

Likewise we may extend the pmf $p_{X_1,X_2}(x_1,x_2)$ over a convenient set by using zero elsewhere.

Hence, we replace

$$\sum_{D}\sum p_{X_1,X_2}(x_1,x_2) \quad by \quad \sum_{x_2}\sum_{x_1} p(x_1,x_2).$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 76

Concept of the Support of a Discrete Random Vector

(explained through an example)

- We begin with the concept of the Support of a random variable:

- The support of a random variable is the set of values that the random variable can take.

- Concentrating on the Support of a discrete variable:

- For discrete random variables, it is the set of all the realizations that have a strictly **positive** probability of being observed.

The above ideas can be extended to develop the Concept of the Support of a Discrete Random **Vector.**

Let us attempt this with the help of the following example.

A coin is tossed three times and our interest is in the ordered number pair (number of heads on first two tosses, number of heads on all three tosses),

Let H and T represent respectively, heads and tails. Then the sample space C is given by

C={TTT,TTH,THT,HTT,THH, HTH,HHT,HHH}.

Let
X_1 denote the **number of H's** on the **first two** tosses
and
X_2 denotes the **number of H's** on **all three** flips.

Then our interest can be represented by the **pair** of random variables (X_1, X_2).

Now,

$X_1(TTT)=0$ and $X_2(TTT)=0$

$X_1(TTH)=0$ and $X_2(TTH)=1$

$X_1(THT)=1$ and $X_2(THT)=1$

$X_1(HTT)=1$ and $X_2(HTT)=1$

$X_1(THH)=1$ and $X_2(THH)=2$

$X_1(HTH)=1$ and $X_2(HTH)=2$

$X_1(HHT)=2$ and $X_2(HHT)=2$

$X_1(HHH)=2$ and $X_2(HHH)=3$

Now, our interest can be represented by the **pair** of random variables (X_1, X_2).

So, the eight possible pairs are

$(X_1,X_2)=(0,0)$,

$(X_1,X_2)=(0,1)$,

$(X_1,X_2)=(1,1)$,

$(X_1,X_2)=(1,1)$,

$(X_1,X_2)=(1,2)$,

$(X_1,X_2)=(1,2)$,

$(X_1,X_2)=(2,2)$,

$(X_1,X_2)=(2,3)$,

So, we write

$$\mathcal{D} = \{(0,0), (0,1),(1,1), (1,2),(2,2),(2,3)\}.$$

with **probabilities**

$P[(X_1,X_2)=(0,0)]= 1/8$

$P[(X_1,X_2)=(0,1)]= 1/8$

$P[(X_1,X_2)=(1,1)]= 2/8$

$P[(X_1,X_2)=(1,2)]= 2/8$

$P[(X_1,X_2)=(2,2)]= 1/8$

$P[(X_1,X_2)=(2,3)]= 1/8$

We can conveniently table the pmf of the random vector (X_1, X_2) as

		Support of X_2			
		0	**1**	**2**	**3**
Support of X_1	**0**	1/8	1/8	0	0
	1	0	2/8	2/8	0
	2	0	0	1/8	1/8

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 77

Continuous Random Vector
and
Joint Probability Density Function

- A **random vector** (X_1, X_2) with space D is of the continuous type if its cdf $F_{X_1,X_2}(x_1, x_2)$ is continuous .

- That is $F_{X_1,X_2}(x_1, x_2)$ can be expressed as

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(w_1, w_2) dw_1 dw_2$$

for all $(x_1, x_2) \in R^2$.

Joint Probability Density Function

We call the integrand

$$F_{X_1,X_2}\left(x_1,x_2\right)=\int_{-\infty}^{x_1}\int_{-\infty}^{x_2} f_{X_1,X_2}\left(w_1,w_2\right)dw_1dw_2$$

the **joint probability density function (PDF)** of (X_1, X_2).

Then

$$\frac{\partial^2 F_{X_1,X_2}(x_1,x_2)}{\partial x_1 \partial x_2} = f_{X_1,X_2}(x_1,x_2)$$

except possibly on events which have probability zero.

Properties

A pdf is essentially characterized by the two properties

$$(i)\ f_{X_1,X_2}(x_1,x_2) \geq 0 \qquad and$$

$$(ii) \int\int_D f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2 = 1. \qquad (1)$$

We may extend the definition of a pdf

$$f_{X_1,X_2}\left(x_1,x_2\right)\ over\ \ R^2$$

by using zero elsewhere.

- We do this consistently so that tedious, repetitious references to the space $\mathcal{D}$ can be avoided. Once this is done, we replace

$$f_{X_1,X_2}(x_1,x_2) \ over \ R^2$$

$$\iint_D f_{X_1,X_2}(x_1,x_2)dx_1dx_2 \quad by \quad \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)dx_1dx_2.$$

- For an event A ϵ Ɗ, we have

$$P\left[\left(X_1, X_2\right) \in A\right] = \iint_A f_{X_1, X_2}\left(x_1, x_2\right) dx_1 dx_2$$

- Note that the $P\left[\left(X_1, X_2\right) \in A\right]$ is just the volume under the surface $z = f_{X_1, X_2}\left(x_1, x_2\right)$ over the set A.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 78

Determination of the probability of an event in the case of a continuous random vector (explained through an example)

Example: Let

$$f\left(x_1, x_2\right) = \begin{cases} 6x_1^2 x_2 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \textit{elsewhere} \end{cases}$$

be the pdf of the random vector (X_1, X_2) where both X_1 and X_2 are random variables of the continuous type.

Suppose, we wish to compute the probability

$$P\left(0 < X_1 < \frac{3}{4}, \frac{1}{3} < X_2 < 2\right)$$

Solution:

Since (X_1, X_2) is a continuous random vector, therefore

$$P\left(0 < X_1 < \frac{3}{4}, \frac{1}{3} < X_2 < 2\right) = \int_{1/3}^{2}\int_{0}^{3/4} f\left(x_1, x_2\right) dx_1 dx_2$$

$$= \int_{1/3}^{1}\int_{0}^{3/4} 6x_1^2 x_2 \, dx_1 dx_2 + \int_{1}^{2}\int_{0}^{3/4} 0 \, dx_1 dx_2$$

$$= 6x_2 \int_{1/3}^{1}\left(\int_{0}^{3/4} x_1^2 dx_1\right) dx_2 + 0$$

$$P\left(0 < X_1 < \frac{3}{4}, \frac{1}{3} < X_2 < 2\right) = 6x_2 \int_{1/3}^{1} \left(\frac{x_1^{2+1}}{2+1} \Big|_0^{3/4} \right) dx_2 + 0 = 6x_2 \int_{1/3}^{1} \left(\frac{\left(\frac{3}{4}\right)^3}{3} - 0 \right) dx_2 + 0$$

$$= 6x_2 \int_{1/3}^{1} \left(\frac{27}{64 \times 3} \right) dx_2 + 0 = 3x_2 \int_{1/3}^{1} \left(\frac{9}{32} \right) dx_2 + 0 \; = \frac{27}{32} \int_{1/3}^{1} x_2 \, dx_2 + 0$$

or

$$P\left(0 < X_1 < \frac{3}{4}, \frac{1}{3} < X_2 < 2\right) = \frac{27}{32}\int_{1/3}^{1} x_2\, dx_2$$

$$= \frac{27}{32}\left(\frac{x_2^{1+1}}{1+1}\Bigg|_{1/3}^{1} = \frac{27}{32}\left(\frac{x_2^2}{2}\Bigg|_{1/3}^{1} = \frac{27}{32}\left(\frac{(1)^2}{2} - \frac{(1/3)^2}{2}\right)\right.\right.$$

$$= \frac{27}{32}\left(\frac{1}{2} - \frac{1}{9\times 2}\right) = \frac{27}{32}\left(\frac{1}{2} - \frac{1}{18}\right) = \frac{27}{32}\left(\frac{9-1}{18}\right)$$

$$= \frac{27}{32}\left(\frac{8}{18}\right) = \frac{3}{4}\left(\frac{1}{2}\right) = \frac{3}{8}.$$

vulearninghub.xyz | by VU Learning HUB

Note that this probability is the **volume** under the surface given by $f(x_1, x_2) = 6x_1^2 x_2$ above the **rectangular** set

$$\left\{(x_1, x_2) : 0 < x_1 < \frac{3}{4}, \frac{1}{3} < x_2 < 1\right\} \in R^2$$

(i.e. the rectangular area given by

$$\left\{(x_1, x_2) : 0 < x_1 < \frac{3}{4}, \frac{1}{3} < x_2 < 1\right\} \in R^2).$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 79

Concept of the Support of a Continuous Random Vector (explained through an example)

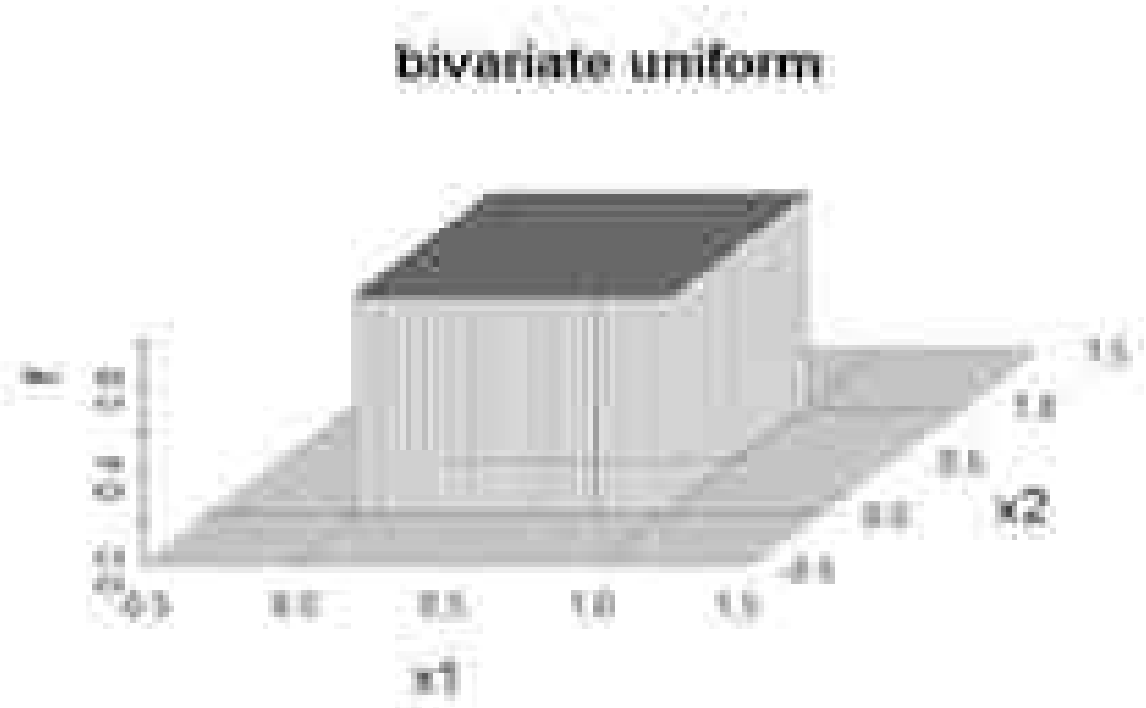
- For a continuous random vector (X_1, X_2), the support of (X_1, X_2) contains all points (x_1, x_2) for which $f(x_1, x_2) > 0$.

Example:

Let us consider the bivariate uniform distribution given by

$$f\left(x_1, x_2\right) = \begin{cases} 1 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \textit{elsewhere} \end{cases}$$

Then:



And

The support of the random vector is

$$\{0 < x_1 < 1,\ 0 < x_2 < 1\}.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 80

Properties
of the
Joint Cumulative Distribution Function

Definition

- The joint cumulative function of two random variables X and Y is defined as

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

Properties:

The joint CDF satisfies the following properties:

1. $F_X(x) = F_{XY}(x, \infty)$, for any x (marginal CDF of X);
2. $F_Y(y) = F_{XY}(\infty, y)$, for any y (marginal CDF of Y);
3. $F_{XY}(\infty, \infty) = 1$;
4. $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$;
5. $P(x_1 < X \le x_2, y_1 < Y \le y_2) =$

$$F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1) + F_{XY}(x_1, y_1);$$

6. If X and Y are independent, then

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

1. $F_X(x) = F_{XY}(x, \infty)$, for any x (marginal CDF of X);

2. $F_Y(y) = F_{XY}(\infty, y)$, for any y (marginal CDF of Y);

Explanation

1. $F_X(x) = F_{XY}(x, \infty)$, for any x (marginal CDF of X);

2. $F_Y(y) = F_{XY}(\infty, y)$, for any y (marginal CDF of Y);

3. $F_{XY}(\infty, \infty) = 1;$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 81

Properties
of the
Joint Cumulative Distribution Function

Definition

- The joint cumulative function of two random variables X and Y is defined as

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

Properties:

The joint CDF satisfies the following properties:

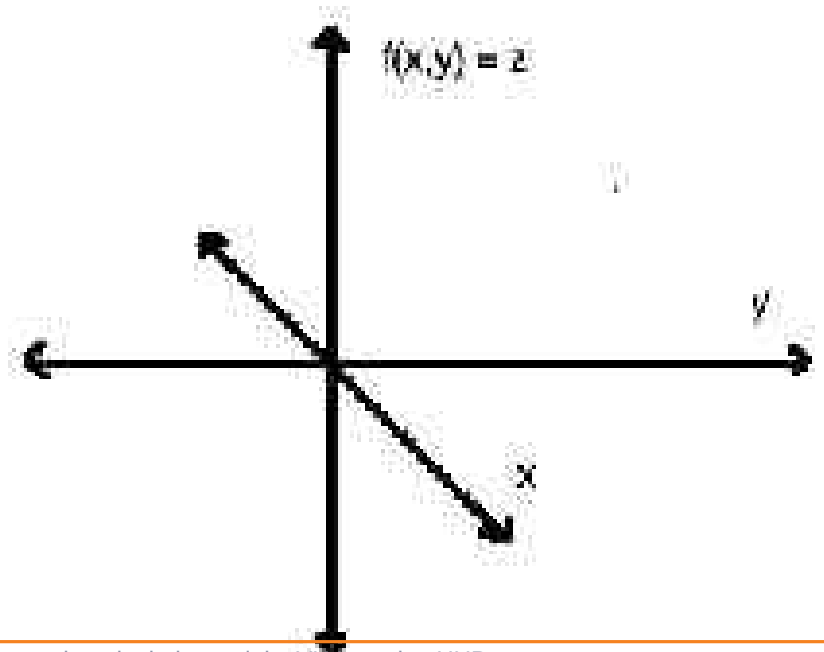
1. $F_X(x) = F_{XY}(x, \infty)$, for any x (marginal CDF of X);
2. $F_Y(y) = F_{XY}(\infty, y)$, for any y (marginal CDF of Y);
3. $F_{XY}(\infty, \infty) = 1$;
4. $F_{XY}(-\infty, y) = F_{XY}(x, -\infty) = 0$;
5. $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) =$

   $F_{XY}(x_2, y_2) - F_{XY}(x_1, y_2) - F_{XY}(x_2, y_1)$

   $+ F_{XY}(x_1, y_1)$;
6. If X and Y are independent, then

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

$$4.\ F_{XY}\left(-\infty, y\right) = F_{XY}\left(x, -\infty\right) = 0;$$



5. $P\left(x_1 < X \leq x_2, y_1 < Y \leq y_2\right) =$

$$F_{XY}\left(x_2, y_2\right) - F_{XY}\left(x_1, y_2\right) - F_{XY}\left(x_2, y_1\right) + F_{XY}\left(x_1, y_1\right);$$

6. If X and Y are independent, then

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

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 82

Marginal

Probability Mass Functions (explained through an example)

- Consider a discrete random vector, that is, a vector whose entries are discrete random variables.

- When one of these entries is taken in **isolation**, its distribution can be characterized in terms of its probability mass function.

- This is called **marginal probability mass function**, in order to distinguish it from the joint probability mass function,

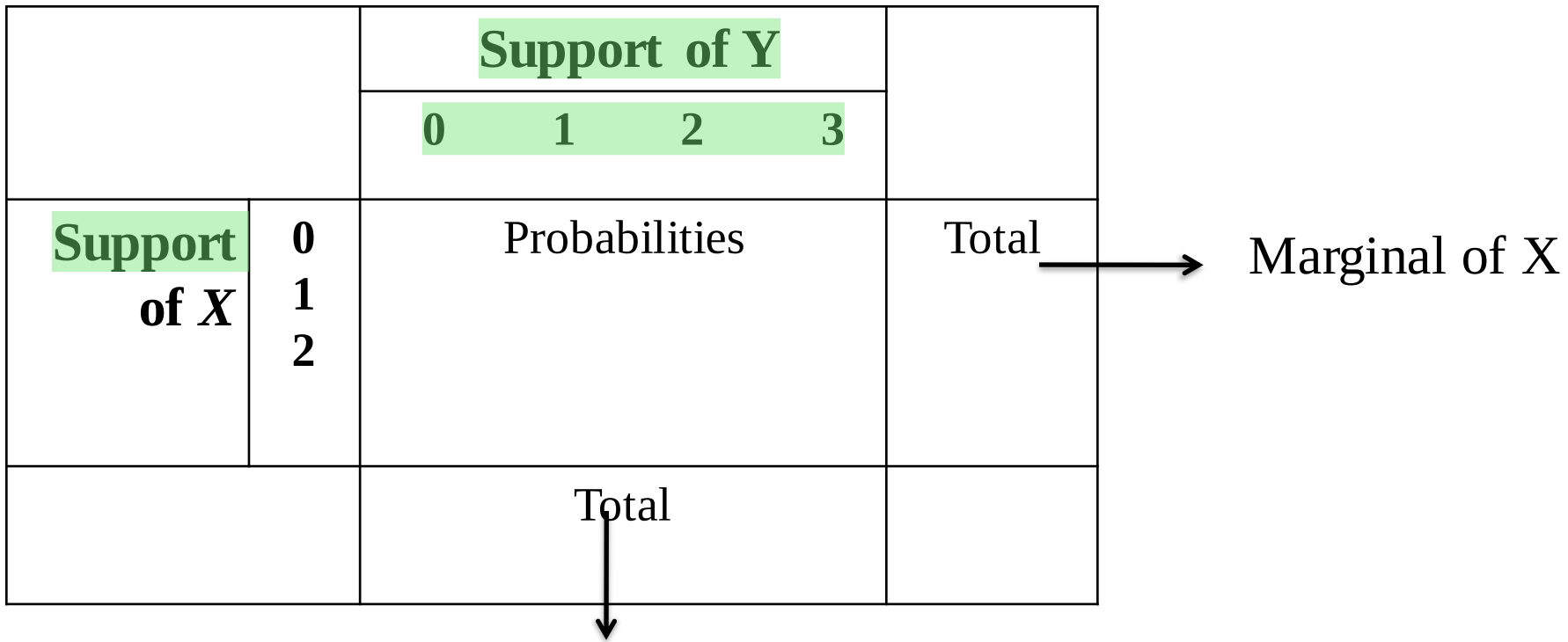
- which is instead used to characterize the joint distribution of all the entries of the random vector considered **together.**

Why the word ‘marginal’?

- In terms of a **tabled** joint pmf with rows comprised of X_1 support values and columns comprised of X_2 support values, this says that the distribution of X_1 can be obtained by **marginal sums of the rows**.

		Support of X_2	
		0 1 2 3	
Support of X_1	0 1 2	Probabilities	Total

Marginal of X_1


Support of Y
0 1 2 3
Support of X
0
1
2
Probabilities
Total
Total
Marginal of X
Marginal of Y

- To find the probability that X_1 is equal to x_1, **keep x_1 fixed ,** in continuous case, instead of summing we are intergrating one variable over the other.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 83

Marginal

Probability Mass Functions

(explained through an example)

Example:

- Let us attempt to understand this with the help of the following **example.**

A coin is tossed three times and our interest is in the ordered number pair (number of heads on first two tosses, number of heads on all three tosses),

Let H and T represent respectively, heads and tails. Then the sample space C is given by

$$C=\{TTT,TTH,THT,HTT,THH,HTH,HHT,HHH\}.$$

Let

X_1 denote the **number of H's** on the **first two** tosses

and

X_2 denotes the **number of H's** on **all three** flips.

Then the space of the discrete random vector (X_1,X_2) is given by

$\mathcal{D} = \{(0,0), (0,1),(1,1), (1,2),(2,2),(2,3)\}.$

with **probabilities**

$P[(X_1,X_2)=(0,0)]= 1/8$

$P[(X_1,X_2)=(0,1)]= 1/8$

$P[(X_1,X_2)=(1,1)]= 2/8$

$P[(X_1,X_2)=(1,2)]= 2/8$

$P[(X_1,X_2)=(2,2)]= 1/8$

$P[(X_1,X_2)=(2,3)]= 1/8$

We can conveniently table the pmf of the random vector (X_1, X_2) as:

		Support of X_2			
		0	**1**	**2**	**3**
Support of X_1	**0**	1/8	1/8	0	0
	1	0	2/8	2/8	0
	2	0	0	1/8	1/8

Now, to find the marginal probabilities, we have:

		Support of X_2				
		0	**1**	**2**	**3**	
Support of X_1	0	1/8	1/8	0	0	$p_{X_1}(x_1)$ 2/8
	1	0	2/8	2/8	0	4/8
	2	0	0	1/8	1/8	2/8
$p_{X_2}(x_2)$		**1/8**	**3/8**	**3/8**	**1/8**	8/8=1

Note that it is not **necessary** to have a **formula** for $p(x_1,x_2)$ in order to obtain the marginal PMFs

But

If we can develop a formula, all the better !

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 84

Marginal Probability Density Functions (explained through and example)

- In case of two continuous random variables X_1 and X_2, the joint probability density function is given by

$$f_{X_1,X_2}\left(x_1, x_2\right)$$

Obviously,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X_1,X_2}\left(x_1, x_2\right)dx_1dx_2 = 1$$

- In the continuous case, the **marginal** pdf of X_1 is found by integrating out x_2 ,i.e.,

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_2 \qquad (1)$$

- Similarly, the marginal pdf of X_2 is found by integrating out x_1.

$$f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2)dx_1 \qquad (2)$$

Example:

Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \textit{elsewhere} \end{cases}$$

It is easy to verify that

$$\int_0^1 \int_0^1 (x_1 + x_2) dx_1 dx_2 = 1$$

i.e.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 = 1$$

The marginal pdf of X_1 is

$$f_1(x_1) = \int_0^1 (x_1 + x_2)\,dx_2 = x_1 \int_0^1 1\,dx_2 + \int_0^1 x_2\,dx_2 = x_1 \left(x_2 \Big|_0^1 + \left(\frac{x_2^2}{2} \Big|_0^1 \right. \right.$$

$$= x_1(1-0) + \left(\frac{1^2}{2} - 0 \right) = x_1(1) + \frac{1}{2}$$

$$f_1(x_1) = x_1 + \frac{1}{2},\ 0 < x_1 < 1$$

zero elsewhere,

and the **marginal** pdf of X_2 is

$$f_2(x_2) = \int_0^1 (x_1 + x_2)\,dx_1 = \int_0^1 x_1\,dx_1 + x_2 \int_0^1 1\,dx_1 = \left(\frac{x_1^2}{2}\Big|_0^1 + x_2 \left(x_1 \right|_0^1 \right.$$

$$= \left(\frac{1^2}{2} - 0 \right) + x_2 (1 - 0) = \frac{1}{2} + x_2 (1)$$

$$f_2(x_2) = \frac{1}{2} + x_2,\ 0 < x_2 < 1,$$

and zero elsewhere.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 85

Another example of Computation of probabilities that can not be found through Marginal PDFs

Example:

Let $f\left(x_1, x_2\right)\begin{cases} = 4x_1x_2, 0 < x_1 < 1, 0 < x_2 < 1, \\ \text{zero elsewhere,} \end{cases}$

be the joint pdf of X_1 and X_2.

Find

i) $P(0 < X_1 < \frac{1}{2}, \frac{1}{4} < X_2 < 1)$,

ii) $P\left(X_1 = X_2\right)$,

iii) $P(X_1 < X_2)$

and

iv) $P(X_1 \leq X_2)$.

Solution:

The probability density function is given by

$$f\left(x_1, x_2\right)\begin{cases} = 4x_1x_2, 0 < x_1 < 1, 0 < x_2 < 1, \\ \text{zero elsewhere,} \end{cases}$$

As such:

$$\text{i) } P(0 < X_1 < \frac{1}{2}, \frac{1}{4} < X_2 < 1) = \int_{\frac{1}{4}}^{1}\int_{0}^{\frac{1}{2}} f\left(x_1, x_2\right) dx_1 dx_2$$

$$= \int_{\frac{1}{4}}^{1}\int_{0}^{\frac{1}{2}} 4x_1x_2 dx_1 dx_2 = 4\int_{\frac{1}{4}}^{1}\left(\int_{0}^{\frac{1}{2}} x_1x_2 dx_1\right) dx_2$$

or

$$P(0 < X_1 < \frac{1}{2}, \frac{1}{4} < X_2 < 1) == 4\int_{\frac{1}{4}}^{1} x_2 \left(\frac{x_1^2}{2}\Big|_0^{1/2} \right) dx_2$$

$$= 4\int_{\frac{1}{4}}^{1} x_2 \left(\frac{(1/2)^2}{2} - \frac{0}{2} \right) dx_2 = 4\int_{\frac{1}{4}}^{1} x_2 \left(\frac{1}{4\times 2} \right) dx_2$$

$$= \frac{1}{2}\int_{\frac{1}{4}}^{1} x_2 dx_2 = \frac{1}{2}\left(\frac{x_2^2}{2}\Big|_{1/4}^{1} \right) = \frac{1}{2}\left(\frac{(1)^2}{2} - \frac{(1/4)^2}{2} \right)$$

$$= \frac{1}{2}\left(\frac{1}{2} - \frac{1}{16\times 2} \right) = \frac{15}{64}$$

ii) Now to find $P(X_1=X_2)$:

$P(X_1=X_2)$ can be re-written as $P(X_1 - X_2 = 0)$

As the random variables, X_1 and X_2 are **continuous random variables**, the **difference** of these random variables is also a **continuous random variable.**

Now, zero is a constant

and

we know that the probability that a continuous random variable assumes a constant value is zero.

Therefore, $P(X_1 - X_2 = 0)$

iii) Find $P(X_1 < X_2)$.

According to the definition,

$$P(X_1 < X_2) = \int_0^1 \int_0^{x_2} f(x_1, x_2) dx_1 dx_2 = \int_0^1 \int_0^{x_2} 4x_1 x_2 dx_1 dx_2$$

$$= 4\int_0^1 x_2 \left[\left(\frac{x_1^2}{2} \Big|_0^{x_2} \right] dx_2 = 4\int_0^1 x_2 \left[\left(\frac{x_2^2}{2} - \frac{0}{2} \right) \right] dx_2 = 4\int_0^1 x_2 . \frac{x_2^2}{2} dx_2$$

$$= 2\int_0^1 x_2^3 dx_2 = 2\left[\left(\frac{x_2^3}{4} \Big|_0^1 \right] = 2\left(\frac{1}{4} - \frac{0}{4} \right) = \frac{1}{2}$$

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 86

Expected Value of a Real-Valued Function of a Random Vector

It is a straightforward extension of the concept of the expected value of a function of a random variable.

Let (X_1, X_2) be a random vector and let $\boldsymbol{Y = g(X_1, X_2)}$ where $g : R^2 \rightarrow R$ is some real-valued function of X_1 and X_2;.

For example:

- $Y = X_1 + X_2,$
- $Y = X_1^2 - e^{X_2}$

- Then Y is a **random variable** and we can determine its expectation by obtaining the **distribution** of Y.

First and foremost, let us determine the **conditions** under which the expectation (or expected value) of Y will **exist**.

Let us commence this discussion with the conditions that are required in the case of a **single** random variable.

A random variable X will be said to have a finite or infinite expectation (or expected value) **according as** $E(X)$ is a finite number or not.

If it is finite, then the expectation exists

and if it is **infinite (**i.e. if it is **not finite**), we shall say that the expectation of X does **not** exist.

So, naturally, we would like to determine the conditions under which E(X) will be **finite**.

By definition,

In the discrete case

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

In the continuous case

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

E(X) will be finite if our summation or integral **converges absolutely.**

Hence, naturally, the next question is: What is meant by **absolute convergence**?

The Concept of Absolute Convergence:

The term ‘Absolutely Convergent’ describes a series that converges **when all terms are replaced by their absolute values.**

Stated a little differently,

The term ‘Absolutely Convergent’ describes a series for which **the sum of all its terms remains finite** when all terms are replaced by their **absolute values.**

Let X be a random variable of the discrete type with probability mass function $p(x_k) = P\{X = x_k\}$, $k = 1,2,\ldots$.

If

$$\sum_{k=1}^{\infty} |x_k| p(x_k) < \infty,$$

then, we say that the expected value of X **exists** and we write

$$\mu = E(X) = \sum_{k=1}^{\infty} x_k p(x_k).$$

Similarly, if X is a random variable of the continuous type with probability density function $f(x)$

Then, if

$$\int_{-\infty}^{\infty} |x| f(x)\,dx < \infty,$$

then, we say that the expected value of X **exists** and we write

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

A similar definition is available for a function $g(X)$ of X.

Thus, if X is of continuous type and has probability density function $f(x)$, we will say that $\mathrm{E}[g(X)]$ exists and equals

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

provided that $\int_{-\infty}^{\infty} |g(x)| f(x) dx < \infty.$

A similar definition is available for a function $g(X_1, X_2)$ of two random variables X_1 and X_2.

If both X_1 and X_2 are of continuous type and have a joint probability density function $f(x_1, x_2)$, we will say that $E[g(x_1, x_2)]$ exists and equals

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2) f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2$$

provided that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left|g(x_1, x_2)\right| f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2 < \infty.$$

Likewise, if the random vector (X_1, X_2) is discrete and we let $\boldsymbol{Y = g(X_1, X_2)}$, then E($Y$) exists if

$$\sum_{x_1}\sum_{x_2}\left|g\left(x_1, x_2\right)\right|p_{X_1,X_2}\left(x_1, x_2\right) < \infty.$$

and is given by

$$E\left(Y\right) = \sum_{x_1}\sum_{x_2} g\left(x_1, x_2\right)p_{X_1,X_2}\left(x_1, x_2\right).$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 87

Yet another example of Computation of probabilities that can not be found through Marginal PDFs

Example:

Let $A_1 = \{(x, y): x \leq 2, y \leq 4\}$,

$A_2 = \{(x , y): x \leq 2, y \leq 1\}$,

$A_3 = \{(x , y): x \leq 0, y \leq 4\}$,

and

$A_4 = \{(x , y): x \leq 0, y \leq 1\}$

be subsets of the space A of two random variables X and Y which is the entire two-dimensional plane.

If $P(A_1)=7/8$, $P(A_2)=4/8$, $P(A_3)=3/8$, and $P(A_4)=2/8$, find $P(A_5)$, where $A_5=\{(x , y): 0 < x \leq 2, 1< y \leq 4\}$.

HINT:

The required probability will be obtained by splitting the integrals representing the various probabilities.

$$P(A_1) = \int_{-\infty}^{2}\int_{\infty}^{4} f(x,y)\,dydx = \frac{7}{8} = \int_{-\infty}^{2}\int_{\infty}^{1} f(x,y)\,dydx + \int_{-\infty}^{2}\int_{1}^{4} f(x,y)\,dydx$$

$$P(A_2) = \int_{-\infty}^{2}\int_{\infty}^{1} f(x,y)\,dydx = \frac{4}{8}$$

So,

$$\frac{7}{8} = P(A_1) = \frac{4}{8} = P(A_2) + \int_{-\infty}^{2}\int_{1}^{4} f(x,y)\,dydx$$

$$\frac{7}{8} = \frac{4}{8} + \int_{-\infty}^{2}\int_{1}^{4} f(x,y)\,dydx$$

yulearninghub.xyz | by VU Learning HUB

$$P(A_3) = \int_{-\infty}^{0} \int_{\infty}^{4} f(x,y)\,dydx$$

So,

$$P(A_1) = P(A_3) + \int_{0}^{2} \int_{-\infty}^{4} f(x,y)\,dydx$$

$$\frac{7}{8} = \frac{3}{8} + \int_{0}^{2} \int_{-\infty}^{4} dydx$$

$$P(A_4) = \int_{-\infty}^{0} \int_{\infty}^{1} f(x,y)\,dydx$$

So,

$$P(A_3) = P(A_4) + \int_{-\infty}^{0} \int_{1}^{4} f(x,y)\,dydx$$

$$\frac{3}{8} = \frac{2}{8} + \int_{-\infty}^{0} \int_{1}^{4} f(x,y)\,dydx$$

So,

$$\int_{-\infty}^{2}\int_{1}^{4} f(x,y)\,dydx = \frac{7}{8} - \frac{4}{8} = \frac{3}{8} = P(A_6)$$

$$\int_{0}^{2}\int_{-\infty}^{4} f(x,y)\,dydx = \frac{7}{8} - \frac{3}{8} = \frac{4}{8} = P(A_7)$$

and

$$\int_{-\infty}^{0}\int_{1}^{4} f(x,y)\,dydx = \frac{3}{8} - \frac{2}{8} = \frac{1}{8} = P(A_8)$$

vulearninghub.xyz | by VU Learning HUB

Hence,

$$P(A_5)\int_{0}^{2}\int_{1}^{4} f(x,y)\,dydx$$

So,

$$\int_{-\infty}^{2}\int_{1}^{4} f(x,y)\,dydx - \int_{-\infty}^{0}\int_{1}^{4} f(x,y)\,dydx$$

$$= \frac{3}{8} - \frac{1}{8} = \frac{2}{8} = \frac{1}{4}$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 88

Determination of the Expectation of the Product of two discrete random variables (explained through an example)

Example:

Let $X_1 X_2$ be two random variables with the joint pmf.

$$p(x_1, x_2) \begin{cases} \dfrac{(x_1 + x_2)}{12} & , \ for \ \ x_1 = 1, 2, x_2 = 1, 2, \\ 0 & elsewhere \end{cases}$$

Compute

$E(X_1), E(X_2)$ *and*

$E(X_1 X_2)$. Is $E(X_1 X_2) = E(X_1) E(X_2)$?

vulearninghub.xyz | by VU Learning HUB

Solution:

First we know that,

$$p(x_1) = \sum_{x_2=1,2} \frac{x_1 + x_2}{12} = \frac{x_1 + 1}{12} + \frac{x_1 + 2}{12} = \frac{x_1 + 1 + x_1 + 2}{12} = \frac{2x_1 + 3}{12} \quad , \quad x_1 = 1, 2$$

In the same way,

$$p(x_2) = \sum_{x_1=1,2} \frac{x_1 + x_2}{12} = \frac{1 + x_2}{12} + \frac{2 + x_2}{12} = \frac{2x_2 + 3}{12} \quad , \quad x_2 = 1, 2$$

Therefore, as X_1 and X_2 are symmetric, they have the same pmfs and

Hence, their mathematical expectations will be equal.

Now computing

$$E(X_1) = \sum_{x_1=1,2} x_1 \frac{2x_1+3}{12} = (1)\frac{2(1)+3}{12} + (2)\frac{2(2)+3}{12} = \frac{5}{12} + \frac{14}{12} = \frac{19}{12}$$

$$E(X_2) = E(X_1) = \sum_{x_2=1,2} x_2 \frac{2x_2+3}{12} = (1)\frac{2(1)+3}{12} + (2)\frac{2(2)+3}{12} = \frac{5}{12} + \frac{14}{12} = \frac{19}{12}$$

Now $E(X_1X_2) = \sum_{x_1,x_2} x_1.x_2 \frac{x_1 + x_2}{12}$

$$= (1\times1)\frac{1+1}{12} + (1\times2)\frac{1+2}{12} + (2\times1)\frac{2+1}{12} + (2\times2)\frac{2+2}{12}$$

$$= \frac{2}{12} + \frac{6}{12} + \frac{6}{12} + \frac{16}{12} = \frac{30}{12} = 2.5$$

On the other hand,

$$E(X_1)E(X_2) = \frac{19}{12} \times \frac{19}{12} = 2.5069$$

Hence

$$E(X_1X_2) \neq E(X_1)E(X_2)$$

vulearninghub.xyz By VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 89

Determination of the Expectation of the Product of two continuous random variables (explained through an example)

Example:

Let X_1, X_2 be two random variables with the joint pdf $f(x_1, x_2) = 4x_1x_2$, $0 < x_1 < 1$, $0 < x_2 < 1$, zero elsewhere. Compute $E(X_1)$, $E(X_2)$

$$E(X_1), E(X_1^2), E(X_2), E(X_2^2), and\ E(X_1X_2).$$

Is $E(X_1X_2) = E(X_1)E(X_2)$?

Find $E(3X_1 - 2X_1^2 + 6X_1X_2)$.

vulearninghub.xyz | by VU Learning HUB

Firstly, $\mathrm{E}(X_1)$

As a rule:

$$E(X_1) = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1) dx_1 = \int_{-\infty}^{\infty} x_1 \left(\int_{-\infty}^{\infty} f(x_1, x_2) dx_2 \right) dx_1$$

$$= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x_1 f(x_1, x_2) dx_2 \right) dx_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f(x_1, x_2) dx_2 dx_1$$

Similarly, $\mathrm{E}(X_2)$

As a rule:

$$E\left(X_2\right)=\int_{-\infty}^{\infty} x_2 f_{X_2}(x_2)dx_2=\int_{-\infty}^{\infty} x_2 f_{X_2}(x_2)dx_2=\int_{-\infty}^{\infty} x_2\left(\int_{-\infty}^{\infty} f(x_1,x_2)dx_2\right)dx_1$$

$$=\int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x_2 f(x_1,x_2)dx_2\right)dx_1=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1,x_2)dx_2dx_1$$

Firstly, E(X_1)

$$E(X_1) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1 f(x_1, x_2)dx_2dx_1 = \int_0^1\int_0^1 x_1.4x_1x_2dx_2dx_1 = \int_0^1\int_0^1 4x_1^2x_2dx_2dx_1$$

$$= \int_0^1\left(\int_0^1 4x_1^2x_2dx_2\right)dx_1 = \int_0^1\left(4x_1^2.\left(\frac{x_2^2}{2}\Big|_0^1\right)dx_1 = \int_0^1\left(4x_1^2.\left(\frac{1}{2}-\frac{0}{2}\right)\right)dx_1\right.$$

$$= \int_0^1\left(2x_1^2\right)dx_1 = 2\left(\frac{x_2^3}{3}\Big|_0^1 = 2\left(\frac{1}{3}-\frac{0}{3}\right)\right.$$

$$\therefore E(X_1) = \frac{2}{3}$$

Then E(X_2)

$$E(X_2)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_2 f(x_1,x_2)dx_1dx_2 = \int_0^1\int_0^1 x_2.4x_1x_2dx_1dx_2 = \int_0^1\int_0^1 4x_1x_2^2dx_1dx_2$$

$$= \int_0^1\left(\int_0^1 4x_2^2x_1dx_1\right)dx_2 = \int_0^1\left(4x_2^2.\left(\frac{x_1^2}{2}\Big|_0^1\right)dx_2 = \int_0^1\left(4x_2^2.\left(\frac{1}{2}-\frac{0}{2}\right)\right)dx_2\right.$$

$$= \int_0^1\left(2x_2^2\right)dx_2 = 2\left(\frac{x_2^3}{3}\Big|_0^1 = 2\left(\frac{1}{3}-\frac{0}{3}\right)\right.$$

$$\therefore E(X_2)=\frac{2}{3}$$

Similarly $E(X_1X_2)$

As a rule:

$$E(X_1X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2\, f_{X_1,X_2}(x_1,x_2)dx_2dx_1 = \int_{-\infty}^{\infty} x_1x_2\left(\int_{-\infty}^{\infty} f(x_1,x_2)dx_2\right)dx_1$$

$$= \int_{-\infty}^{\infty}\left(\int_{-\infty}^{\infty} x_1x_2\, f(x_1,x_2)dx_2\right)dx_1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2\, f(x_1,x_2)dx_2dx_1$$

Now E(X_1X_2)

$$E(X_1X_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x_1x_2\, f(x_1,x_2)dx_1dx_2 = \int_0^1\int_0^1 x_1x_2.4x_1x_2dx_1dx_2 = \int_0^1\int_0^1 4x_1^2x_2^2dx_1dx_2$$

$$= \int_0^1\left(\int_0^1 4x_1^2x_2^2dx_1\right)dx_2 = \int_0^1\left(4x_2^2.\left(\frac{x_1^3}{3}\Big|_0^1\right)dx_2 = \int_0^1\left(4x_2^2.\left(\frac{1}{3}-\frac{0}{3}\right)\right)dx_2\right.$$

$$= \frac{4}{3}\int_0^1 x_2^2dx_2 = \frac{4}{3}\left(\frac{x_2^3}{3}\Big|_0^1 = \frac{4}{3}\left(\frac{1}{3}-\frac{0}{3}\right) = \frac{4}{9}\right.$$

$$\therefore E(X_1X_2) = E(X_1)E(X_2) = \left(\frac{2}{3}\right)\left(\frac{2}{3}\right) = \frac{4}{9}$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 90

Determination of the Expectation of the Ratio of two continuous random variables (explained through an example)

Example:

Let X_1 and X_2 have the pdf

$$f(x_1, x_2) = \begin{cases} 8x_1x_2 & 0 < x_1 < x_2 < 1 \\ 0 & \textit{elsewhere} \end{cases}$$

Suppose the random variable Y is defined by $Y=X_1/X_2$ and we are interested to determine the $E(Y)$.

vulearninghub.xyz | by VU Learning HUB

Solution:

We can determining $E(Y)$ in two ways.

The first way is **by definition,** i.e., find the **distribution** of Y and then determine its expectation.

The cdf of Y, for $0 < y < 1$, is derived as follows:

$$F_Y(y) = P(Y \le y) = P(X_1 \le yX_2) = \int_0^1 \int_0^{yx_2} 8x_1x_2\,dx_1\,dx_2$$

$$= \int_0^1 \left(\int_0^{yx_2} 8x_1x_2\,dx_1 \right) dx_2 = \int_0^1 \left(8x_2 \left(\frac{x_1^2}{2}\Big|_0^{yx_2} \right) \right) dx_2 = \int_0^1 \left(8x_2 \left(\frac{y^2x_2^2}{2} - \frac{0}{2} \right) \right) dx_2$$

$$= \int_0^1 \left(8x_2 \left(\frac{y^2x_2^2}{2} \right) \right) dx_2 = \int_0^1 4y^2x_2^3\,dx_2 = 4y^2 \left(\frac{x_2^4}{4}\Big|_0^1 \right) = 4y^2 \left(\frac{1}{4} - \frac{0}{4} \right)$$

$$= 4y^2 \left(\frac{1}{4} \right) = y^2.$$

Hence

$$F_Y(y) = \begin{cases} y^2 & 0 < y < 1 \\ 0 & elsewhere \end{cases}$$

And hence, the pdf of Y is

$$f_Y(y) = F_Y'(y) = \begin{cases} 2y & 0 < y < 1 \\ 0 & elsewhere \end{cases}$$

which leads to

$$E(Y) = \int_0^1 y(2y)dy$$

$$= 2\int_0^1 y^2 dy = 2\left(\frac{y^3}{3}\Big|_0^1 = 2\left(\frac{1}{3} - \frac{0}{3}\right) = 2\left(\frac{1}{3}\right)\right.$$

$$\therefore E(Y) = \frac{2}{3},$$

$$E(Y) = E\left(\frac{X_1}{X_2}\right) = \int_0^1 \left\{ \int_0^{x_2} \left(\frac{x_1}{x_2}\right) 8 x_1 x_2 \, dx_1 \right\} dx_2 = \int_0^1 \left\{ \int_0^{x_2} 8 x_1^2 \, dx_1 \right\} dx_2$$

$$= \int_0^1 \left\{ 8 \left(\frac{x_1^3}{3} \Big|_0^{x_2} \right\} dx_2 = \int_0^1 \left\{ 8 \left(\frac{x_2^3}{3} - \frac{0}{3} \right) \right\} dx_2 = \int_0^1 \left\{ 8 \left(\frac{x_2^3}{3} \right) \right\} dx_2$$

$$= \int_0^1 \frac{8}{3} x_2^3 \, dx_2 = \frac{8}{3} \left(\frac{x_2^4}{4} \Big|_0^1 = \frac{8}{3} \left(\frac{1}{4} - \frac{0}{4} \right) = \frac{8}{3} \times \frac{1}{4} = \frac{2}{3}. \right.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 91

How to obtain product moments and simple moments from the MGF of a random vector?

In a simplified notation

$$E(X) = \frac{\partial M(0,0)}{\partial t_1}, \qquad E(Y) = \frac{\partial M(0,0)}{\partial t_2}$$

$$E(X^2) = \frac{\partial^2 M(0,0)}{\partial t_1^2}, \quad E(Y^2) = \frac{\partial^2 M(0,0)}{\partial t_2^2}$$

and

$$E(XY) = \frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2}$$

Therefore

$$\mu_1 = E(X) = \frac{\partial M(0,0)}{\partial t_1},$$

$$\mu_2 = E(Y) = \frac{\partial M(0,0)}{\partial t_2}$$

and

$$\sigma_1^2 = E(X^2) - [E(X)]^2 = \left(\frac{\partial^2 M(0,0)}{\partial t_1^2}\right) - \left(\frac{\partial M(0,0)}{\partial t_1}\right)^2$$

$$\sigma_2^2 = E(Y^2) - [E(Y)]^2 = \left(\frac{\partial^2 M(0,0)}{\partial t_2^2}\right) - \left(\frac{\partial M(0,0)}{\partial t_2}\right)^2$$

vulearninghub.xyz | by VU Learning HUB

And, as far as the Co variance is concerned we have

$$E\left[(X-\mu_1)(Y-\mu_2)\right] = E(XY) - E(X)E(Y)$$

$$= \frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2} - \left(\frac{\partial M(0,0)}{\partial t_1}\right)\left(\frac{\partial M(0,0)}{\partial t_2}\right)$$

And from the above we can compute the correlation coefficient ρ.

- Thus the **correlation coefficient** may be computed by using the **mgf** of the joint distribution if that function is readily available.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 92

Expected Value of a Random Vector

(defined in terms of Component wise expectation)

Definition

(Expected Value of a Random Vector)

Let $\boldsymbol{X} = \left(X_1, X_2\right)'$ be a random vector.

Then the expected value of $\boldsymbol{X}$ exists if the expectations of X_1 and X_2 exist.

If $\mathrm{E}(X_1)$ and $\mathrm{E}(X_2)$ exist, then the expected value of $\boldsymbol{X}$ is given by

$$\mathrm{E}(\boldsymbol{X}) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix}$$

Similarly, if we toss two fair coins together, then expected value of $X_1 = 1$

AND

if we toss a fair die, then expected value of $X_2 = 3.5$

If $E(X_1)$ and $E(X_2)$ exist, then the expected value of $\boldsymbol{X}$ is given by

$$E(X) = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}$$

If we consider the two experiments together, then the expected value of $\boldsymbol{X}$ is given by

$$E[X] = \begin{bmatrix} E(X_1) \\ E(X_2) \end{bmatrix} = \begin{bmatrix} 1 \\ 3.5 \end{bmatrix}$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 93

Linear Combination of Expected Values of Real-Valued Functions of a Random Vector (explained through an example)

Theorem:

Let (X_1, X_2) be a random vector, and,

let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be random variables the expectations of which **exist**.

Then, for all real numbers k_1 and k_2,

$$E\left(k_1 Y_1 + k_2 Y_2\right) = k_1 E\left(Y_1\right) + k_2 E\left(Y_2\right) \qquad (1)$$

Before we begin the proof, let us remind ourselves what is meant by the **existence** of a function of two random variables.

If $\boldsymbol{Y=g(X_1, X_2)}$ is a function of two random variables X_1 and X_2 both of which are of the **continuous** type and have a joint probability density function $f(x_1, x_2)$, we will say that $E[g(x_1, x_2)]$ exists and equals

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_1, x_2) f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2$$

provided that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g(x_1, x_2)| f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2 < \infty.$$

In other words, if $\boldsymbol{Y = g(X_1, X_2)}$,

then

$$E(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f_{X_1, X_2}(x_1, x_2) dx_1 dx_2$$

provided that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x_1, x_2)| f_{X_1, X_2}(x_1, x_2) dx_1 dx_2 < \infty.$$

Focusing on the theorem again:

Theorem: Let (X_1, X_2) be a random vector, and let $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ be random variables the expectations of which exist.

Then, for all real numbers k_1 and k_2,

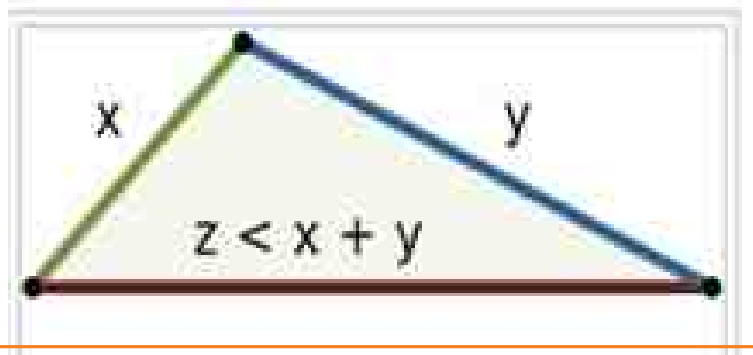
$$E\left(k_1 Y_1 + k_2 Y_2\right) = k_1 E\left(Y_1\right) + k_2 E\left(Y_2\right) \qquad (1)$$

Proof:

For the continuous case, the **existence** of the expected value of $k_1Y_1+k_2Y_2$ follows directly from the **triangle inequality** and **linearity of integrals.**

So, let us first see what is meant by the Triangle Inequality:

- For any triangle the sum of the length of any two sides must be greater than or equal to the length of the third side.



So now according to Euclidian geometry:

- If *x* and *y* are two real numbers than the absolute value (the modulus) of |x+y| is less than or equal to the modulus of |x| and modulus of |y|

$$\|x + y\| \leq \|x\| + \|y\|$$

Reverting back to the Proof:

For the continuous case, the **existence** of the expected value of $k_1Y_1+k_2Y_2$ follows directly from the **triangle inequality** and **linearity of integrals.**

Now, according to the triangle inequality, for any two real numbers x and y,

$$|x+y| \le |x| + |y|.$$

This implies that

$$|k_1 g_1(x_1, x_2) + k_2 g_1(x_1, x_2)| \le |k_1 g_1(x_1, x_2)| + |k_2 g_2(x_1, x_2)|$$

$$\Rightarrow |k_1 g_1(x_1, x_2) + k_2 g_1(x_1, x_2)| \le |k_1||g_1(x_1, x_2)| + |k_2||g_2(x_1, x_2)|$$

$$\Rightarrow |k_1 g_1(x_1, x_2) + k_2 g_1(x_1, x_2)| f_{X_1, X_2}(x_1, x_2)$$

$$\le \left(|k_1||g_1(x_1, x_2)| + |k_2||g_2(x_1, x_2)|\right) f_{X_1, X_2}(x_1, x_2)$$

or

$$|k_1 g_1(x_1, x_2) + k_2 g_1(x_1, x_2)| f_{X_1, X_2}(x_1, x_2)$$

$$\le \left(|k_1||g_1(x_1, x_2)| + |k_2||g_2(x_1, x_2)|\right) f_{X_1, X_2}(x_1, x_2)$$

Now $\left|k_1 g_1\left(x_1, x_2\right)+k_2 g_1\left(x_1, x_2\right)\right| f_{X_1, X_2}\left(x_1, x_2\right)$

$$\leq\left(\left|k_1\right|\left|g_1\left(x_1, x_2\right)\right|+\left|k_2\right|\left|g_2\left(x_1, x_2\right)\right|\right) f_{X_1, X_2}\left(x_1, x_2\right)$$

can be re-written as

$$\left|k_1 g_1\left(x_1, x_2\right)+k_2 g_1\left(x_1, x_2\right)\right| f_{X_1, X_2}\left(x_1, x_2\right)$$

$$\leq\left|k_1\right|\left|g_1\left(x_1, x_2\right)\right| f_{X_1, X_2}\left(x_1, x_2\right)+\left|k_2\right|\left|g_2\left(x_1, x_2\right)\right| f_{X_1, X_2}\left(x_1, x_2\right)$$

....(1)

Now, let us consider the concept of Linearity of Integration:

- In calculus, the integral of any linear combination of functions equals the same linear combination of the integrals of functions,; this property is known as linearity of integration.

- Linearity of integration is related to the linearity of summation, since integrals are thought of as infinite sums.

As such, applying integrals on both sides of inequality (1), we have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|k_1 g_1(x_1,x_2)+k_2 g_1(x_1,x_2)\right| f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$$

$$\leq |k_1|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|g_1(x_1,x_2)\right| f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$$

$$+|k_2|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|g_2(x_1,x_2)\right| f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$$

vtulearninghub.xyz | by VTU Learning HUB

Now, according to the statement of the Theorem,

$Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$

are random variables the expectations of which exist

i.e.

$$E(Y_1) = E[g_1(X_1, X_2)] \text{ exists} \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g_1(x_1, x_2)| f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2 < \infty.$$

$$E(Y_2) = E[g_2(X_1, X_2)] \text{ exists} \Rightarrow \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |g_2(x_1, x_2)| f_{X_1, X_2}(x_1, x_2)\, dx_1 dx_2 < \infty$$

Hence $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|k_1 g_1(x_1,x_2)+k_2 g_1(x_1,x_2)\right| f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$

$$\leq |k_1|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|g_1(x_1,x_2)\right| f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$$

$$+|k_2|\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|g_2(x_1,x_2)\right| f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2 < \infty$$

Therefore $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[k_1 g_1(x_1,x_2)+k_2 g_1(x_1,x_2)\right] f_{X_1,X_2}(x_1,x_2)\,dx_1 dx_2$

vulearninghub.xyz | by VU Learning HUB

$=\mathrm{E}\left[k_1 g_1(x_1,x_2)+k_2 g_1(x_1,x_2)\right]$ exists.

By **once** **again** using linearity of the integral, we have

$$E\left(k_1Y_1+k_2Y_2\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[k_1g_1\left(x_1,x_2\right)+k_2g_1\left(x_1,x_2\right)\right]f_{X_1,X_2}\left(x_1,x_2\right)dx_1dx_2$$

$$=k_1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g_1\left(x_1,x_2\right)f_{X_1,X_2}\left(x_1,x_2\right)dx_1dx_2+k_2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g_2\left(x_1,x_2\right)f_{X_1,X_2}\left(x_1,x_2\right)dx_1dx_2$$

$$=k_1E\left(Y_1\right)+k_2E\left(Y_2\right),$$

i.e., the desired result.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 94

Transformation in the case of a Bivariate Probability Mass Function (explained through an example)

- Let (X_1,X_2) be a random vector.

- Suppose we know the joint distribution of (X_1,X_2) and we seek the distribution of a **transformation** of (X_1,X_2), say, **$Y=g(X_1,X_2)$** or **$Y_1=g_1(X_1,X_2)$, $Y_2=g_2(X_1,X_2)$**

- We may be able to obtain the cdf of Y. Another way is to use a transformation.

It is best to discuss the discrete and continuous cases **separately.**

DISCRETE CASE:

- Let $p_{X1,X2}(x_1,x_2)$ be the joint pmf of two discrete-type random variables X_1 and X_2 with S the (two-dimensional) set of points at which $p_{X1,X2}(x_1,x_2) > 0$; i.e., S is the **support** of (X_1,X_2).

- Let $y_1=u_1(x_1,x_2)$ and $y_2 = u_2(x_1,x_2)$ define a **one-to-one transformation** that maps S onto T

The joint pmf of the two new random variables

$Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$

is given by

$$p_{Y_1,Y_1}(y_1, y_2) = \begin{cases} p_{X_1,X_1}\left[w_1(y_1, y_2), w_2(y_1, y_2)\right] & (y_1, y_2) \in T \\ 0 & \textit{elsewhere,} \end{cases}$$

where $x_1 = w_1(y_1, y_2), x_2 = w_2(y_1, y_2)$

is the single-valued inverse of

$y_1 = u_1(X_1, X_2)$ and $y_2 = u_2(X_1, X_2)$.

In using this **change of variable technique**, it should be emphasized that we need **two "new" variables** to replace the two "old" variables.

After we have found the joint pmf $p_{Y_1,Y_2}(y_1, y_2)$, we may obtain the **marginal** pmf of Y_1 by summing on y_2 or the **marginal** pmf Y_2 by summing on y_1.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 95

Transformation in the case of a Bivariate Probability Mass Function (explained through an example)

Example:

Let X_1 and X_2 have the joint pmf

$$p_{X_1,X_2}(x_1,x_2)=\begin{cases}\dfrac{\mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1-\mu_2}}{x_1!x_2!}, & x_1=0,1,2,3,\ldots,\ x_2=0,1,2,3,\ldots\\ zero\ elsewhere\end{cases}$$

where μ_1 and μ_2 are fixed positive real numbers.

Note: The space S is the set of points (x_1,x_2), where each of x_1 and x_2 is a **non-negative integer.**

vulearninghub.xyz | by VU Learning HUB

- Suppose that we wish to find the pmf of $Y_1=X_1+X_2$.

- If we use the change of variable technique, we need to define a **second random variable Y_2.**

Because Y_2 is of **no interest** to us, let us choose it in such a way that we have a **simple** one-to-one transformation.

For example, take $\boldsymbol{Y_2 = X_2}$.

Then $\boldsymbol{y_1 = x_1 + x_2}$ **and** $\boldsymbol{y_2 = x_2}$ represent a one-to-one transformation that maps S onto

$$T = \{(y_1, y_2) : y_2 = 0, 1, \ldots, y_1 \text{ and } y_1 = 0, 1, 2, \ldots\}$$

Note that if $(y_1, y_2) \in T$, then $0 \le y_2 \le y_1$.

vulearninghub.xyz | by VU Learning HUB

The inverse functions are given by

$$x_1 = y_1 - y_2 \quad and \quad x_2 = y_2.$$

Thus the joint pmf of Y_1 and Y_2 is

$$p_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \dfrac{\mu_1^{y_1-y_2}\,\mu_2^{y_2}\,e^{-\mu_1-\mu_2}}{(y_1-y_2)!\,y_2!}, & (y_1, y_2) \in T, \\ zero\ elsewhere \end{cases}$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 96

Transformation in the case of a Bivariate Probability Density Function using the Jacobian of Transformation (explained through an example)

Example:

Suppose (X_1, X_2) have the joint pdf

$$f_{X_1,X_2}(x_1, x_2) = \begin{cases} 1 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & elsewhere. \end{cases}$$

The support of (X_1, X_2) is then the set

$$S=\{(x_1,x_2) : 0 < x_1 < 1,\ 0 < x_2 < 1\}$$

Suppose $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

where,

$$y_1 = u_1(x_1, x_2) = x_1 + x_2$$

$$y_2 = u_2(x_1, x_2) = x_1 - x_2.$$

- This transformation is **one-to-one.**
- We first determine the set T in the y_1y_2-plane that is the mapping of S under this this transformation.

- Inverse function

Now $y_1 = x_1 + x_2,\ y_2 = x_1 - x_2$

implies that

$$y_1 + y_2 = 2x_1 \Rightarrow x_1 = w_1(y_1, y_2) = \frac{y_1 + y_2}{2}$$

&

$$y_1 - y_2 = 2x_2 \Rightarrow x_2 = w_2(y_1, y_2) = \frac{y_1 - y_2}{2}.$$

vulearninghub.xyz | by VU Learning HUB

Next, the Jacobian is given by

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4} - \frac{1}{4} = -\frac{2}{4} = -\frac{1}{2}$$

$$\Rightarrow |J| = \frac{1}{2}.$$

Now, using the inequalities $0 < x_1 < 1$ *and* $0 < x_2 < 1$ we can write

$$0 < \frac{y_1 + y_2}{2} < 1 \quad \text{and} \quad 0 < \frac{y_1 - y_2}{2} < 1$$

and it is easy to see that these are equivalent to

$$-y_1 < y_2, \quad y_2 < 2 - y_1, \quad y_2 < y_1, \quad y_1 - 2 < y_2$$

and these define the set T.

Hence, the joint pdf of (Y_1, Y_2) is given by

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} f_{X_1,X_2}\left(\frac{y_1+y_2}{2}, \frac{y_1-y_2}{2}\right).|J|, & (y_1,y_2) \in T \\ 0 & \textit{elsewhere} \end{cases}$$

or

$$f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} 1.\frac{1}{2}, & (y_1,y_2) \in T \\ 0 & \textit{elsewhere} \end{cases}$$

In other words, the joint pdf of (Y_1, Y_2) is given by

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} \frac{1}{2}, & (y_1, y_2) \in T \\ 0 & \textit{elsewhere} \end{cases}$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 97

Transformation in the case of a Bivariate Probability Mass Function using the MGF technique (explained through an example)

Example:

Here X_1 and X_2 have the joint pmf

$$p_{X_1,X_2}(x_1,x_2)=\begin{cases}\dfrac{\mu_1^{x_1}\mu_2^{x_2}e^{-\mu_1}e^{-\mu 2}}{x_1!x_2!} & x_1=0,1,2,3,\ldots,x_2=0,1,2,3,\ldots,\\ 0 & elsewhere,\end{cases}$$

where μ_1 and μ_2 are fixed positive real numbers.

Vulearninghub.xyz by VU Learning HUB

MGF of X_1 *and* X_1

$$E\left(e^{tX_1}\right)=\left[e^{\mu_1\left(e^t-1\right)}\right]$$

$$E\left(e^{tX_2}\right)=\left[e^{\mu_2\left(e^t-1\right)}\right]$$

Let $Y = X_1 + X_2$ and consider

$$E\left(e^{tY}\right) = \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} e^{t\left(x_1+x_2\right)} p_{X_1,X_2}\left(x_1, x_2\right)$$

$$= \sum_{x_1=0}^{\infty} e^{tx_1} \frac{\mu^{x_1} e^{-\mu_1}}{x_1!} \sum_{x_2=0}^{\infty} e^{tx_2} \frac{\mu^{x_2} e^{-\mu_2}}{x_2!}$$

$$E\left(e^{tY}\right)=\left[e^{-\mu_1}\sum_{x_1=0}^{\infty}\frac{\left(e^t\mu_1\right)^{x_1}}{x_1!}\right]\left[e^{-\mu_2}\sum_{x_2=0}^{\infty}\frac{\left(e^t\mu_2\right)^{x_2}}{x_2!}\right]$$

$$=\left[e^{\mu_1\left(e^t-1\right)}\right]\left[e^{\mu_2\left(e^t-1\right)}\right]$$

$$=e^{\left(\mu_1+\mu_2\right)\left(e^t-1\right)}.$$

Note that the factors in the brackets in the next-to-least equality are mgfs of X_1 and X_2, respectively.

Hence, the mgf of Y is the same as that of X_1 except μ_1 has been replaced by $\mu_1+\mu_2$.

Therefore,

By the uniqueness of mgfs, the pmf of Y must be

$$p_Y(y) = e^{-(\mu_1+\mu_2)} \frac{(\mu_1+\mu_2)^y}{y!}, \quad y = 0,1,2,\ldots,$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 98

Characteristic Function

The Characteristic function of a probability distribution is denoted by $\varphi_X(t)$ and is defined as the expected value of e^{itX} i.e.

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right)$$

where t is an arbitrary real number and i is the imaginary number given by $i = \sqrt{-1}$.

If the PDF of X is given by $f(x)$, then

$$\varphi_X(t) = \mathrm{E}\left(e^{itX}\right) = \int_{-\infty}^{\infty} e^{itx} f(x)dx.$$

The characteristic function exists for **every** distribution and it possesses the uniqueness property. In other words, every distribution has **unique** characteristic function.

In other words, the characteristic function **uniquely** identifies the probability distribution.

In fact, it **completely determines** the behavior and properties of the distribution.

Algebraic expressions of Some well-known Characteristic Functions

Characteristic Functions of some well-known discrete distributions:

Distribution	Cumulative function $\phi(t)$
Bernoulli $P(X=1)=p$	$1-p+pe^{it}$

Characteristic Functions of some well-known discrete distributions:

Distribution	Cumulative function $\phi(t)$
Binomial $B(n, p)$	$\left(1-p+pe^{it}\right)^n$

Characteristic Functions of some well-known discrete distributions:

Distribution	Cumulative function $\phi(t)$
Poisson $Pois(\lambda)$	$e^{\lambda\left(e^{it}-1\right)}$

Characteristic Functions of some well-known discrete distributions:

Distribution	Cumulative function $\phi(t)$
Geometric $(1-p)^{k-1}\,p$	$\dfrac{pe^{it}}{1-(1-p)e^{it}}$

Characteristic Functions of some well-known discrete distributions:

Distribution	Cumulative function $\phi(t)$
Uniform (discrete) $\mathrm{U}(a,b)$	$\dfrac{e^{ait}-e^{(b+1)it}}{(b-a+1)(1-e^{it})}$

Characteristic Functions of some well-known continuous distributions:

- Exponential
- Normal
- Chi-squared
- Gamma
- Cauchy
- Laplace

Distribution	Characteristic Function $\varphi(t)$
Exponential $Exp(\lambda)$	$\left(1-it\lambda^{-1}\right)^{-1}$
Normal $N(\mu, \sigma^2)$	$e^{it\mu}-\frac{1}{2}\sigma^2t^2$

Distribution	Characteristic Function $\varphi(t)$
Chi-squared X^2_k	$(1-2it)^{-\frac{k}{2}}$
Gamma $\Gamma(k, \theta)$	$(1-it\theta)^{-k}$
Cauchy $Cauchy(\mu, \theta)$	$e^{it\mu-\theta\lvert t\rvert}$
Laplace $L(\mu, b)$	$\frac{e^{it\mu}}{1+b^2t^2}$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 99

Derivation of mean and variance of a distribution from its Characteristic Function

(by repeated differentiation of the Characteristic Function)

If X has a distribution with characteristic function $\varphi(t)$, then, if $E(X)$ and $E(X^2)$ exist, they are given respectively, by

$$iE(X) = \varphi'(0) \text{ and } i^2E(X^2) = \varphi''(0).$$

Example:

- Use the characteristic function of the Exponential distribution in order to find the mea and variance.

Solution:

CF of exponential distribution is given by

$$\varphi(t) = \frac{\lambda}{\lambda - it}$$

Now, taking the first derivative w.r.t 't', we get

$$\varphi'(t) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - it}\right) = \lambda \frac{d}{dt}(\lambda - it)^{-1} = \lambda(-1)(\lambda - it)^{-1-1} \frac{d}{dt}(\lambda - it)$$

$$= -\lambda(\lambda - it)^{-2} \frac{d}{dt}(0 - i(1)) = -\lambda(\lambda - it)^{-2}(-i) = \frac{i\lambda}{(\lambda - it)^2}$$

Now putting t=0, we have

$$\varphi'(0) = \frac{i\lambda}{(\lambda - i(0))^2} = \frac{i\lambda}{(\lambda - 0)^2} = \frac{i\lambda}{(\lambda)^2} = \frac{i}{\lambda}$$

Therefore,

$$\varphi'(0) = \frac{i}{\lambda} = i\left(\frac{1}{\lambda}\right) = iE(X).$$

Now, taking the second derivative w.r.t 't', we get

$$\varphi''(t) = \frac{d}{dt}\left(\frac{i\lambda}{(\lambda - it)^2}\right) = i\lambda \frac{d}{dt}(\lambda - it)^{-2} = \lambda(-2)(\lambda - it)^{-2-1} \frac{d}{dt}(\lambda - it)$$

$$= -2\lambda(\lambda - it)^{-3} \frac{d}{dt}(0 - i(1)) = \frac{2i^2\lambda}{(\lambda - it)^3} = \frac{-2\lambda}{(\lambda - it)^3} \qquad (\because i = \sqrt{-1} \Rightarrow i^2 = -1)$$

Now putting t=0, we have

$$\varphi''(0) = \frac{-2\lambda}{(\lambda - 0)^3} = \frac{-2\lambda}{(\lambda)^3} = \frac{-2}{\lambda^2} = \frac{-2}{\lambda^2}$$

Therefore,

$$\varphi'(0) = -1\left(\frac{2}{\lambda^2}\right) = i^2\left(\frac{2}{\lambda^2}\right) = i^2 E\left(X^2\right). \ (\because i = \sqrt{-1}, i^2 = -1)$$

As we know that,

$$V(X) = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

Therefore, putting the values in the above

$$\frac{1}{\lambda^2} = E\left(X^2\right) - \left[\frac{1}{\lambda}\right]^2 \Rightarrow E\left(X^2\right) = \frac{1}{\lambda^2} + \frac{1}{\lambda^2} = \frac{2}{\lambda^2}$$

For any distribution, if a higher moment exists, then all lower moments necessarily exists.

It is worthwhile to note that

$$\varphi(t) = M(it)$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 100

Concept of Conditional distribution

with the help of an example

Introduction

The conditional probability of even A, given event B, as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided $P(B) \neq 0$.

Suppose now that A and B are the events $X=x$ and $Y=y$ so that we can write

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

provided $P(Y=y) = h(y) \neq 0$, where $f(x , y)$ is the value of the **joint probability distribution** of X and Y at (x , y).

Definition

If $f(x, y)$ is the value of the joint probability distribution of the discrete random variables X and Y at (x, y), and $\boldsymbol{h(y)}$ is the value of the **marginal distribution** of Y of y, the function given by

$$f(x \mid y) = \frac{f(x, y)}{h(y)}, \quad h(y) \neq 0$$

for each x within the range of X, is called the conditional distribution of X given $Y=y$.

vulearninghub.xyz | by VU Learning HUB

Example:

Let $f_{1|2}\left(x_1 \mid x_2\right) = 2x_1 / x_2^2, 0 < x_1 < x_2, 0 < x_2 < 1$ zero elsewhere, and $f_2\left(x_2\right) = 5x_2^4, 0 < x_2 < 1,$ zero elsewhere, denote, respectively, the conditional pdf of X_1, given $X_2 = x_2$, and the marginal pdf of X_2.

Determine: $(d) \quad P\left(\frac{1}{4} < X_1 < \frac{1}{2}\right).$

Solution:

First we will find the marginal of x_1

$$f_{X_1}(x_1) = \int_{x_1}^{1} 10 x_1 x_2^2 dx_2 = 10 x_1 \int_{x_1}^{1} x_2^2 dx_2$$

$$= 10 x_1 \left(\frac{x_2^3}{3} \Big|_{x_1}^{1} = 10 x_1 \left(\frac{1}{3} - \frac{x_1^3}{3} \right) = \frac{10 x_1}{3} - \frac{10 x_1^4}{3} = \frac{10}{3} x_1 \left(1 - x_1^3\right) \right.$$

Now,

$$P\left(\frac{1}{4} < X_1 < \frac{1}{2}\right) = \frac{10}{3}\int_{1/4}^{1/2}\left(x_1 - x_1^4\right)dx_1 = \frac{10}{3}\left(\frac{x_1^2}{2}\Big|_{1/4}^{1/2} - \frac{x_1^5}{5}\Big|_{1/4}^{1/2}\right)$$

$$= \frac{10}{3}\left[\left(\frac{1}{4\times 2} - \frac{1}{16\times 2}\right) - \left(\frac{1}{32\times 5} - \frac{1}{1024\times 5}\right)\right]$$

$$P\left(\frac{1}{4} < X_1 < \frac{1}{2}\right) = \frac{10}{3}\left[\left(\frac{1}{8} - \frac{1}{32}\right) - \left(\frac{1}{160} - \frac{1}{5120}\right)\right]$$

$$= \frac{10}{3}\left[\left(\frac{4-1}{32}\right) - \left(\frac{32-1}{5120}\right)\right] = \frac{10}{3}\left[\frac{3}{32} - \frac{31}{5120}\right]$$

$$= \frac{10}{3}\left(\frac{480-31}{5120}\right) = \frac{4490}{3\times 5120} = \frac{449}{1536}.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 101

The Concept of Conditional Expectation, Conditional Mean and Conditional Variance (explained through an example)

If $u(X_2)$ is a function of X_2, the conditional expectation of $u(X_2)$, given that $X_1=x_1$, if it exists, is given by

$$E\left[u\left(X_2\right)|\, x_1\right]=\int_{-\infty}^{\infty} u\left(x_2\right) f_{2|1}\left(x_2\,|\,x_1\right) dx_2.$$

Note that $E[u(X_2)\,|\,x_1]$ is a function of x_1.

If they do it exists, then $E(X_2|\, x_1)$ is the mean

and

$$E\left\{\left[X_2 - E\left(X_2 \mid x_1\right)\right]^2 \mid x_1\right\},$$

is the variance of the conditional distribution of X_2 given $X_1=x_1$, which can be written more simply as $\mathrm{Var}(X_2 \mid x_1)$.

It is convenient to refer to these as the “conditional mean” and the “ conditional variance” of X_2, given X_1= x_1.

Also, we have

$$Var\left(X_2 \mid x_1\right) = E\left(X_2^2 \mid x_1\right) - \left[E\left(X_2 \mid x_1\right)\right]^2$$

from an earlier result.

The conditional expectation of $u(X_1)$, given $X_2 = x_2$, if it exists, is given by

$$E\left[u\left(X_1\right)|\, x_2\right] = \int_{-\infty}^{\infty} u\left(x_1\right) f_{1|2}\left(x_1 \,|\, x_2\right) dx_1.$$

In the case of when random variables of the discrete type, these conditional probabilities and conditional expectations are computed by using summation instead of integration

Example:

Let X_1 and X_2 have the joint pdf

$$f(x_1, x_2) = \begin{cases} 2 & 0 < x_1 < x_2 < 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

Then the marginal probability density functions are, respectively,

$$f_1(x_1) = \begin{cases} \int_{x_1}^{1} 2dx_2 = 2(1-x_1) & 0 < x_1 < 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

and

$$f_2(x_2) = \begin{cases} \int_{0}^{x_2} 2dx_1 = 2x_2 & 0 < x_2 < 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

The conditional pdf of X_1, given $X_2 = x_2$, $0 < x_2 < 1$, is

$$f_{1|2}\left(x_1 \mid x_2\right) = \begin{cases} \dfrac{2}{2x_2} = \dfrac{1}{x_2} & 0 < x_1 < x_2 < 1 \\ 0 & \textit{elsewhere.} \end{cases}$$

Here the conditional mean and the conditional variance of X_1 ,given $X_2 = x_2$, are respectively,

$$E\left[X_1 \mid x_2\right] = \int_{-\infty}^{\infty} x_1 f_{1|2}\left(x_1 \mid x_2\right) dx_1$$

$$= \int_0^{x_2} x_1 \left(\frac{1}{x_2}\right) dx_1 = \frac{x_2}{2}, \quad 0 < x_2 < 1,$$

and

$$Var\left(X_1 \mid x_2\right) = \int_0^{x_2} \left(x_1 - \frac{x_2}{2}\right)^2 \left(\frac{1}{x_2}\right) dx_1$$

$$= \frac{x_2^2}{12}, \quad 0 < x_2 < 1.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 102

Rigorous proof of the theorem containing the results

$$E[E(X_2|X_1)]=E(X_2)$$

Theorem:

Let (X_1, X_2) be a random vector such that $E(X_2)$is finite.

Then,

$$E[E(X_2|X_1)]= E(X_2).$$

Proof:

$$E(X_2) = \int_{-\infty}^{\infty} x_2 h(x_2)\, dx_2$$

But

$$h(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$$

Therefore, $E(X_2)$ can be written as

$$E(X_2) = \int_{-\infty}^{\infty} x_2 \left(\int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 \right) dx_2$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f(x_1, x_2)\, dx_1 dx_2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_2 f(x_1, x_2)\, dx_2 dx_1$$

Multiplying and dividing by $f_1(x_1)$, $E\left(X_2\right)$ can be written as

$$E\left(X_2\right)=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}x_2\,\frac{f\left(x_1,x_2\right)}{f_1\left(x_1\right)}\,dx_2\right]f_1\left(x_1\right)dx_1$$

But $\int_{-\infty}^{\infty}x_2\,\frac{f\left(x_1,x_2\right)}{f_1\left(x_1\right)}\,dx_2=E\left(X_2\,|\,x_1\right)$

Therefore

$$E\left(X_2\right)=\int_{-\infty}^{\infty}E\left(X_2\,|\,x_1\right)f_1\left(x_1\right)dx_1=E\left[E\left(X_2\,|\,x_1\right)\right].$$

vulearninghub.xyz | by VU Learning HUB

Hence, we can write

$$E\left(X_2\right) = E\left[E\left(X_2 \mid X_1\right)\right].$$

This is sometimes called **the law of total expectation** and sometimes it is called **the law of iterated expectations**.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 103

The Correlation Coefficient

(explained through an example)

- Let X and Y have the joint pdf $f(x, y)$.

- If $u(x, y)$ is a function of x and y, then $E[u(X,Y)]$ can be determined, if it exists.

- Here, let us assume that all mathematical expectations **exist.**

The means of X and Y, say μ_1 and μ_2, are obtained by taking $u(x, y)$ to be x and y, respectively;

i.e.

$\mu_1 = E[u(X,Y)] = E[X]$

&

$\mu_2 = E[u(X,Y)] = E[Y]$

and

the variances of X and Y, say σ_1^2 and σ^2_2, are obtained by setting the function $u(X,Y)$ equal to $(x - \mu_1)^2$ and $(x - \mu_2)^2$, respectively.

Covariance

- Consider the mathematical expectation

$$E\left[\left(X-\mu_1\right)\left(Y-\mu_2\right)\right]=E\left(XY-\mu_2X-\mu_1Y+\mu_1\mu_2\right)$$
$$=E\left(XY\right)-\mu_2E\left(X\right)-\mu_1E\left(Y\right)+\mu_1\mu_2$$
$$=E\left(XY\right)-\mu_1\mu_2$$

- This number is called the **covariance** of X and Y and is often denoted by $\text{cov}(X,Y)$.

Correlation

If each of σ_1 and σ_2 is positive, the number

$$\rho = \frac{E\left[\left(X-\mu_1\right)\left(X-\mu_2\right)\right]}{\sigma_1\sigma_2} = \frac{\operatorname{cov}\left(X,Y\right)}{\sigma_1\sigma_2}$$

is called the correlation coefficient of X and Y.

Example:

Let the random variables X and Y have the joint pdf

$$f(x,y)=\begin{cases} x+y & 0<x<1, 0<y<1 \\ 0 & \textit{elsewhere.} \end{cases}$$

We next compute the correlation coefficient ρ of X and Y.

vulearninghub.xyz | by Vu Learning HUB

Now

$$\mu_1 = E(X) = \int_0^1 \int_0^1 x(x+y)dxdy = \frac{7}{12}$$

&

$$\sigma_1^2 = E(X^2) - \mu_1^2 = \int_0^1 \int_0^1 x^2(x+y)\,dxdy - \left(\frac{7}{12}\right)^2 = \frac{11}{144}.$$

Similarly,

$$\mu_2 = E(Y) = \frac{7}{12} \quad and \quad \sigma_2^2 = E(Y^2) - \mu_2^2 = \frac{11}{144}.$$

Therefore, the covariance of X and Y is

$$E(XY) - \mu_1\mu_2 = \int_0^1 \int_0^1 xy(x+y)\,dxdy - \left(\frac{7}{12}\right)^2 = -\frac{11}{144}.$$

Accordingly, the correlation coefficient of X and Y is

$$\rho = \frac{-\frac{1}{144}}{\sqrt{\left(\frac{11}{144}\right)\left(\frac{11}{144}\right)}} = -\frac{1}{11} = -0.09$$

Implying that their exists a very weak negative linear correlation between X and Y.

vulearninghub.xyz by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 104

Properties

of the

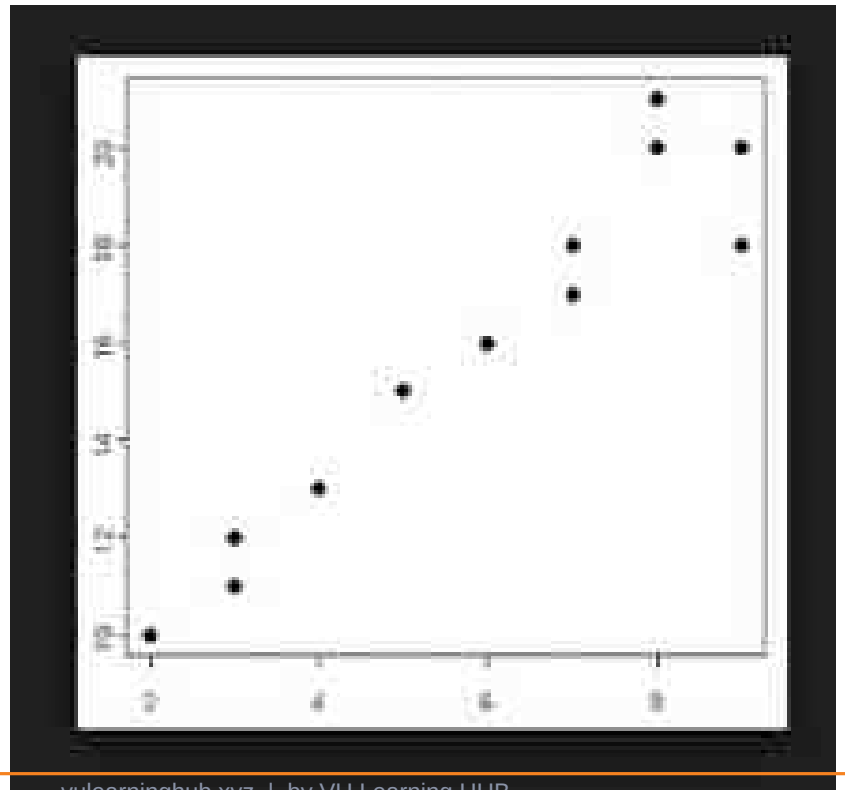
Correlation Coefficient

1. Coefficient of Correlation lies between -1 and +1:

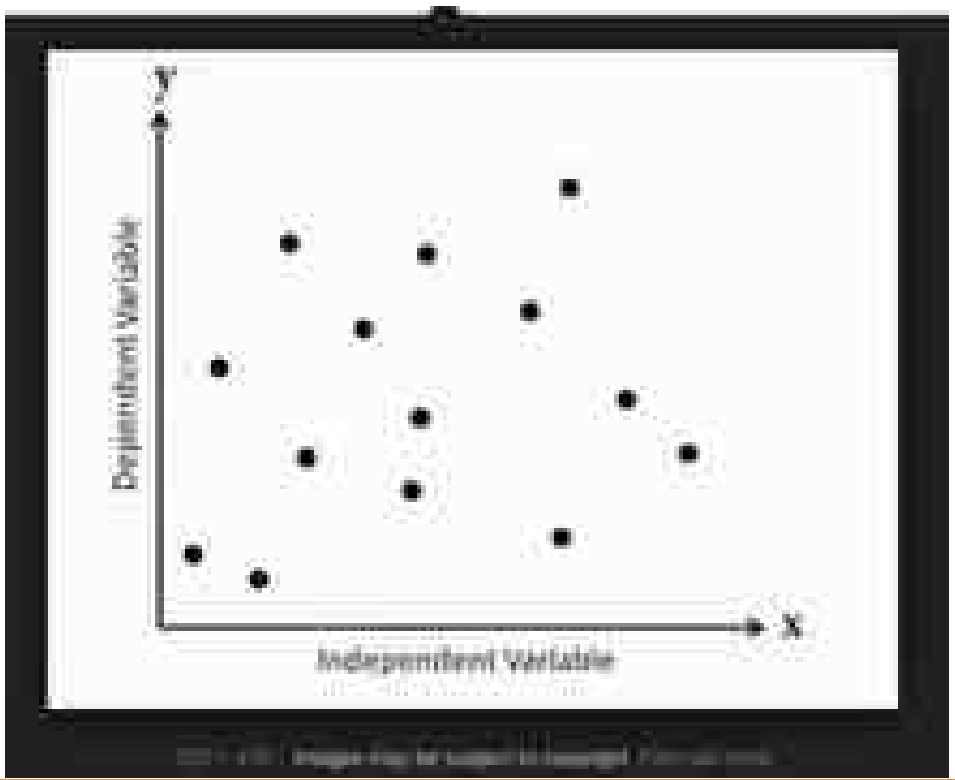
The coefficient of correlation cannot take value less than -1 or more than one +1. Symbolically,

$$-1 \leq \rho \leq +1 \text{ or } |\rho| < 1.$$

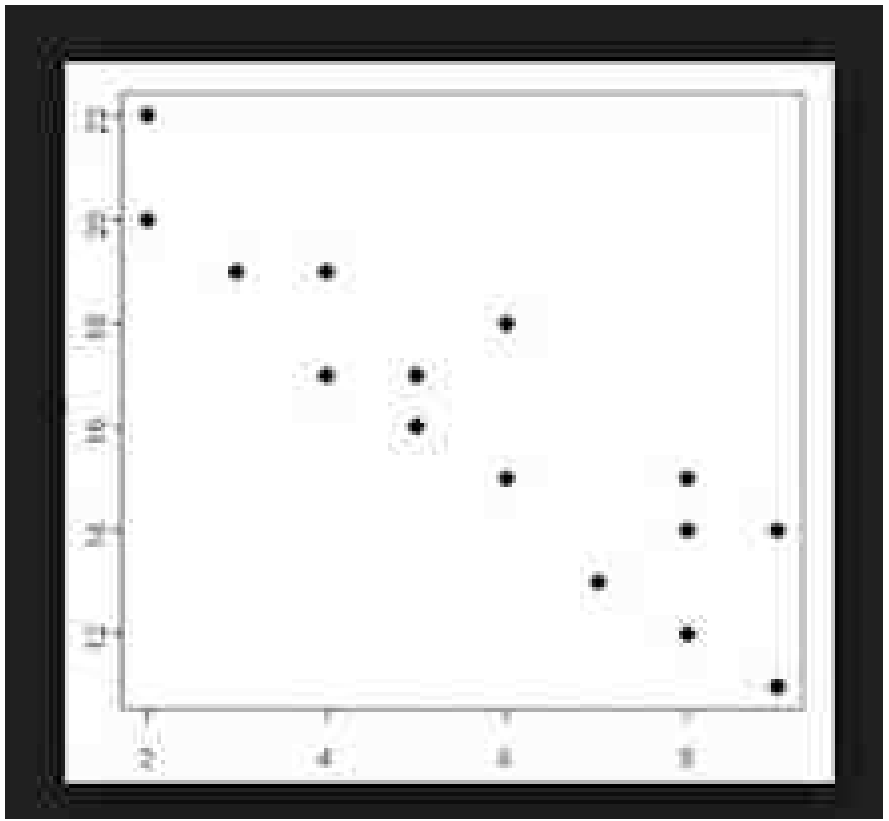
‘Upward-going’ scatter-diagram: $0 < \rho \leq 1$



Neither upward nor downward scatter-diagram: ρ =0



‘Downward-going’ scatter-diagram: $-1 \leq \rho < 0$



Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 105

Properties

of the

Correlation Coefficient

1. Coefficients of Correlation are independent of Change of Origin

2. Coefficient of Correlation is independent of Change of Scale

2. Coefficients of Correlation are independent of Change of Origin:

This property reveals that

if we subtract any constant from all the values of X (or add any constant to all the values of X)

and

subtract any constant from all the values of Y (or add any constant to all the values of X),

it will not affect the coefficient of correlation.

In other words,

$$\rho_{D_X D_Y} = \rho_{XY}$$

where

$$D_X = X - \mathrm{A} \;\; \& \;\; D_Y = Y - \mathrm{B}$$

2. Coefficient of Correlation is independent of Change of Scale:

This property reveals that if we divide or multiply all the values of X by any constant and divide or multiply all the values of Y by any constant, it will not affect the coefficient of correlation. $\rho_{UV} = \rho_{XY}$

where

$$U = \frac{X}{A} \quad \& \quad V = \frac{Y}{B}$$

vulearninghub.xyz (by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 106

Properties

of the

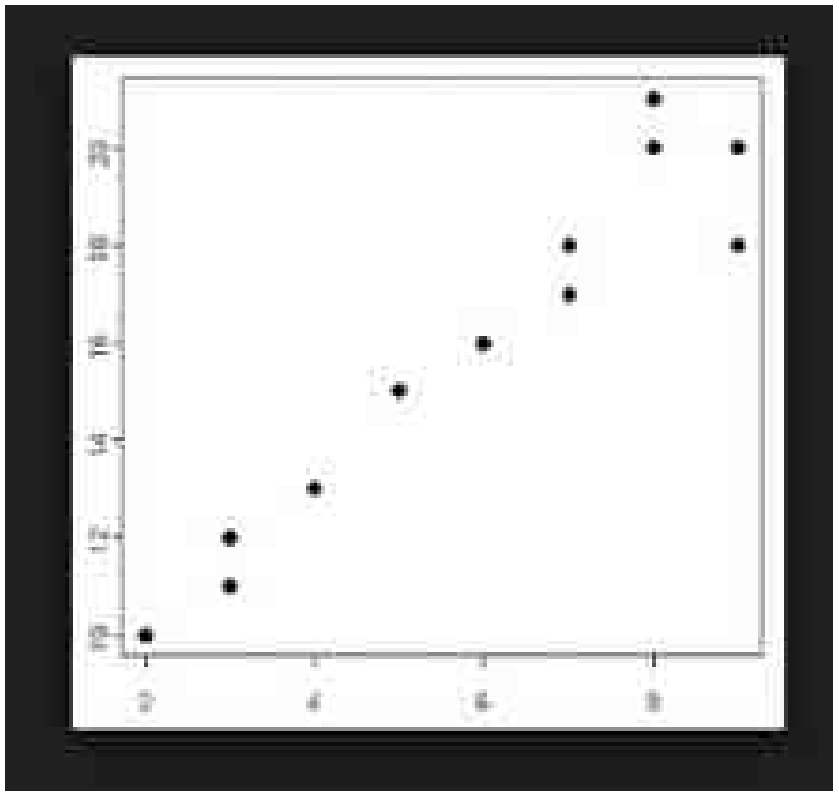
Correlation Coefficient

4. Coefficient of Correlation possess the property of symmetry

$$\rho_{XY} = \rho_{YX}$$

5. Co-efficient of correlation measures only **linear** correlation between X and Y.

‘Upward-going’ scatter-diagram: $0 < \rho \leq 1$



6. If two variables X and Y are **independent**, coefficient of correlation between them will be **zero**.

Topic No. 107

vulearninghub.xyz | By VU Learning HUB

Scanned with CamScanner

Scanned with CamScanner

Computation of Correlation Coefficient

- "Complicated" domain:

In this regard, the first thing to note is that $0 < x < y, 0 < y < 1$

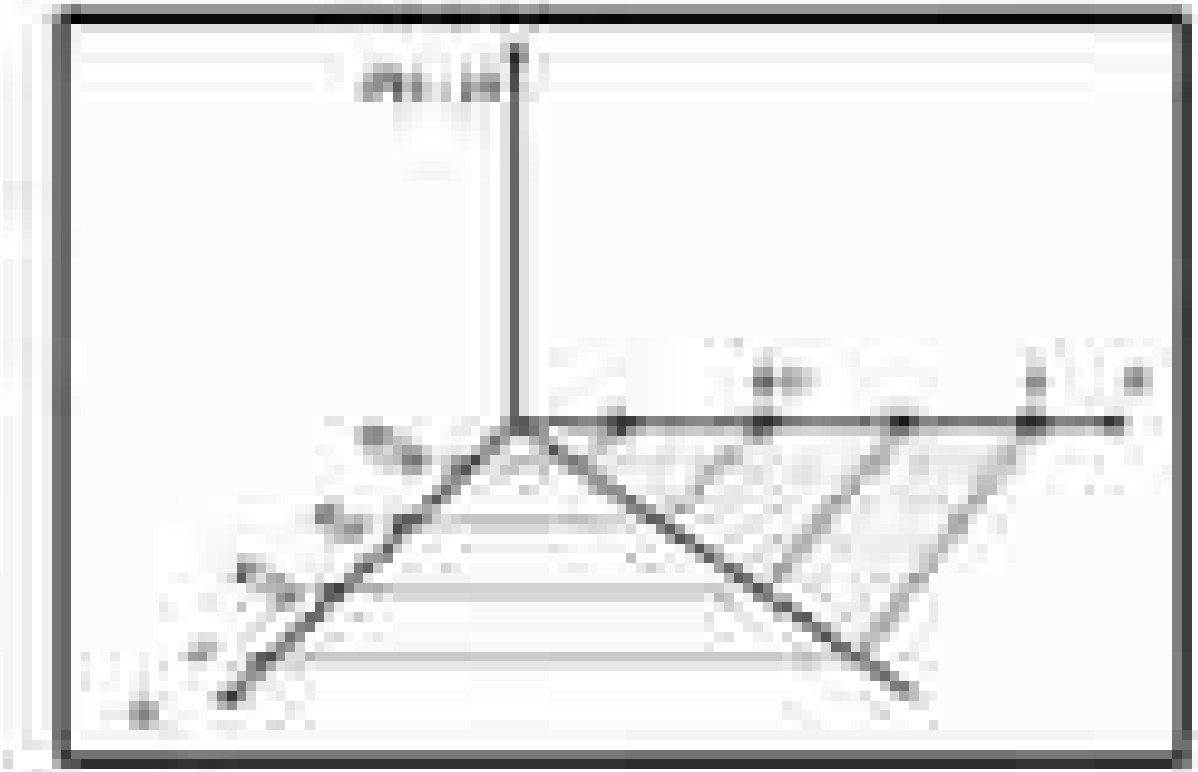
can be re-written as

$$0 < x < y < 1$$

and $0 < x < y < 1$ can be re-written as $0 < x < 1, x < y < 1$

Scanned with CamScanner

Computation of Correlation Coefficient



Scanned with CamScanner

Computation of Correlation Coefficient

First and foremost, we need to find each of the two marginal distributions:

Scanned with CamScanner

Computation of Correlation Coefficient

vulearninghub.xyz | by VU Learning HUB

Scanned with CamScanner

Computation of Correlation Coefficient

As such:

Marginal distribution of X:

$$f_X(x) = \int_x^1 2dy = 2[y]_x^1 = 2(1-x), \quad 0 < x < 1$$

and

Marginal distribution of Y:

$$f_Y(y) = \int_0^y 2dx = 2[x]_0^y = 2(y-0) = 2y, \quad 0 < y < 1$$

Scanned with CamScanner

Computation of Correlation Coefficient

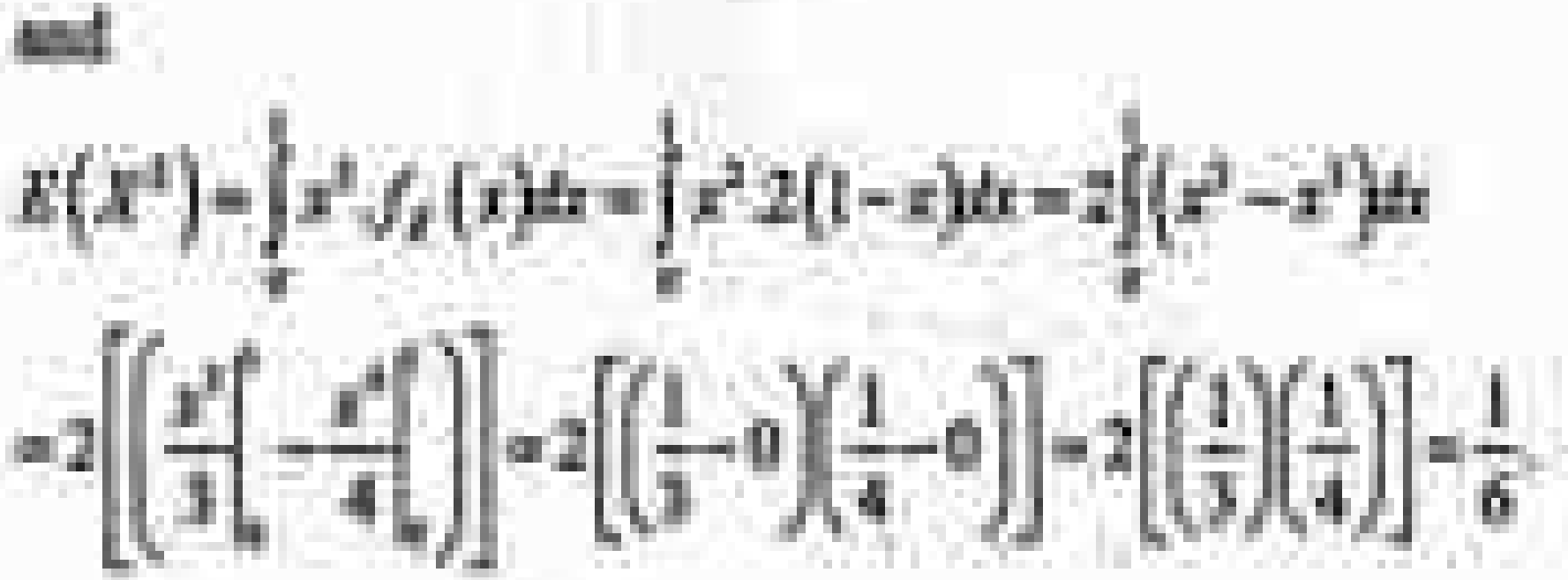
Now

$$E(XY) = \int_0^1 x f_3(x)\,dx = \int_0^1 x\,2(1-x)\,dx = 2\int_0^1 (x-x^2)\,dx$$

$$= 2\left[\left(\frac{x^2}{2}\right)_0^1 - \left(\frac{x^3}{3}\right)_0^1\right] = 2\left[\left(\frac{1}{2}-0\right)\left(\frac{1}{3}-0\right)\right] = 2\left[\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)\right] = \frac{1}{3}$$

Scanned with CamScanner

Computation of Correlation Coefficient



Scanned with CamScanner

Computation of Correlation Coefficient

Therefore,

$$\text{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

$$= \frac{1}{6} - \left(\frac{1}{3}\right)^2 = \frac{1}{6} - \frac{1}{9} = \frac{9-6}{54} = \frac{3}{54} = \frac{1}{18}$$

Scanned with CamScanner

Computation of Correlation Coefficient

Now, the product moment $E(XY)$:

$$E(XY) = \int_0^1 \left(\int_0^x xy\, f(x,y)\, dx \right) dy = \int_0^1 \int_0^x xy(2)\, dx\, dy = 2 \int_0^1 \int_0^x xy\, dx\, dy$$

$$= 2 \int_0^1 y \left(\frac{x^2}{2} \Big|_0^x \right) dx = 2 \int_0^1 y \left(\frac{y^2}{2} - 0 \right) dx = \int_0^1 y^3\, dx = \frac{y^4}{4} \Big|_0^1 = \frac{1}{4} - 0 = \frac{1}{4}$$

Scanned with CamScanner

Computation of Correlation Coefficient

Hence

$$\text{Cov}(X,Y) = E(XY) - E(X)E(Y)$$

$$= \frac{1}{4} - \frac{1}{3}\cdot\frac{2}{3} = \frac{1}{4} - \frac{2}{9} = \frac{9-8}{36} = \frac{1}{36}$$

Scanned with CamScanner

Computation of Correlation Coefficient

Therefore,

$\rho_{X,Y}$ the correlation coefficient of X and Y,

is given by:

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{\frac{1}{36}}{\sqrt{\left(\frac{1}{18}\right)\left(\frac{1}{18}\right)}} = \frac{\frac{1}{36}}{\sqrt{\left(\frac{1}{18}\right)^2}} = \frac{\frac{1}{36}}{\frac{1}{18}} = \frac{1/2}{1} = \frac{1}{2}$$

Scanned with CamScanner

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

vulearninghub.xyz | by VU Learning HUB

Topic No. 108

Example of computation of the variance of a conditional distribution in the case of joint PDF of two random variables

Example:

Given the pdf

$$f(x,y) = \begin{cases} 2, & 0 < x < y,\ 0 < y < 1, \\ zero & elsewhere \end{cases}$$

Show that the **variance** of the conditional distribution of Y, given $X=x$, is $(1-x)^2/12$, $0 < x < 1$,

and that the **variance** of the conditional distribution of X, given $Y = y$, is $y^2/12$, $0 < y < 1$.

1. “Complicated” domain:

In this regard, the first thing to note is that $0 < x < y,\ 0\ < y < 1$

can be re-written as

$$0 < x < y < 1$$

and that $0 < x < y < 1$ can be re-written as $0 < x < 1,\ x < y < 1$

First and foremost, we need to find each of the two **marginal** distributions:

Because of the nature of the support of the bivariate density function,
when finding **marginal** distributions:

the limits of the **integral** with respect to ***x*** are to be **0 to *y***
(and this will give us the marginal distribution of y i.e. $f_Y(y)$)

and

the limits of the **integral** with respect to ***y*** are to be ***x*** **to 1.**
(and this will give us the marginal distribution of x i.e. $f_X(x)$)

As such:

Marginal distribution of X:

$$f_X(x) = \int_x^1 2dy = 2.[y]_x^1 = 2(1-x),\ 0 < x < 1$$

and

Marginal distribution of Y:

$$f_Y(y) = \int_0^y 2dx = 2.[x]_0^y = 2(y-0) = 2y,\ 0 < y < 1$$

Then, the conditional distributions are as follows:

$$f\left(x \mid y\right) = \frac{f\left(x,y\right)}{f_Y\left(y\right)} = \frac{2}{2y} = \frac{1}{y}, \;\; 0 < x < y$$

and

$$f\left(y \mid x\right) = \frac{f\left(x,y\right)}{f_X\left(x\right)} = \frac{2}{2\left(1-x\right)} = \frac{1}{1-x}, \;\; x < y < 1$$

vulearninghub.xyz | by VU Learning HUB

(Note the domains (i.e. supports) of the two distributions.)

Now, since

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{2}{2y} = \frac{1}{y}, \quad 0 < x < y$$

Therefore the conditional mean of X given $Y = y$ is

$$E(X|Y = y) = \int_0^y x.f(x|y)dx = \int_0^y x.\frac{1}{y}dx = \frac{1}{y}\left(\frac{x^2}{2}\Big|_0^y\right) = \frac{1}{y}\left(\frac{y^2}{2} - 0\right)$$

$$= \frac{1}{y}.\frac{y^2}{2} = \frac{y}{2}, \quad 0 < y < 1$$

Similarly

$$E\left(X^2 \mid y\right) = \int_0^y x^2 . f\left(x \mid y\right) dx = \int_0^y x^2 . \frac{1}{y} dx = \frac{1}{y}\left(\frac{x^3}{3} \right|_0^y$$

$$= \frac{1}{y}\left(\frac{y^3}{3} - 0 \right) = \frac{1}{y} . \frac{y^3}{3} = \frac{y^2}{3}, \quad 0 < y < 1$$

Therefore,

$$\mathrm{Var}(X \mid y) = E(X^2 \mid y) - [E(X \mid y)]^2 = \frac{y^2}{3} - \left(\frac{y}{2}\right)^2$$

$$= \frac{y^2}{3} - \frac{y^2}{4} = \frac{4y^2 - 3y^2}{12} = \frac{y^2}{12}, \quad 0 < y < 1.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 109

The Conditional Mean of Y given X that is linear in X

Let $f(x,y)$ denote the joint pdf of two random variables X and Y and let $f_1(x)$ denote the marginal pdf of X.

So, the **conditional pdf** of Y, given $X=x$, is given by

$$f_{2|1}(y \mid x) = \frac{f(x,y)}{f_1(x)}$$

at the points where $f_1(x) > 0$;

vulearninghub.xyz | by VU Learning HUB

and the **conditional mean** of Y, given $X = x$, is given by

$$E(Y \mid x) = \int_{-\infty}^{\infty} y f_{2|1}(y \mid x)\, dy = \int_{-\infty}^{\infty} \frac{y f(x, y)}{f_1(x)} dy = \frac{\int_{-\infty}^{\infty} y f(x, y)\, dy}{f_1(x)}$$

This conditional mean of Y, given $X=x$, *i.e.* $E(Y|X)$ is, of course,

a **function** of x, say $u(x)$.

Why ?

$$\frac{\int_{-\infty}^{\infty} yf(x,y)\,dy}{f_1(x)}.$$

Have **a close look** at the expression

In case $u(x)$ is a **linear** function of x,

say $$u(x) = a + bx$$

Then we can say that **the conditional mean of *Y* given *x* is linear in *x*;**

or that $E(Y|x)$ is **a linear conditional mean.**

If $E(Y|X)$ is linear in X then there are some **theoretical results** pertaining to $E(Y|X)$ and $\mathrm{E}(\mathrm{Var}(Y|X))$:

Theorem:

Suppose (X,Y) are a joint distribution with the variances of X and Y **finite** and **positive.**

Denote the mean and variances of X and Y by μ_1,μ_2 and σ_1^2,σ_2^2, respectively,

and let ρ be the correlation coefficient between X and Y.

If $E(Y|X)$ is linear in X **then**

$$E\left(Y \mid X\right) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}\left(X - \mu_1\right) \qquad (1)$$

and

$$E\left(\mathrm{Var}\left(Y \mid X\right)\right) = \sigma_2^2\left(1 - \rho^2\right) \qquad (2)$$

and

$$E\left(\mathrm{Var}\left(Y \mid X\right)\right) = \sigma_2^2\left(1-\rho^2\right) \qquad (2)$$

The Conditional Variance can be interpreted as a Random Variable:

Actually, both $E(Y|X)$ and $Var(Y|X)$ can be considered as random variables.

Let us consider an **example**:

Suppose we are interested in the mean heights of people of various races and also in the variances of the heights of people of various races.

(Race *is a socially meaningful category of people who share biologically transmitted traits that are obvious and considered important.)*

If Y = ***height*** and X = ***race*** for persons in a certain population,

then

E($Y \mid X$) =E(*height* | *race*) is the **variable** which assigns to each person in the population the mean height for that person's race.

and

Var($Y \mid X$) =Var(*height* | *race*) is the **variable** which assigns to each person in the population the variance of height for that person's race.

Then, obviously,

E[E($Y \mid X$)] =E[E(*height* | *race*)] is the **expected value** of the variable E($Y \mid X$) i.e. the **expected value** of **the mean heights of various races**.

and

E[Var($Y \mid X$)] =E[Var(*height* | *race*)] is the expected value of the variable

Var($Y \mid X$) i.e. the **expected value** of **the variances of the heights of various races.**

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 110

The Concept of Independent Random Variables (explained through an example)

Let X_1 and X_2 denote the random variables of the continuous type which have the joint pdf $f(x_1,x_2)$ and marginal probability density functions $f_1(x_1)$ and $f_2(x_2)$, respectively.

Then, in accordance with the definition of the conditional pdf $f_{2|1}(x_2|x_1)$, we many write the joint pdf $f(x_1,x_2)$ as

$$f\left(x_1, x_2\right) = f_1\left(x_1\right) f_{2|1}\left(x_2 \mid x_1\right) \qquad (1)$$

Suppose that we have an instance where $f_{2|1}(x_2|x_1)$ **does not** depend upon x_1.

Then, for random variables of the **continuous** type, the marginal pdf of X_2 is given by :

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1 = \int_{-\infty}^{\infty} f_{2|1}(x_2 \mid x_1)\, f_1(x_1)\, dx_1$$

$$= f_{2|1}(x_2 \mid x_1) \int_{-\infty}^{\infty} f_1(x_1)\, dx_1 = f_{2|1}(x_2 \mid x_1).$$

Accordingly,

$$f_2\left(x_2\right) = f_{2|1}\left(x_2 \mid x_1\right)$$

And hence, eq.(1) can be re-written as

$$f\left(x_1, x_2\right) = f_1\left(x_1\right) f_2\left(x_2\right)$$

(when $f_{2|1}(x_2|x_1)$ does not depend upon x_1).

In other words, if the conditional distribution of X_2, given $X_1 = x_1$, is **independent** of any assumption about x_1, then

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

Formal Definition of Independence

in the case of

Discrete Random Variables

Let the random variables X_1 and X_2 have the joint pmf $p(x_1,x_2)$

and the marginal pmfs $p_1(x_1)$ and $p_2(x_2)$ respectively.

The random variables X_1 and X_2 are said to be independent **if and only if**

$$p(x_1,x_2) = p_1(x_1)p_1(x_2).$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 111

Example related to the Concept of Independent Random Variables

Example:

Let the joint pdf of X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \textit{elsewhere} \end{cases}$$

Show that X_1 and X_2 are dependent.

Here the marginal probability density functions are obtained as follows:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 = \int_0^1 (x_1 + x_2)\, dx_2$$

$$= \left[x_1 x_2 + \frac{x_2^2}{2} \right]_0^1 = x_1 + \frac{1}{2} - 0 - 0 = x_1 + \frac{1}{2}$$

so that

$$f_1(x_1) = x_1 + \frac{1}{2}, \quad 0 < x < 1$$

and

$$f_2(x_2) = \int_0^1 (x_1 + x_2)\, dx_1 = \left[\frac{x_1^2}{2} + x_1 x_2 \right]_0^1$$

$$= \frac{1}{2} + x_2 - 0 - 0 = \frac{1}{2} + x_2$$

so that

$$f_2(x_2) = x_2 + \frac{1}{2}, \quad 0 < x_2 < 1.$$

Now

$$f_1(x_1)\,f_2(x_2) = \left(x_1 + \frac{1}{2}\right)\left(x_2 + \frac{1}{2}\right) = x_1^2 + \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{4}$$

whereas

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

Since $f(x_1, x_2) \neq f_1(x_1)\,f_2(x_2)$, we can say that the random variables X_1 and X_2 are dependent (i.e. **not** independent).

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 112

Theorem stating that

X and Y are independent

if and only if

$$F(x, y) \equiv F_X(x) F_Y(y)$$

Theorem:

If X and Y are random variables with joint distribution function F and marginal distribution functions F_X and F_Y, then X and Y are independent if and only if

$$F(x , y) = F_X(x)F_Y(y)$$

For all $(x , y) \in R^2$.

Proof:

If X and Y are independent, then

$$F(x,y) = P(X \le x, Y \le y) = P(X \le x)P(Y \le y) = F_X(x)F_Y(y).$$

Now suppose

$$F(x,y) = F_X(x)F_Y(y) \text{ for all } (x,y) \in \mathbb{R}^2 \text{ and let } A = (a,b]$$

$$\text{and } B = (c,d].$$

$$P\left(X \in A, Y \in B\right) = P\left(a < X \leq b, c < Y \leq d\right)$$

$$= F\left(b,d\right) - F\left(a,d\right) - F\left(b,c\right) + F\left(a,c\right)$$

$$= F_X\left(b\right)F_Y\left(d\right) - F_X\left(a\right)F_Y\left(d\right) - F_X\left(b\right)F_Y\left(c\right) + F_X\left(a\right)F_Y\left(c\right)$$

$$P\left(X \in A, Y \in B\right) = \left(F_X\left(b\right) - F_X\left(a\right)\right)\left(F_Y\left(d\right) - F_Y\left(c\right)\right)$$

$$= P\left(a < X \leq b\right) P\left(c < Y \leq d\right)$$

$$= P\left(X \in A\right) P\left(Y \in B\right).$$

It may now be shown that $P(X \epsilon A, Y \epsilon B) = P(X \epsilon A)P(Y \epsilon B)$ for any sets $A \square \mathbb{R}$ and $B \square \mathbb{R}$. Thus X and Y are independent.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 113

Theorem stating that

if X_1 and X_2 are independent then,

$$E[u(X_1)v(X_2)] = E[u(X_1)]\,E[v(X_2)]$$

provided that the expectations exist.

Theorem:

Suppose X_1 and X_2 are independent and that $E(u(X_1))$ and $E(v(X_2))$ exist.

Then

$$E\left[u\left(X_1\right)v\left(X_2\right)\right]=E\left[u\left(X_1\right)\right]E\left[v\left(X_2\right)\right].$$

Proof:

The independence of X_1 and X_2 implies that the joint pdf of X_1 and X_2 is $f_1(x_1)\, f_2(x_2)$.

Thus we have, by definition of expectation,

$$E\left[u(X_1)v(X_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)\, f(x_1, x_2)\, dx_1 dx_2$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)\, f_1(x_1)\, f_2(x_2)\, dx_1 dx_2$$

Now,

$$E\left[u(X_1)v(X_2)\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} u(x_1)v(x_2)\, f_1(x_1)\, f_2(x_2)\, dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} v(x_2)\, f_2(x_2) \int_{-\infty}^{\infty} u(x_1) f_1(x_1)\, dx_1 dx_2$$

$$= \left[\int_{-\infty}^{\infty} u(x_1)\, f_1(x_1)\, dx_1\right]\left[\int_{-\infty}^{\infty} v(x_2)\, f_2(x_2)\, dx_2\right]$$

$$= E\left[u(X_1)\right]E\left[v(X_2)\right].$$

Hence proved.

Upon taking the functions u(.) and v(.) to be the identity functions in in the theorem we note that for independent random variables X_1 and X_2,

$$E(X_1X_2) = E(X_1)E(X_2). \qquad (1)$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 114

Theorem stating that X_1 and X_2 are independent if and only if

the joint MGF is identically equal to

the product of the marginal MGFs

i.e.

$$M(t_1,t_2) = M(t_1,0)\, M(0,t_2)$$

Theorem:

Suppose the joint mgf, $M(t_1,t_2)$, exists for the random variables X_1 and X_2.

Then, X_1 and X_2 are independent if and only if

$$M\left(t_1,t_2\right)=M\left(t_1,0\right)M\left(0,t_2\right);$$

That is, the joint mgf is identically equal to the product of the marginal mgfs.

Proof:

1. Suppose that X_1 and X_2 are independent.

The joint MGF of X_1 and X_2 is given by

$$M(t_1, t_2) = E\left(e^{t_1 X_1 + t_2 X_2}\right) = E\left(e^{t_1 X_1} e^{t_2 X_2}\right)$$

However, we already know that, if X_1 and X_2 are independent, then, for any functions $u(X_1)$ of X_1 and $v(X_2)$ of X_2 respectively, we have

$$E\left[u(X_1)v(X_2)\right] = E\left[u(X_1)\right]E\left[v(X_2)\right].$$

Therefore, if X_1 and X_2 are independent, then we can write

$$M(t_1, t_2) = E\left(e^{t_1 X_1 + t_2 X_2}\right) = E\left(e^{t_1 X_1} e^{t_2 X_2}\right)$$
$$= E\left(e^{t_1 X_1}\right) E\left(e^{t_2 X_2}\right)$$
$$= E\left(e^{t_1 X_1 + 0 X_2}\right) E\left(e^{0 X_1 + t_2 X_2}\right)$$
$$= M(t_1, 0) M(0, t_2).$$

Thus, the independence X_1 and X_2 implies that mgf of the joint distribution factors into the product of moment-generating functions of the two marginal distributions.

2. Suppose next that the mgf of the joint distribution of X_1 and X_2 is given by

$$M(t_1, t_2) = M(t_1, 0) M(0, t_2);$$

Now X_1 has a unique mgf, which, in the continuous case is given by

$$M(t_1, 0) = \int_{-\infty}^{\infty} e^{t_1 x_1} f_1(x_1) dx_1.$$

Similarly, a unique mgf of X_2, in the continuous case, is given by

$$M(0, t_2) = \int_{-\infty}^{\infty} e^{t_2 x_2} f_2(x_2) dx_2.$$

Thus we have

$$M(t_1,0)M(0,t_2)=\left[\int_{-\infty}^{\infty} e^{t_1x_1} f_1(x_1)dx_1\right]\left[\int_{-\infty}^{\infty} e^{t_2x_2} f_2(x_2)dx_2\right]$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1}e^{t_2x_2} f_1(x_1) f_2(x_2)dx_1dx_2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{t_1x_1+t_2x_2} f_1(x_1) f_2(x_2)dx_1dx_2.$$

We are given that,

$$M\left(t_1,t_2\right)=M\left(t_1,0\right)M\left(0,t_2\right);$$

Thus

$$M\left(t_1,t_2\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x_1+t_2x_2}\,f_1\left(x_1\right)f_2\left(x_2\right)dx_1dx_2 \qquad (1)$$

But, by definition, $M(t_1,t_2)$ is the mgf of X_1 and X_2

i.e.

$$M\left(t_1,t_2\right)=E\left(e^{t_1X_1+t_2X_2}\right)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{t_1x_1+t_2x_2}\,f\left(x_1,x_2\right)dx_1dx_2 \qquad (2)$$

The uniqueness of the mgf implies that the two distributions of probability that are described by $f_1(x_1)\,f_2(x_2)$ and $f(x_1, x_2)$ are the same.

Thus

$$f(x_1, x_2) \equiv f_1(x_1)\,f_2(x_2).$$

That is, if $M(t_1, t_2) = M(t_1, 0)M(0, t_2)$, then X_1 and X_2 are independent.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 115

Discrete Uniform Distribution

The **discrete uniform distribution** is a **symmetric** probability distribution whereby a finite number of values are **equally likely** to be observed;

In other words, if the discrete random variable X can assume n distinct values, and every one of the n values has **equal probability** $1/n$, we are dealing with a discrete uniform distribution.

Example 1:

Let X denote the number of heads that we can have if a fair coin is tossed once. Then, we have

X	$p(x)$
0	½
1	½

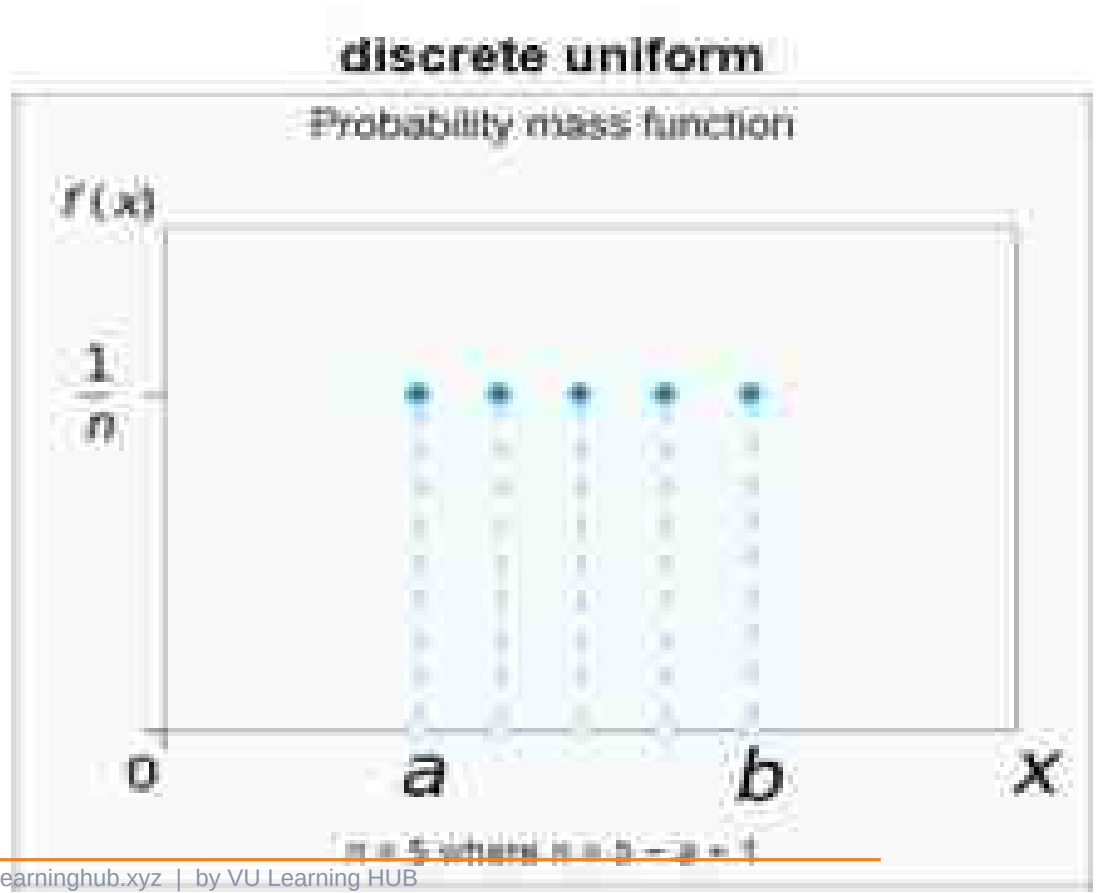
Example 2:

Let X denote the number that we observe on the uppermost face if we roll a fair die once.

Then, we have

X	$p(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

SHAPE OF THE DISTRIBUTION



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 116

Example related to the Concept of Independent Random Variables

Example:

Let the joint pdf of X_1 and X_2 be

$$f(x_1, x_2) = \begin{cases} x_1 + x_2 & 0 < x_1 < 1, 0 < x_2 < 1 \\ 0 & \textit{elsewhere} \end{cases}$$

Show that X_1 and X_2 are dependent.

Here the marginal probability density functions are obtained as follows:

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 = \int_0^1 (x_1 + x_2)\, dx_2$$

$$= \left[x_1 x_2 + \frac{x_2^2}{2} \right]_0^1 = x_1 + \frac{1}{2} - 0 - 0 = x_1 + \frac{1}{2}$$

so that

$$f_1(x_1) = x_1 + \frac{1}{2}, \quad 0 < x < 1$$

vulearninghub.xyz | by VU Learning HUB

and

$$f_2(x_2) = \int_0^1 (x_1 + x_2)\, dx_1 = \left[\frac{x_1^2}{2} + x_1 x_2 \right]_0^1$$

$$= \frac{1}{2} + x_2 - 0 - 0 = \frac{1}{2} + x_2$$

so that

$$f_2(x_2) = x_2 + \frac{1}{2}, \quad 0 < x_2 < 1.$$

Now

$$f_1(x_1)\,f_2(x_2) = \left(x_1 + \frac{1}{2}\right)\left(x_2 + \frac{1}{2}\right) = x_1^2 + \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{4}$$

whereas

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

Since $f(x_1, x_2) \neq f_1(x_1)\,f_2(x_2)$, we can say that the random variables X_1 and X_2 are dependent (i.e. **not** independent).

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 117

CDF

of the

Discrete Uniform Distribution

Example :

Suppose that we are rolling one fair die one time and we want to have the cdf of the distribution in the mathematic form, algebraic form and want to draw its graph.

Cumulative Distribution Function:

Step 1: Find the cumulative probabilities:

X	$p(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Cumulative Distribution Function:

Step 1: Find the cumulative probabilities:

X	*p(x)*	*cumulative*
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6

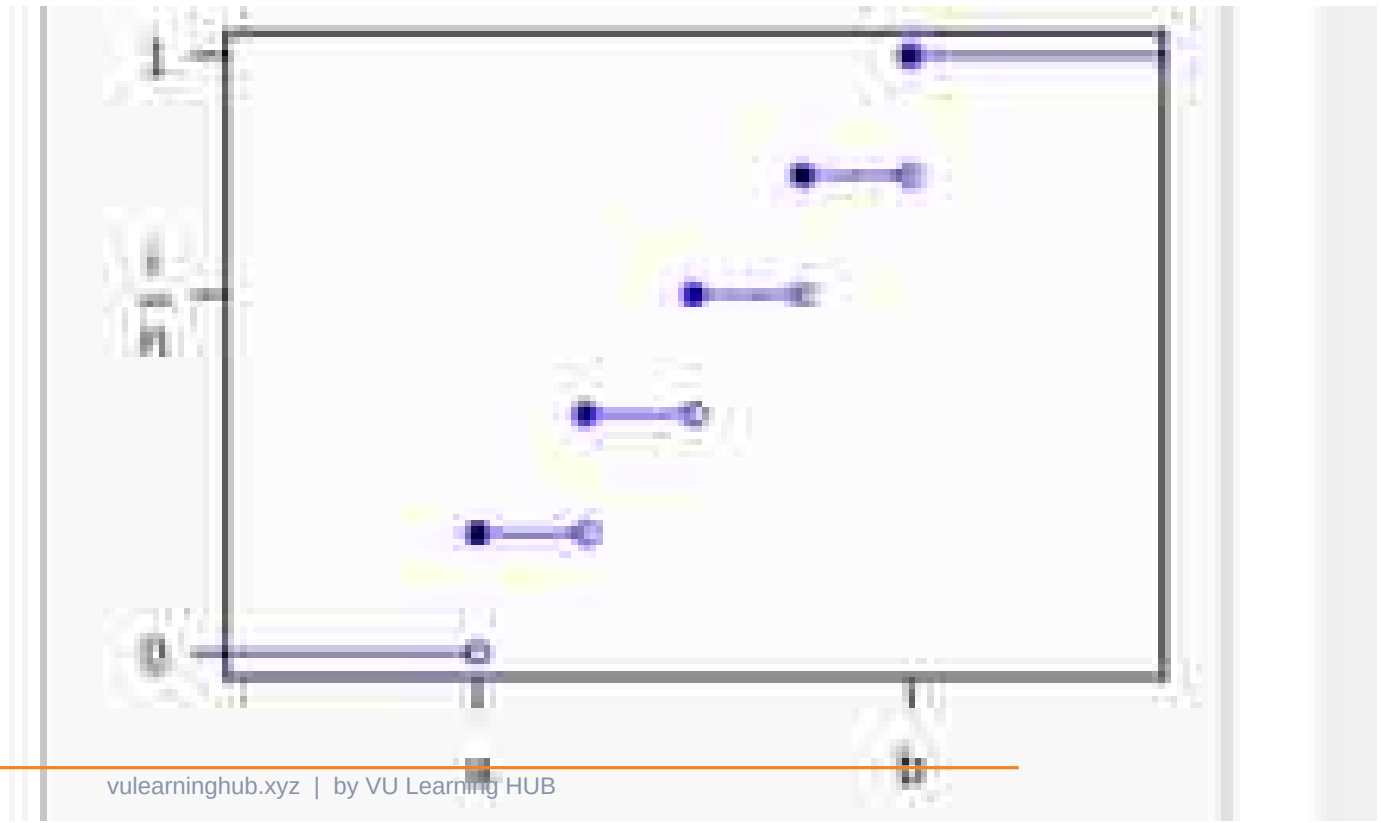
Step 2: Write as follows:

$$p(x)=\begin{cases}0 & -\infty < x < 1\\ 1/6 & 1 \le x < 2\\ 2/6 & 2 \le x < 3\\ 3/6 & 3 \le x < 4\\ 4/6 & 4 \le x < 5\\ 5/6 & 5 \le x < 6\\ 6/6 & 6 \le x < \infty\end{cases}$$

vulearninghub.xyz | by VU Learning HUB

The entire X -axis (from $-\infty$ to ∞) must be covered.

Graph of the CDF is a step function:



Step 2: Write as follows:

$$p(x)=\begin{cases}0 & -\infty < x < 1\\ 1/6 & 1 \le x < 2\\ 2/6 & 2 \le x < 3\\ 3/6 & 3 \le x < 4\\ 4/6 & 4 \le x < 5\\ 5/6 & 5 \le x < 6\\ 6/6 & 6 \le x < \infty\end{cases}$$

vulearninghub.xyz | by VU Learning HUB

The entire X -axis (from $-\infty$ to ∞) must be covered.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 118

Derivation of the Mean and Variance of the Discrete Uniform Distribution

We derive the formula for the mean of the discrete random variable X ~ U(1,*n*):

$$\mu = \sum_{x=1}^{n} x p_X(x) = \sum_{x=1}^{n} x.\frac{1}{n}$$

$$= \frac{1}{n}(1+2+...+n) = \frac{1}{n} \times \frac{n(n+1)}{2}$$

or

$$\mu = \frac{n+1}{2}$$

where we have used the famous identity $1+2+\ldots+n = n(n+1)/2$.

$$Var(X) = E(X^2) - (E(X))^2$$

First

$$E(X^2) = \frac{1}{n}\sum_{x=1}^{n} x^2 = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

And we already know that

$$\mu = E(X) = \frac{n+1}{2}.$$

Therefore $\sigma^2 = E\left(X^2\right) - \left(E\left(X\right)\right)^2$

$$= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4}$$

$$= \frac{4(n+1)(2n+1) - 6(n+1)^2}{24} = \frac{2(n+1)\left[2(2n+1) - 3(n+1)\right]}{24}$$

$$= \frac{2(n+1)\left[(4n+2) - (3n+3)\right]}{24} = \frac{2(n+1)(n-1)}{24}$$

$$= \frac{2(n^2-1)}{24} = \frac{n^2-1}{12}$$

The square root of this expression $\frac{n^2-1}{12}$ gives the

Standard Deviation of the discrete uniform distribution i.e.

$$\sigma = \sqrt{\frac{n^2-1}{12}}$$

Example:

Let us consider the rolling of a fair die, when X is representing the number of dots on the uppermost face of the die , the values would be 1, 2,3,4,5,6. therefore, n = 6

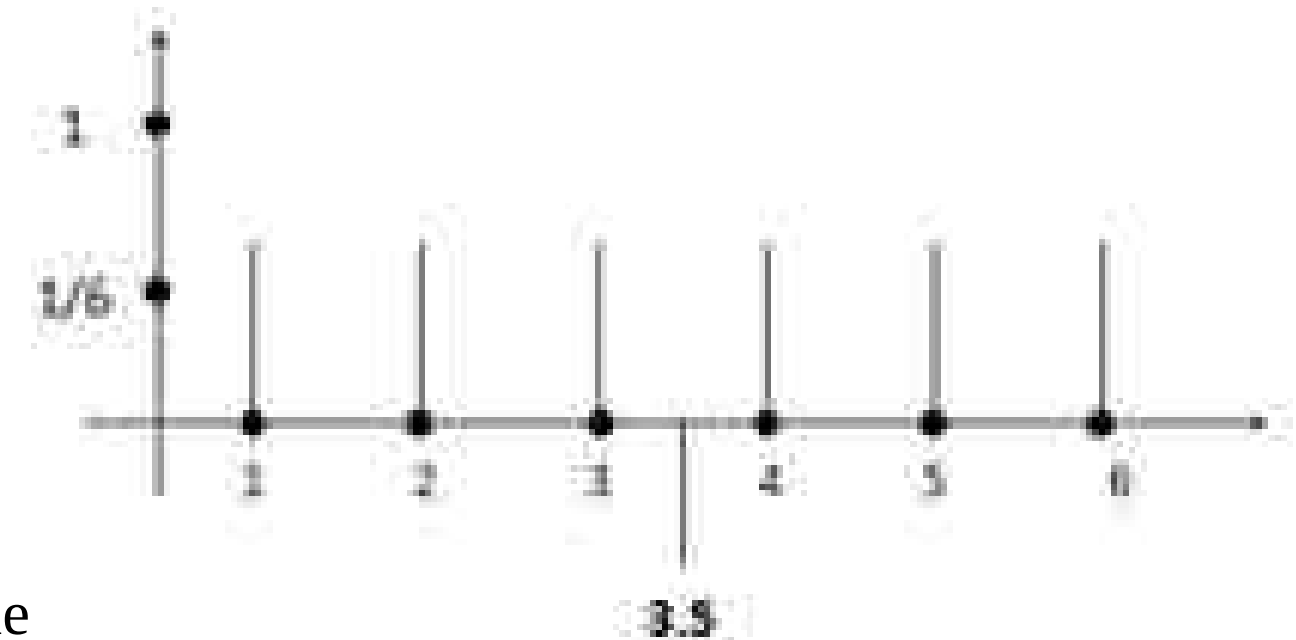
The probabilities of which are 1/6

Mean =μ = 3.5

We know that

$$\mu = E(X) = \frac{n+1}{2}$$

when n= 6 we get the mean as 3.5



Absolutely symmetric distribution

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 119

Derivation of the MGF of the Discrete Uniform distribution

A discrete uniform distribution defined on the set $\{a, a+1, a+2,....,b-1, b\}$ is denoted by the symbol $\boldsymbol{U(a, b)}$.

The constants a and b are called the ***parameters*** of the discrete uniform distribution.

We consider the discrete uniform distribution defined on the set $\{1,2,3,\ldots, n\}$.

Clearly, we have n possible outcomes, and since each is equally likely, we have $p(x)=1/n$.

The moment generating function is found by evaluating $E(e^{tX})$.

In other words, by definition,

$$M(t) = E\left(e^{tx}\right) = \sum_{x=1}^{N}\left(e^{tx}\frac{1}{n}\right) = \frac{1}{n}\sum_{x=1}^{n}\left(e^{t}\right)^{x}.$$

Now $\sum_{x=1}^{n} \left(e^{t}\right)^{x}$

is a geometric series with first term e^{t} and common ratio e^{t}.

We know that, for a geometric series with first term a and common ratio $r \neq 1$:

$$\text{Sum of the series} = \frac{a\left(1-r^{n}\right)}{\left(1-r\right)}$$

Here $a = e^{t}$, $r = e^{t}$ and note that $e^{t} \neq 1$

$$\text{So } \sum_{x=1}^{n}\left(e^{t}\right)^{x} = \frac{e^{t}\left(1-\left(e^{t}\right)^{n}\right)}{\left(1-\left(e^{t}\right)\right)} = \frac{e^{t}\left(1-e^{tn}\right)}{\left(1-e^{t}\right)}$$

vulearninghub.xyz by VU Learning HUB

Now,

since $M(t) = \frac{1}{n}\sum_{x=1}^{n}\left(e^{t}\right)^{x}$

$$\therefore\ M(t) = \frac{1}{n}\left(\frac{e^{t}\left(1-e^{tn}\right)}{\left(1-e^{t}\right)}\right) = \frac{e^{t}\left(1-e^{tn}\right)}{n\left(1-e^{t}\right)}$$

This can also be written as

$$M(t) = \frac{e^{t}-e^{(n+1)t}}{n\left(1-e^{t}\right)}.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 120

Binomial distribution

(Binomial experiment and PMF of the distribution)

Binomial Experiment

An experiment is called a **binomial experiment** if it possesses the following **four** properties:

1. The outcome of each trial may be classified into one of two categories, conventionally called **success** **(*S*)** and **Failure (*F*).**

It is to be noted that the outcome of **interest** is called a **success** an the other, a **failure**.

2. The probability of success, denoted by *p*; remains **constant** for all trials.

3. The successive trials are all **independent.**

4. The experiment is repeated a **fixed** number of times, say *n*.

Binomial Distribution

Let the random variable X denote the number of successes in *n* trials --- when a binomial experiment is performed.

Then the PMF of X is given by

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & x = 0,1,2,\ldots,n \\ 0 & \textit{elsewhere.} \end{cases} \qquad (1)$$

where *p* is the probability of success in each trial and *n* is the number of trials.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 121

Binomial distribution

(Binomial experiment and PMF of the distribution)

EXAMPLE

Example:

An event has the probability $p = 3/8$.
Find the complete binomial distribution for $n=5$ trials.

Here $p=3/8$, so that,

$q = 1- p = 5/8$; and $n = 5$.

Hence the desired probabilities are the successive terms in the binomial expansion of $(5/8+3/8)^5$, i.e.

$$\left[\binom{5}{0}\left(\frac{5}{8}\right)^{5}\left(\frac{3}{8}\right)^{0}+\binom{5}{1}\left(\frac{5}{8}\right)^{4}\left(\frac{3}{8}\right)^{1}+\binom{5}{2}\left(\frac{5}{8}\right)^{3}\left(\frac{3}{8}\right)^{2}+\binom{5}{3}\left(\frac{5}{8}\right)^{2}\left(\frac{3}{8}\right)^{3}\right.$$
$$\left.+\binom{5}{4}\left(\frac{5}{8}\right)^{1}\left(\frac{3}{8}\right)^{4}+\left(\frac{3}{8}\right)^{5}\right]$$

$$\left[\left(\frac{5}{8}\right)^5+\binom{5}{1}\left(\frac{5}{8}\right)^4\left(\frac{3}{8}\right)^1+\binom{5}{2}\left(\frac{5}{8}\right)^3\left(\frac{3}{8}\right)^2+\binom{5}{3}\left(\frac{5}{8}\right)^2\left(\frac{3}{8}\right)^3\right.$$
$$\left.+\binom{5}{4}\left(\frac{5}{8}\right)^1\left(\frac{3}{8}\right)^4+\left(\frac{3}{8}\right)^5\right]$$

We can now write these probabilities in the form of a probability table as below:

x	0	1	2	3	4	5
$P(X=x)$	0.0954	0.2861	0.3233	0.2060	0.0618	0.0074

Binomial Theorem

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

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 122

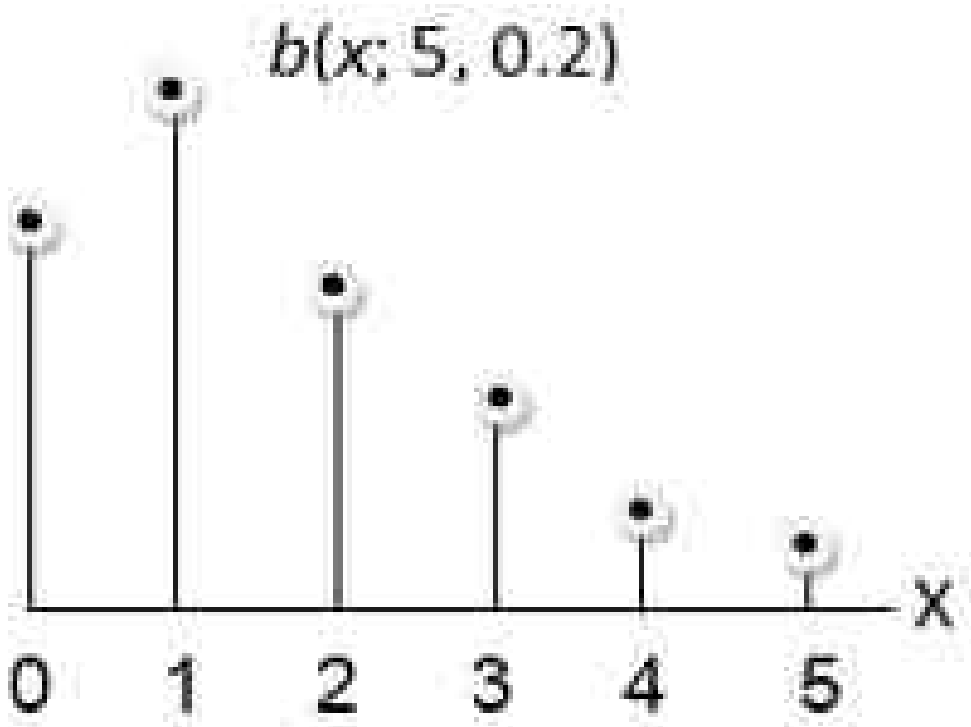
Shape

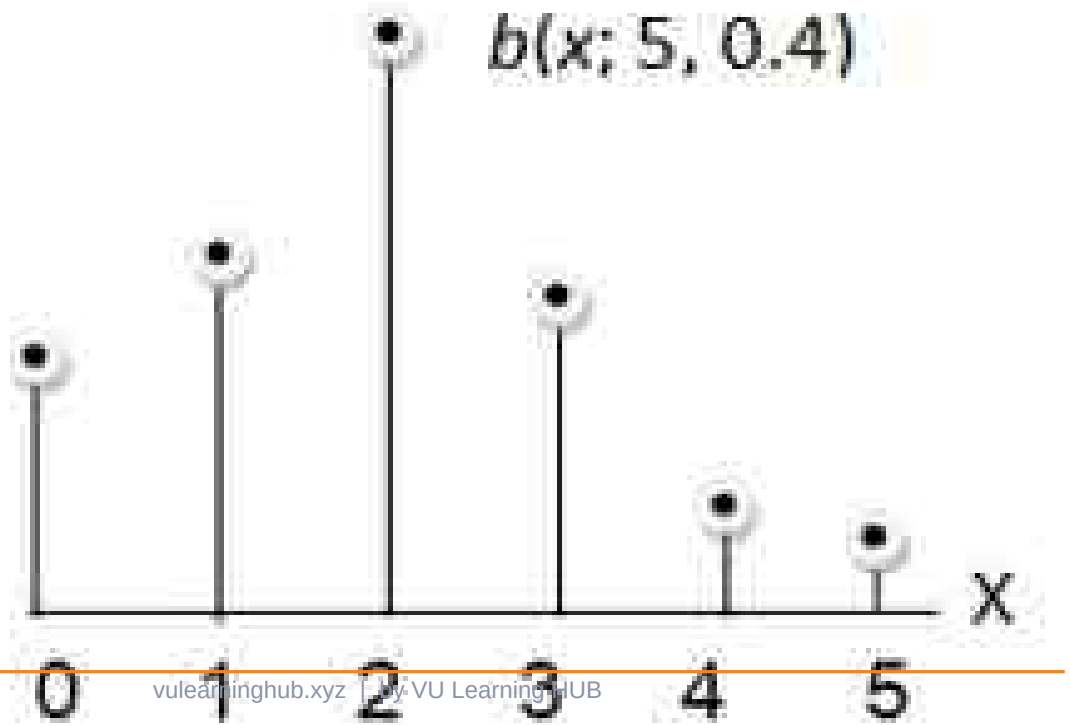
of the

Binomial distribution

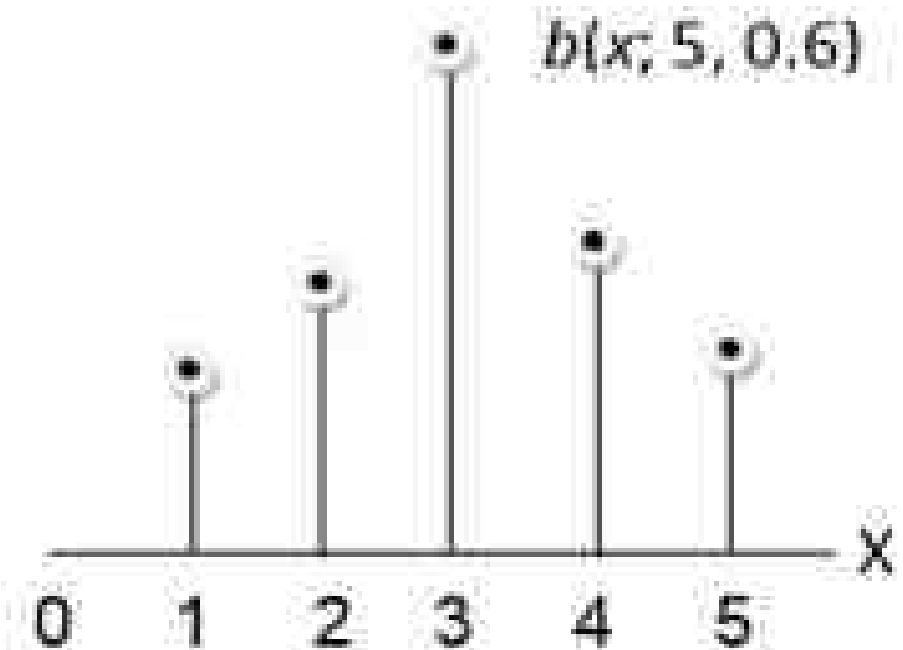
- The **shape** of the binomial probability distribution depends on the values of the **two parameters** p and n.

- The sketches given indicate the influence of p and n on the shape of the distribution.


b(x; 5, 0.2)
X
0
1
2
3
4
5



vulearninghub.xyz | by VU Learning HUB


b(x; 5, 0.6)
X
0
1
2
3
4
5

- Thus, we observe that, when p < ½, the distribution is positively skewed and when p > ½, the distribution becomes negatively skewed.

- When $p = 1/2$, the distribution is always symmetrical.

In general:

- As n, the number of trials, increases to ∞,

  $\beta_1 \rightarrow 0$ and $\beta_2 \rightarrow 3$.

- Hence for large n, the binomial probability distribution is symmetrical and mesokurtic.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 123

An example pertaining to the Binomial Distribution

Example:

If Y is $b(n,1/3)$, then $P(Y \geq 1) = 1 - P(Y=0) = 1 - (2/3)^n$.

Find the smallest value of n that yields $P(Y \geq 1) > 0.80$.

Solution:

Let us first see how

$$P(Y \geq 1) = 1 - P(Y<1) = 1 - P(Y=0) = 1 - (2/3)^n.$$

We know that $P(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}$

so $P(Y = 0) = \binom{n}{0} p^0 (1-p)^{n-0} = (1-p)^n$

Here $p = 1/3$ and n

So $P(Y = 0) = (1-1/3)^n = (2/3)^n$

Now, we are required to find the smallest value of n that yields $P(Y \geq 1) > 0.80$.

For this, we proceed as follows:

The condition can be written as $\mathbf{1-(2/3)^n > 0.80}$ implying that $0.20 > (2/3)^n$.

Taking log of both sides, we have

$$\ln(0.20) > n \ln(2/3)$$

implying that

$$-1.6094 > n(-0.4055)$$

or $$-1.6094 > -0.4055n$$

or $$1.6094 < 0.4055n$$

or $$n > 1.6094/0.4055 \quad \text{i.e. } n > 3.97.$$

vulearninghub.xyz | by VU Learning HUB

Thus, we see that $n = 4$ is the solution.

Interpretation:

The probability is greater than 0.80 i.e. 80% that **at least one success** will be obtained in ***n* = 4** independent repetitions of a random experiment for which the probability of success *p* is equal to 1/3.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 124

Mean and Variance of the Binomial Distribution

explained through

an

Example

Example

The probability that a planted radish seed germinates is 0.80. A gardener plants **nine** seeds.

Let X denote the number of radish seeds that successfully germinate.

(i) What is the **average number** of seeds the gardener could expect to germinate?

(ii) What is the **standard deviation** of X?

vulearninghub.xyz | by VU Learning HUB

(iii) What is the probability that, out of the nine seeds, at least 8 will germinate ?

Solution:

Assuming that the nine seeds are selected at random from a **large** number of seeds, it is easy to see that X can be regarded as a **binomial** random variable --- as **follows**:

1. Either the seed will germinate (**success**) or not germinate (**failure**)
2. The germination/non-germination of each seed is **independent** of each other
3. The probability of germination of each seed is the same i.e. 0.80
4. The number of seeds is fixed i.e. 9.

Now, if X is a binomial random variable, then we know that the mean of X is np.

Therefore,

the gardener could expect, on average, $9 \times 0.80 = 7.2$ seeds to germinate.

Also, if X is a binomial random variable, then we know that the variance of X is:

$$np(1-p) = 9 \times 0.80 \times 0.2 = 1.44$$

therefore,

the standard deviation of X is the square root of 1.44, or 1.20.

What is the probability that, out of the nine seeds, at least 8 will germinate ?

We have

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{9}{x} 0.80^x (0.20)^{9-x}$$

$$So\ P(X = 8\ or\ X = 9) = \binom{9}{8} 0.80^8 (0.20)^{9-8} + \binom{9}{9} 0.80^9 (0.20)^{9-9}$$

$$= 9 \times 0.80^8 (0.20) + 1 \times 0.80^9 \times 1 = 9 \times 0.1678 \times 0.20 + 0.1342$$

$$= 0.3020 + 0.1342 = 0.4362 = 43.62\%$$

$$(a\ non-trivial\ amount\ of\ probability)$$

vulearning.edu.vn | by VU Learning Edu

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 125

Derivation of the

MGF

of a

Binomial distribution

The mgf of a binomial distribution is easily obtained as follows:

$$M(t) = \sum_x e^{tx} p(x) = \sum_{x=0}^{n} e^{tx} \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \binom{n}{x} \left(e^t p\right)^x (1-p)^{n-x}$$

$$= \left[(1-p) + pe^t\right]^n$$

for all real values of t.

Now, the **mean** μ and **variance** σ^2 of X may be computed from $M(t)$.

We know that $\mu = M'(0)$

and

$\sigma^2 = M''(0) - \mu^2$

Since

$$M'(t) = n\left[(1-p) + pe^{t}\right]^{n-1}\left(pe^{t}\right)$$

and

$$M''(t) = n\left[(1-p)+pe^{t}\right]^{n-1}\left(pe^{t}\right)+n(n-1)\ .\left[(1-p)+pe^{t}\right]^{n-2}\left(pe^{t}\right)^{2},$$

Since

$$M'(t) = n\left[(1-p) + pe^{t}\right]^{n-1}\left(pe^{t}\right)$$

so

$$M'(t) = n\left[(1-p) + pe^{t}\right]^{n-1}\left(pe^{t}\right)$$

$$\Rightarrow \mu = M'(0) = n\left[(1-p) + pe^{0}\right]^{n-1}\left(pe^{0}\right)$$

$$= n\left[(1-p) + p\right]^{n-1}(p) = n[1]^{n-1}(p) = np$$

vulearninghub.xyz | by VU Learning HUB

And since

$$M''(t) = n\left[(1-p)+pe^{t}\right]^{n-1}\left(pe^{t}\right)+n(n-1)\ .\left[(1-p)+pe^{t}\right]^{n-2}\left(pe^{t}\right)^{2}$$

so

$$M''(0) = n\left[(1-p)+pe^{0}\right]^{n-1}\left(pe^{0}\right)+n(n-1)\ .\left[(1-p)+pe^{0}\right]^{n-2}\left(pe^{0}\right)^{2}$$

$$= n\left[(1-p)+p\right]^{n-1}\left(p\right)+n(n-1)\ .\left[(1-p)+p\right]^{n-2}\left(p\right)^{2}$$

$$= n\left[1\right]^{n-1}\left(p\right)+n(n-1)\ .\left[1\right]^{n-2}\left(p\right)^{2}$$

$$= np+n(n-1)p^{2}$$

implying that

$$\sigma^2 = M''(0) - \mu^2$$

$$= \left[np + n(n-1)p^2 \right] - (np)^2$$

$$= np + \left(n^2 - n \right) p^2 - (np)^2$$

$$= np + n^2 p^2 - np^2 - n^2 p^2$$

$$= np - np^2 = np(1-p) = npq.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 126

Recognizing the values of the parameters of a Binomial distribution through the expression of its MGF

(explained through an example)

Example:

If the mgf of a random variable X is

$$M(t) = \left(\frac{2}{3} + \frac{1}{3} e^{t} \right)^{5},$$

Then, X has a binomial distribution with $n = 5$ and $p = 1/3$;

that is, the **pmf** of X is

$$p(x) = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{5-x} & x = 0,1,2,\ldots,5 \\ 0 & \textit{elsewhere.} \end{cases}$$

and

$$\mu = np = \frac{5}{3} \quad and \quad \sigma^2 = np(1-p) = \frac{10}{9}.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 127

The sum of m independent Binomial random variables with same value of p is also Binomial

Theorem:

Let $X_1, X_2, \ldots, X_m$ be **independent** random variables such that X_i has binomial $b(n_i, p)$ distribution, for $i = 1,2,\ldots,m$.

Let, $$Y = \sum_{i=1}^{m} X_i$$

Then Y has a binomial $b\left(\sum_{i=1}^{m} n_i p\right)$ distribution.

vulearninghub.xyz | by VUL earning HUB

Proof:

In general, the mgf of a binomial random variable X is

$$M_X(t) = \left(1 - p + pe^t\right)^n.$$

So, in case of m binomial random variables $X_i,\ i = 1, 2, ..., m,$

with $p_1 = p_2 = p_3 ..... p_m = p$ (i.e. a constant)

the mgf of X_i is $M_{X_i}(t) = \left(1 - p + pe^t\right)^{n_i}$

By **independence** it follows that

$$M_Y(t) = \prod_{i=1}^{m} \left(1 - p + pe^t\right)^{n_i} = \left(1 - p + pe^t\right)^{\sum_{i=1}^{m} n_i} = \left(q + pe^t\right)^{\sum_{i=1}^{m} n_i}.$$

Hence, $Y = \sum_{i=1}^{m} X_i$ has a binomial $b\left(\sum_{i=1}^{m} n_i, p\right)$ distribution.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 128

Negative Binomial Distribution

(PMF, Shape of the distribution, Mean and Variance)

Consider **Bernoulli trials** — that is,

(1) there are two possible outcomes,

(2) the trials are independent,

and

(3) p, the probability of success, remains the same from trial to trial.

In a sequence of such trials, let the random variable X denote the number of the trial at which the rth **success** occurs, where r is a fixed integer.

Then the **probability mass function** of *X* is:

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \; x = r, r+1, r+2, \ldots$$

and we say that X has a **negative binomial(*r*, *p*)** distribution.

vulearninghub.xyz | by VU Learning Hub

Example:

Suppose that a game consists of tossing a fair coin until we obtain the 3rd head.

Then

head is ‘success’,

***p=P(H)*=0.5**

&

***r*=3**

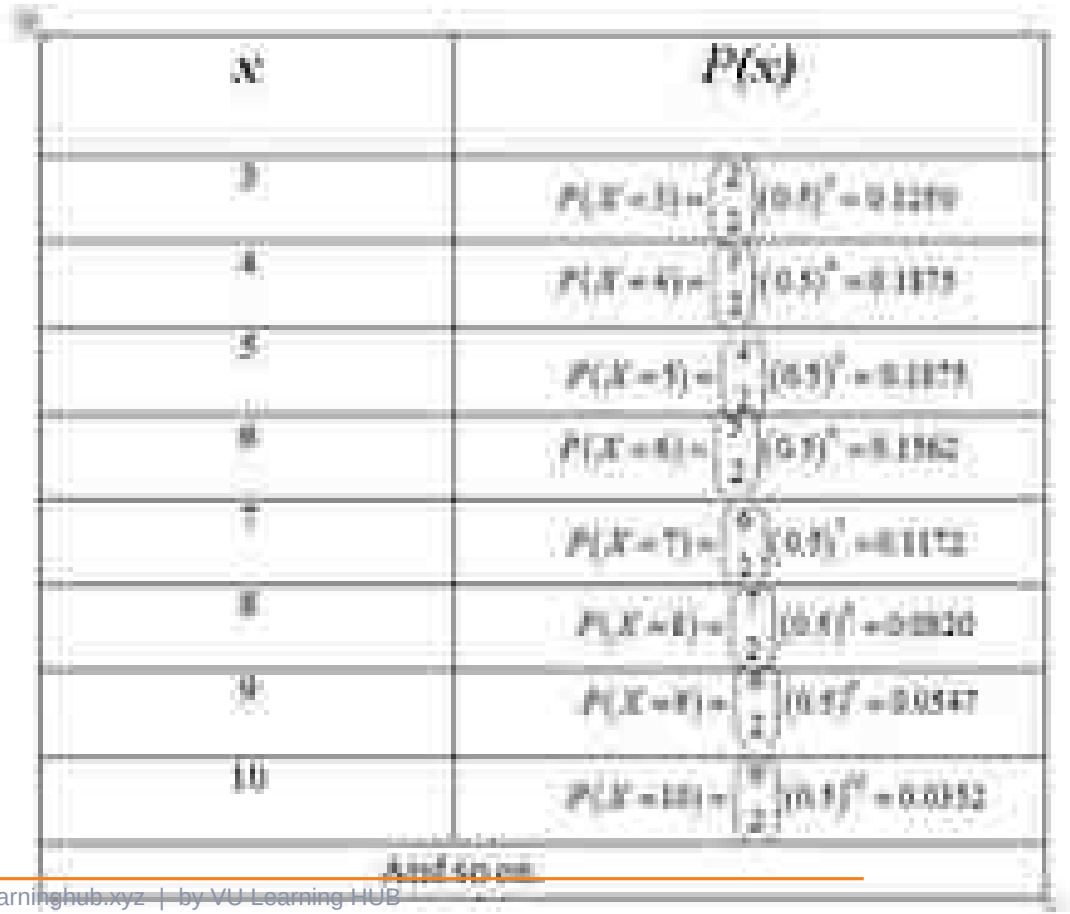
In this case, the negative binomial distribution is given by

$$P(X=x)=\binom{x-1}{3-1}(0.5)^3(1-0.5)^{x-3},\ x=3,3+1,3+2,\ldots$$

or

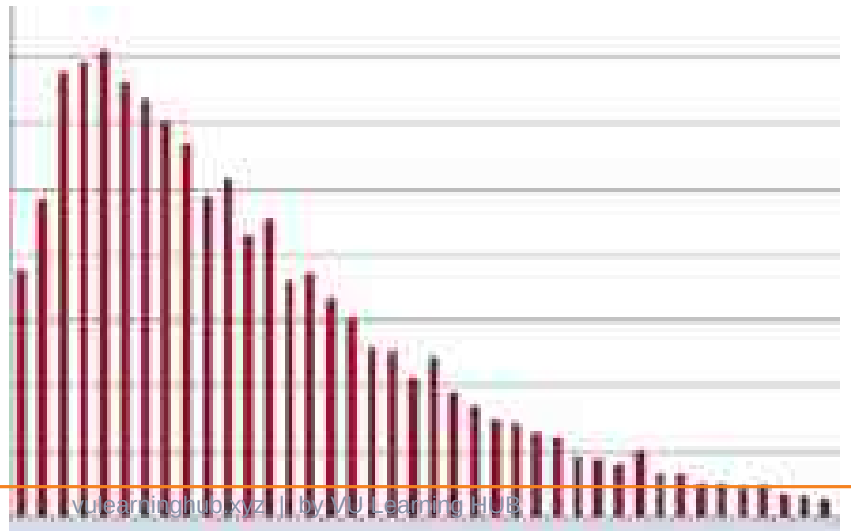
$$P(X=x)=\binom{x-1}{2}(0.5)^x,\ x=3,4,5,\ldots$$

Calculation of Probabilities



x	$P(x)$
3	$P(X=3)=\binom{2}{2}(0.5)^3=0.1250$
4	$P(X=4)=\binom{3}{2}(0.5)^4=0.1875$
5	$P(X=5)=\binom{4}{2}(0.5)^5=0.1875$
6	$P(X=6)=\binom{5}{2}(0.5)^6=0.1562$
7	$P(X=7)=\binom{6}{2}(0.5)^7=0.1172$
8	$P(X=8)=\binom{7}{2}(0.5)^8=0.0820$
9	$P(X=9)=\binom{8}{2}(0.5)^9=0.0547$
10	$P(X=10)=\binom{9}{2}(0.5)^{10}=0.0352$
And so on.	

Shape of the distribution



In general,

Shape of the Negative Binomial distribution

It is obvious that the negative binomial distribution is moderately positively skewed and that, as p tends to zero, the shape of the distribution tends to **normality**.

Mean and Variance of the Negative Binomial Distribution

The mean of the distribution is given by

$$E(X) = \frac{r(1-p)}{p}$$

and the Variance is

$$\text{Var}(X) = \frac{r(1-p)}{p^2}$$

MGF of the Negative Binomial Distribution

The moment generating function of the distribution is given by

$$M(t) = \left(\frac{1-p}{1-pe^{t}} \right)^{r} \quad for \quad t < -\log p$$

This can be used to find the mean, variance and higher moments of the distribution.

A Special Case of the Negative Binomial Distribution: Geometric Distribution

In a sequence of independent Bernoulli(p) trials,
let the random variable X denote the number of the trial
at which the ***1st*** **success** occurs.

Then the **probability mass function** of X is:

$$P(X = x) = \begin{pmatrix} x-1 \\ 0 \end{pmatrix} p^1 (1-p)^{x-1}, \; x = 1, 2, 3, \ldots$$

and we say that X has a **Geometric(p) distribution.**

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 129

Application of the Negative Binomial Distribution (explained through an example)

Example derived from
https://newonlinecourses.science.psu.edu/stat414/node/80/

Example

An oil company conducts a geological study that indicates that an exploratory oil well should have a **20% chance** of striking oil.

The company is wanting success **twice** i.e. two oil wells from where they will be able to pull out oil --- and is willing to try upto five times.

What is the probability that the **second success** comes on the **fifth well drilled**?

vulearninghub.xyz | by VU Learning HUB

Solution:

We know that, in a sequence **of independent Bernoulli(p) trials**, if the random variable X denotes the number of the trial at which the ***r*th success** occurs, then then the **probability mass function** of X is:

$$P(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \; x = r, r+1, r+2, \ldots$$

Here, **we do have** a sequence of Bernoulli trials as --- whenever we will dig an exploratory oil well, either we will or will not strike oil.

We can assume that the trials are **independent,**

and, **due to the independence** of the trials, we can assume that *p*, the probability of success, remains the **same** from trial to trial.

If we regard Striking oil as Success, then ***p*=0.20** and ***r*=2**

and, as such, the PMF of the negative binomial distribution will be given by

$$P(X=x)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r},\ x=r,r+1,r+2,\ldots$$

putting the values

$$P(X=x)=\binom{x-1}{2-1}(0.2)^{2}(0.8)^{x-2},\ x=2,3,4,\ldots$$

or $$P(X=x)=\binom{x-1}{1}(0.2)^{2}(0.8)^{x-2},\ x=2,3,4,\ldots$$

or $$P(X=x)=(x-1)(0.2)^{2}(0.8)^{x-2},\ x=2,3,4,\ldots$$

To find the required probability, we need to find ***P*(*X* = 5).**

$$P(X = x) = (x-1)(0.2)^2(0.8)^{x-2}$$

$$\Rightarrow P(X = 5) = (5-1)(0.2)^2(0.8)^{5-2}$$

$$= 4(0.2)^2(0.8)^3$$

$$= 0.082$$

$$= 8.2\%$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 130

Geometric Distribution (PMF and shape of the distribution)

Geometric Distribution
(PMF and shape of the distribution)

Assume Bernoulli trials — that is,

(1) there are two possible outcomes,

(2) the trials are independent, and

(3) *p*, the probability of success, remains the same from trial to trial.

Geometric Distribution
(PMF and shape of the distribution)

(4) *n*, the number of trials is fixed in advance

and

the binomial distribution is given by

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-1}, \; x = 0,1,2,3,\ldots,n$$

When n=1,

there are two possible outcomes,

and p, the probability of success, is the only parameter of the Bernoulli distribution given by

$$p_X(x) = \binom{1}{x} p^x (1-p)^{1-x}, \; x = 0,1$$

or

$$p_X(x) = p^x (1-p)^{1-x}, \; x = 0,1$$

If we let X denote the number of trials **until the first success.**

Then, the **probability mass function** of X is:

$$p_X(x) = (1-p)^{x-1} p,\ x = 1, 2, 3, \ldots$$

Example:

Suppose that a game consists of tossing a fair coin until we obtain the first head.

Then $p=P(H)=0.5$

And, in this case, the geometric distribution is given by

$$p_X(x) = (1-0.5)^{x-1}(0.5) = (0.5)^x, \quad x = 1, 2, 3, \ldots$$

or

$$p_X(x) = \frac{1}{2^x}, \quad x = 1, 2, 3, \ldots$$

Representation in tabular form:

X	$P(x)$
1	$1/2^1=1/2=0.5$
2	$1/2^2=1/4=0.25$
3	$1/2^3=1/8=0.125$
4	$1/2^4=1/16=0.0625$
5	$1/2^5=1/32=0.03125$
6	$1/2^6=1/64=0.015625$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 131

Application

of the Geometric Distribution

(explained through an example)

Example:

If the probability that a person will believe a rumor about the retirement of a certain politician is 0.25, what is the probability that

the sixth person to hear the rumor will be the first to believe it.

Solution:

In general, we know that if X denotes the number of trials **until the first success**

then, the probability mass function of X is:

$$p_X(x) = (1-p)^{x-1} p, \; x = 1, 2, 3, \ldots \qquad (1)$$

Here, let X denote the number of a person who hears the rumor and is the first one to believe it.

Then $p=0.25$ and we have

$$p_X(x) = (1-0.25)^{x-1}(0.25),\ x = 1, 2, 3, \ldots$$

or

$$p_X(x) = (0.75)^{x-1}(0.25),\ x = 1, 2, 3, \ldots$$

vulearninghub.xyz | by VU Learning HUB

Since the sixth person is the first to believe the rumor, i.e. the first success occurs on the sixth trial, therefore we will put $x=6$.

Hence

$$P(X=6)=(0.75)^{6-1}(0.25)=(0.75)^{5}(0.25)$$
$$=0.2373\times 0.25=0.059=5.9\%$$

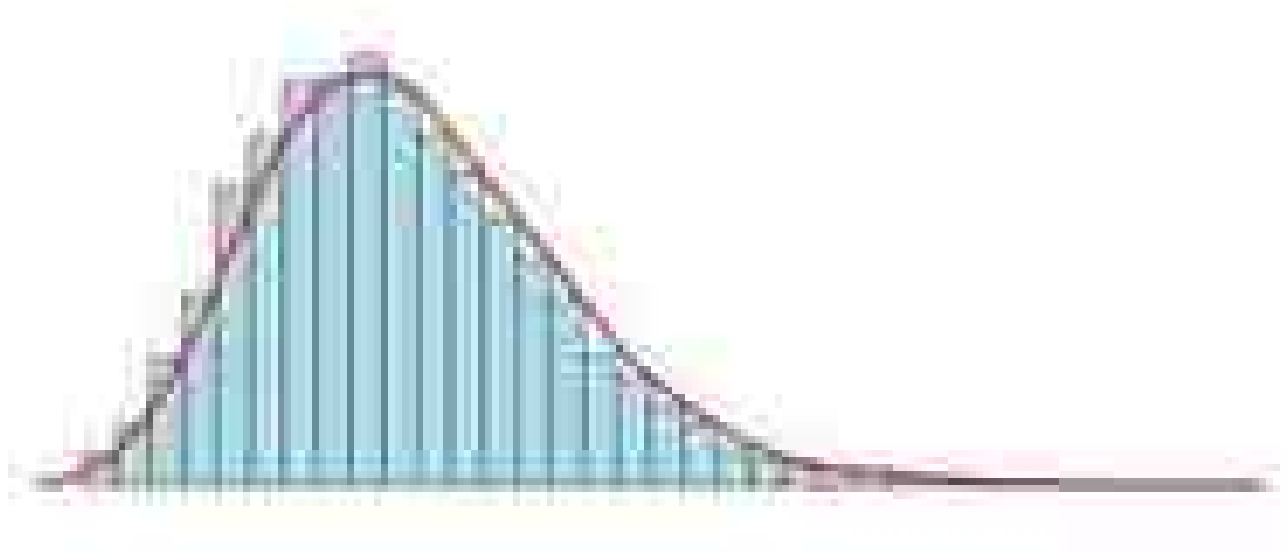
Note that,

If the first person is the first to believe the rumor, i.e. the first success occurs on the first trial, therefore we will put $x=1$.

Hence

$$P(X=1)=(0.75)^{1-1}(0.25)=(0.75)^{0}(0.25)$$

$$=1\times 0.25=0.25=25\%$$



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 132

Mean and Variance of the Geometric Distribution (explained through an example)

Example:

- A representative from an ice-cream producing company randomly stops people on a randomly selected street in the city of Lahore until he finds a person who actually purchased and tried the latest product of this company meaning a new kind of ice-cream they have just now brought into the market.

- Let *p*, the probability that, for any one of the persons stopped, he succeeds in finding such a person, who has actually tried that product equals 0.20.

- And, let X denote the number of people he stops until he finds his first success.

- How many people should we **expect** the marketing representative needs to stop before he finds one who actually purchased and tried this ice-cream?

- And, while we're at it, what is the variance?

Solution:

'How many people should we expect' means 'what is the average number).

For the geometric random variable X, the average number is given by:

$$\mu = E(X) = \frac{1}{p} = \frac{1}{0.20} = 5$$

So here

$$\mu = E(X) = \frac{1}{0.20} = 5$$

That is, we should expect the marketing representative to have to stop 5 people before he finds one who actually purchased and tried that particular ice-cream.

If this particular experiment of stopping people is repeated again and again, then on the **average** we can expect he would have to stop 5 people and the fifth one will be the one who had tried it.

For the geometric random variable X, the average number is given by:

$$\sigma^2 = Var(X) = \frac{1-p}{p^2}$$

so here

$$\sigma^2 = Var(X) = \frac{1-p}{p^2} = \frac{1-0.20}{0.20^2} = \frac{0.80}{0.04} = 20$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 133

Concept of Multinomial Distribution

Concept of Multinomial Distribution

The binomial distribution is generalized to be the multinomial distribution as follows:

Let a random experiment be repeated n independent times.

On each repetition, the experiment results in one of k mutually exclusive and exhaustive outcomes, says C_1, C_2, …, C_k.

Let p_i be the probability that the outcomes is an element of C_i and let p_i remain constant throughout the n independent repetitions, i= 1,2,…,k.

Now, we define the random variable X_i to be equal to the number of outcomes that are elements of C_i, $i = 1,2,\ldots, k-1$.

Furthermore,

Let $x_1, x_2, \ldots, x_{k-1}$ be nonnegative integers such that,

$$x_1 + x_2 + \ldots + x_{k-1} \leq n.$$

Suppose that a die is tossed 30 times, Then, x_1 represents the event how many times 1 came, x_2 represents how many times 2 came and so on upto x_6, k = 6. Suppose that $x_1 = 2$, $x_2 = 3$, $x_3 = 4$, $x_4 = 5$, $x_4 = 6$.

By adding all these numbers we get 2+3+4+5+6 = 20.

Total tosses were 30, therefore the last possibility of getting a 6 is 10 times.

and hence exactly $n - (x_1 + \ldots + x_{k-1})$

$= 30 - (2+3+4+5+6) = 30 - 20 = 10$

- The **probability mass function** of this multinomial distribution is:

$$f\left(x_1,\cdots,x_k;n,p_1,\cdots,p_k\right)$$

$$=\begin{cases}\dfrac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}\times p_2^{x_2}\cdots\times p_k^{x_k}, & \text{when } \sum_{i=1}^{k}x_i=n\\ 0 & \text{otherwise.}\end{cases}$$

for non-negative integers $x_1, \ldots, x_k$.

vulearninghub.xyz | by VU Learning HUB

This is the multinomial pmf of k-1 random variables $X_1, X_2,\ldots, X_{k-1}$ of the discrete type.

Properties:

The **mean** of the Multinomial distribution

$$E\left(X_i\right) = np_i$$

The **variance** of the Multinomial distribution

$$\mathrm{Var}\left(X_i\right) = np_i\left(1 - p_i\right)$$

The **covariance** of the Multinomial distribution

$$\operatorname{cov}\left(X_i, X_j\right) = -np_i p_j, \qquad i \neq j$$

Note:

The binomial distribution is **a special case** of the multinomial distribution.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 134

The binomial distribution is a special case of the multinomial distribution.

The binomial distribution is a special case of the multinomial distribution.

We have

$$f\left(x_1,\cdots,x_k;n,p_1,\cdots,p_k\right)$$

$$=\begin{cases}\dfrac{n!}{x_1!x_2!\cdots x_k!}\,p_1^{x_1}\times p_2^{x_2}\cdots\times p_k^{x_k}, \text{ when } \sum_{i=1}^{k}x_i=n\\ 0 \qquad \text{otherwise.}\end{cases}$$

Putting k=2, we obtain

$$f\left(x_1, x_2; n, p_1, p_2\right) = \begin{cases} \dfrac{n!}{x_1!\,x_2!}\, p_1^{x_1} p_2^{x_2}, & \text{when } \sum_{i=1}^{2} x_i = n \\ 0 & \text{otherwise.} \end{cases}$$

But

$$\sum_{i=1}^{2} x_i = n \text{ means } x_1 + x_2 = n \Rightarrow x_2 = n - x_1$$

And

$$\sum_{i=1}^{2} p_i = 1 \text{ means } p_1 + p_2 = 1 \Rightarrow p_2 = 1 - p_1$$

Hence, the above equation can be re-written as

$$f\left(x_1; n, p_1\right) = \frac{n!}{x_1!\left(n - x_1\right)!} p_1^{x_1} \left(1 - p_1\right)^{n - x_1}$$

or, simply

$$f\left(x; n, p\right) = \frac{n!}{x!\left(n - x\right)!} p^{x} \left(1 - p\right)^{n - x}$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 135

Real-Life Example of the Application of Multinomial Distribution

Example:

Suppose that the racial/ethnic distribution in a very large Western city is given by the table that follows:

White	Black	Hispanic
65%	20%	15%

Suppose that a **jury** of **twelve members** is to be chosen from this city in such a way that each resident has an **equal probability** of being selected **independently** of every other resident.

What is the probability that the jury contains:

(i) **Seven** White, **three** Black and **two** Hispanic members;

(ii) **Four** White and **eight** Other members;

(iii) At the most **one** Black member.

Solution

It is easy to see that, in solving each of the three questions, we will be dealing with **special cases** of the **Multinomial** distribution.

Part (i):

What is the probability that the jury contains **seven** White, **three** Black and **two** Hispanic members?

Solution:

Here, we are dealing with the Multinomial distribution with $k = 3$.

The resulting distribution is known as the **Trinomial** distribution.

To solve this problem, consider the random **vector**

$\boldsymbol{X} = (X_1, X_2, X_3)$

where

X_1 = number of White members,

X_2 = number of Black members and

X_3 = number of Hispanic members

Then $\boldsymbol{X}$ has a Trinomial distribution with parameters n = 12 and p_1=0.65, p_2 =0.20, p_3 =0.15.

Hence, the answer to the first question is:

$$P\left(X_1 = 7, X_2 = 3, X_3 = 2\right) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$= \frac{12!}{7!3!2!}\left(0.65\right)^7\left(0.20\right)^3\left(0.15\right)^2 = 0.0699 = 6.99\% \text{ or } 7\%.$$

Part (ii):

What is the probability that the jury contains **four** White and **eight** Other members?

Solution:

Here, we are dealing with the Multinomial distribution with $k = 2$, as follows:

White	Others (Black & Hispanic combined)
65%	20%+15%=35%

The resulting distribution is none other than the **Binomial** distribution.

Hence, the answer to the second question is:

$$P\left(X_1 = 4, X_2 = 8\right) = \frac{n!}{x_1!x_2!} p_1^{x_1} p_2^{x_2}$$

$$= \frac{12!}{4!8!}\left(0.65\right)^4\left(0.35\right)^8 = 0.0199 = 1.99\% \text{ or } 2\%.$$

Part (iii):

What is the probability that the jury contains at the most **one** Black member.

Solution:

Here again, we are dealing with the Multinomial distribution with $k = 2$, as follows:

Black	Others (White & Hispanic combined)
20%	65%+15%=80%

Once again, the resulting distribution is none other than the **Binomial** distribution.

With reference to question (iii), **regarding Black as 'success',** note that:

- We are dealing with a **binomial** distribution with n =12 and p = 0.20.
- X represents the **number of Black members** of the jury
- "**at the most one** Black member" means X = 0 or X= 1.

Using the binomial distribution, we have

$$P(X=0)=\frac{12!}{0!12!}(0.20)^0(0.80)^{12}=0.069$$

$$\text{and } P(X=1)=\frac{12!}{1!11!}(0.20)^1(0.80)^{11}=0.206$$

Therefore, the answer is:

$$P(X=0 \text{ or } X=1)=P(X=0)+P(X=1)$$

$$=0.069+0.206=0.275=27.5\%.$$

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 136

MGF

of the Multinomial Distribution

Recall: Concept of Multinomial Distribution

Let a random experiment be repeated n independent times.

On each repetition, the experiment results in one of k mutually exclusive and exhaustive outcomes, says C_1, C_2, …, C_k.

Then the **probability mass function** of this multinomial distribution is given by:

$$f\left(x_1,\cdots,x_k;n,p_1,\cdots,p_k\right)$$

$$=\begin{cases}\dfrac{n!}{x_1!x_2!\cdots x_k!}p_1^{x_1}\times p_2^{x_2}\cdots\times p_k^{x_k}, & \text{when } \sum_{i=1}^{k}x_i=n\\ 0 & \text{otherwise.}\end{cases}$$

for non-negative integers $x_1, \ldots, x_k$.

vulearninghub.xyz | by VU Learning HUB

For this purpose, we begin with the **definition** of the **mgf** of a **multivariate** distribution (in general):

Definition of Moment Generating Function of a Random Vector having *k* components:

Let $\underline{X} = (X_1, X_2, ..., X_k)$ be a random vector.

If $E\left(e^{t_1X_1+t_2X_2+...+t_kX_k}\right)$ exists for $|t_1| < h_1, |t_2| < h_2, ..., |t_k| < h_k$,

where $h_1, h_2, ..., h_k$ are positive, then $E\left(e^{t_1X_1+t_2X_2+...+t_kX_k}\right)$ is denoted by

$M_{X_1,X_2,...,X_k}(t_1, t_2, ..., t_k)$

and is called the moment generating function (mgf)

of the random vector X.

vulearninghub.xyz | by VU Learning HUB

As far as the moment generating function of the **multinomial distribution** is concerned, we have:

$$M_{\underline{X}}(t) = M_{\underline{X}}(t_1, t_2, \ldots, t_k) = E\left(\mathrm{e}^{\sum_{i=1}^{k} t_i X_i} \right)$$

$$= \left(\sum_{i=1}^{k} p_i \, \mathrm{e}^{t_i} \right)^n$$

The **cumulant generating function** of the Multinomial distribution is given by

$$K_{\underline{X}}(t) = K_{\underline{X}}(t_1, t_2, \ldots, t_k) = \log\left[M_{\underline{X}}(t_1, t_2, \ldots, t_k)\right]$$

$$= n \log\left(\sum_{i=1}^{k} p_i e^{t_i}\right)$$

Coming back to the **MGF,** the question is:

How to **utilize** the mgf of the Multinomial distribution?

For this, RECALL:

The **general** method of obtaining **Product Moments** and **simple Moments** from the **MGF** of a random vector

Suppose that k=3.

Then the **means** of the random variables X_1, X_2 and X_3 are given by:

$$E(X_1) = \frac{\partial M(0,0,0)}{\partial t_1},$$

$$E(X_2) = \frac{\partial M(0,0,0)}{\partial t_2},$$

$$E(X_3) = \frac{\partial M(0,0,0)}{\partial t_3}$$

Also, let us suppose that the trinomial experiment is being repeated **only 2 times**.

The MGF of the trinomial distribution is then given by

$$M_{\underline{X}}(t) = M_{X_1,X_2,X_3}(t_1,t_2,t_3) = E\left(\mathrm{e}^{\sum_{i=1}^{3} t_i X_i}\right)$$

$$= \left(\sum_{i=1}^{3} p_i \,\mathrm{e}^{t_i}\right)^2 = \left(p_1 e^{t_1} + p_2 e^{t_2} + p_3 e^{t_3}\right)^2$$

Now we know that $(a+b+c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc$

Hence

$$M_{X_1,X_2,X_3}(t_1,t_2,t_3)$$

$$= p_1{}^2 e^{2t_1} + p_2{}^2 e^{2t_2} + p_3{}^2 e^{2t_3} + 2p_1p_2e^{t_1+t_2} + 2p_1p_3e^{t_1+t_3} + 2p_2p_3e^{t_2+t_3}$$

Then the **means** of the random variables X_1, X_2 and X_3 are given by:

$$E(X_1) = \frac{\partial M(0,0,0)}{\partial t_1},$$

$$E(X_2) = \frac{\partial M(0,0,0)}{\partial t_2},$$

$$E(X_3) = \frac{\partial M(0,0,0)}{\partial t_3}$$

Hence, taking the partial derivative of the MGF with respect to t_1, we have:

$$\frac{\partial}{\partial t_1}\left(M_{X_1,X_2,X_3}\left(t_1,t_2,t_3\right)\right)=\frac{\partial}{\partial t_1}\left(p_1{}^2e^{2t_1}+p_2{}^2e^{2t_2}+p_3{}^2e^{2t_3}+2p_1p_2e^{t_1+t_2}+2p_1p_3e^{t_1+t_3}+2p_2p_3e^{t_2+t_3}\right)$$

$$=2p_1{}^2e^{2t_1}+0+0+2p_1p_2e^{t_1}e^{t_2}+2p_1p_3e^{t_1}e^{t_3}+0=2p_1{}^2e^{2t_1}+2p_1p_2e^{t_1}e^{t_2}+2p_1p_3e^{t_1}e^{t_3}$$

As such,

$$\frac{\partial}{\partial t_1}\left(M_{X_1,X_2,X_3}\left(0,0,0\right)\right)=2p_1{}^2e^{2t(0)}+2p_1p_2e^0e^0+2p_1p_3e^0e^0$$

$$=2p_1{}^2\left(1\right)+2p_1p_2\left(1\right)\left(1\right)+2p_1p_3\left(1\right)\left(1\right)$$

$$=2p_1{}^2+2p_1p_2+2p_1p_3=2p_1(p_1+p_2+p_3)=2p_1(1)=2p_1$$

Taking the partial derivative of the MGF with respect to t_2, we have:

$$\frac{\partial}{\partial t_2}\left(M_{X_1,X_2,X_3}\left(t_1,t_2,t_3\right)\right)=\frac{\partial}{\partial t_2}\left(p_1{}^2e^{2t_1}+p_2{}^2e^{2t_2}+p_3{}^2e^{2t_3}+2p_1p_2e^{t_1+t_2}+2p_1p_3e^{t_1+t_3}+2p_2p_3e^{t_2+t_3}\right)$$

$$=0+2p_2{}^2e^{2t_2}+0+2p_1p_2e^{t_1}e^{t_2}+0+2p_2p_3e^{t_2}e^{t_3}=2p_2{}^2e^{2t_2}+2p_1p_2e^{t_1}e^{t_2}+2p_2p_3e^{t_2}e^{t_3}$$

Hence

$$\frac{\partial}{\partial t_2}\left(M_{X_1,X_2,X_3}\left(0,0,0\right)\right)=2p_2{}^2e^{2(0)}+2p_1p_2e^0e^0+2p_2p_3e^0e^0$$

$$=2p_2{}^2(1)+2p_1p_2(1)(1)+2p_2p_3(1)(1)$$

$$=2p_2{}^2+2p_1p_2+2p_2p_3=2p_2(p_2+p_1+p_3)=2p_2(1)=2p_2$$

And, taking the partial derivative of the MGF with respect to t_3, we have:

$$\frac{\partial}{\partial t_3}\left(M_{X_1,X_2,X_3}\left(t_1,t_2,t_3\right)\right)=\frac{\partial}{\partial t_3}\left(p_1{}^2e^{2t_1}+p_2{}^2e^{2t_2}+p_3{}^2e^{2t_3}+2p_1p_2e^{t_1+t_2}+2p_1p_3e^{t_1+t_3}+2p_2p_3e^{t_2+t_3}\right)$$

$$=0+0+2p_3{}^2e^{2t_3}+0+2p_1p_3e^{t_1}e^{t_3}+2p_2p_3e^{t_2}e^{t_3}=2p_3{}^2e^{2t_3}+2p_1p_3e^{t_1}e^{t_3}+2p_2p_3e^{t_2}e^{t_3}$$

Hence

$$\frac{\partial}{\partial t_3}\left(M_{X_1,X_2,X_3}\left(0,0,0\right)\right)=2p_3{}^2e^{2(0)}+2p_1p_3e^0e^0+2p_2p_3e^0e^0$$

$$=2p_3{}^2(1)+2p_1p_3(1)(1)+2p_2p_3(1)(1)$$

$$=2p_3{}^2+2p_1p_3+2p_2p_3=2p_3(p_3+p_1+p_2)=2p_3(1)=2p_3$$

But we already know that, for the Multinomial distribution:

$$E(X_i) = np_i$$

Here n=2.

$$\therefore E(X_i) = 2p_i$$

i.e.

$$E(X_1) = 2p_1, E(X_2) = 2p_2 \ \& \ E(X_3) = 2p_3$$

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 137

Hypergeometric Distribution (PMF and shape of the distribution)

Hypergeometric Distribution

The following conditions characterize the hypergeometric distribution:

1. The result of each draw (the elements of the population being sampled) can be classified into one of two mutually exclusive categories

   (e.g. Pass/Fail or Employed/Unemployed).

2. The successive draws are **not independent** of each other.

3. The number of draws is **fixe**d in advance.

4. The probability of a success **changes** on each draw, as each draw decreases the population (*sampling without replacement* from a finite population).

Example:

- Suppose that we have a class of Phd students in a small educational institution in some particular subject and the total number of PhD students of that subject is 5.

  So, N = 5

- We would like to draw 3 students to send on a Conference then n= 3.
- Success =achieved Grade A failure = not achieved Grade A

Success = achieved Grade A

Failure = not achieved Grade A (C or D)

If that population initially has 2 students who have achieved A Grade and 3 students who have not achieved A grade then we have another parameter that is k, which is the number of successes in the population at the beginning of the experiment which is k = 2.

Taking the sample of size 3 without replacement.

Probability of selecting a student

Taking the sample of size 3 without replacement.

Probability of selecting a student who has got an A grade = 2/5

Suppose that student is an A grade student and will select the second student having A grade.

What is the probability that the second student is also an A grade holder = 1/4

Definition

A random variable X follows the hypergeometric distribution if its probability mass function (pmf) is given by

$$P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}, \; x = 0,1,2,\ldots,K \; if \; K < n \;\; \& \;\; x = 0,1,2,\ldots,n \; if \; n < K$$

where,

- N is the population size,

- *K* is the number of success in the population at the beginning of the experiment ,- *n* is the number of draws (i.e. quantity drawn in each trial),

- *k* is the number of observed successes

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 138

Hypergeometric Distribution (EXAMPLE)

Example:

The names of 5 men and 5 women are written on slips of paper and placed in a hat. Four names are drawn.

What is the probability that 2 are men and 2 are women?

Let X denote the number of **men.**

Then,

N=5+5=10 names to be drawn from;

K=5 & n=4,

and here the possible values of x are 0,1,2,3,4, i.e. n)

Hence the hypergeometric distribution is

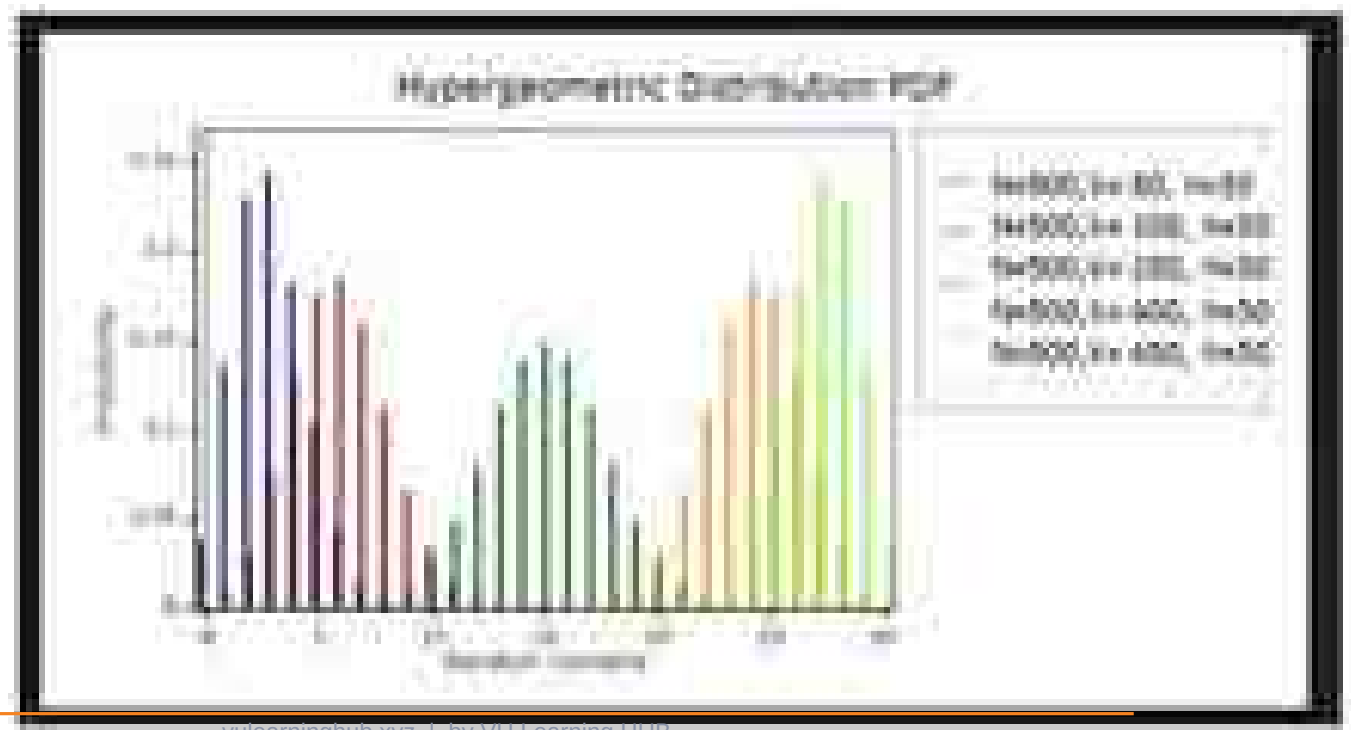
$$h(x;10,4,5)=\frac{\binom{5}{x}\binom{5}{4-x}}{\binom{10}{4}}$$

and the required probability, i.e. $P(X=2)$ is

$$h(2;10,4,5)=\frac{\binom{5}{2}\binom{5}{4-2}}{\binom{10}{4}}=\frac{\binom{5}{2}\binom{5}{2}}{\binom{10}{4}}=\frac{10}{21}.$$

Shape of the Distribution

https://www.google.com.pk/search?q=hypergeometric+distribution&rlz=1C1CHBD_enPK785PK785&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjYx9SouNTeAhWBDcAKHZ3hDOkQ_AUIDigB&biw=1280&bih=657#imgrc=6MdjpMk-oHbjKM:



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 139

Derivation of the Mean and Variance of the Hypergeometric Distribution

Two important properties of the hypergeometric probability distribution are given here.

The mean & variance of the Hypergeometric distribution are as follows:

$$\text{Mean} = E(X) = n\frac{K}{N}$$

&

$$\text{Variance} = n\frac{K}{N}\frac{(N-K)}{N}\frac{N-n}{N-1}$$

vulearninghub.xyz | by VU Learning HUB

Or --- in other words --- the mean and variance of the hypergeometric probability distribution are

$$\mu = np \text{ and } \sigma^2 = npq\left(\frac{N-n}{N-1}\right),$$

$$\text{where } p = \frac{k}{N} \text{ and } q = \frac{N-k}{N}.$$

Derivation of the Mean:

Let X have the hypergeometric probability distribution given by

$$h(x;N,n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

for x such that, $0 \le x \le n$ and $0 \le x \le k$.

Then the mean, μ, is given by

$$\mu = E(X)$$

$$= \sum_{x=0}^{n} \frac{x \cdot \binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = 0 + \sum_{x=1}^{n} \frac{x \cdot \binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \sum_{x=1}^{n} \frac{x \cdot k!\binom{N-k}{n-x}}{x!(k-x)!\binom{N}{n}}$$

or

$$\mu = E(X) = \sum_{x=1}^{n} \frac{x \cdot k(k-1)!\binom{N-k}{n-x}}{x(x-1)!(k-1+1-x)!\binom{N}{n}} = \sum_{x=1}^{n} \frac{k(k-1)!\binom{N-k}{n-x}}{(x-1)!((k-1)-(x-1))!\binom{N}{n}}$$

$$= \sum_{x=1}^{n} \frac{k\binom{k-1}{x-1}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{k}{\binom{N}{n}} \sum_{x=1}^{n} \binom{k-1}{x-1}\binom{N-k}{n-x}$$

Let $y = x - 1$ implying that $x = y + 1$, then

$$\mu = \frac{k}{\binom{N}{n}} \sum_{y=0}^{n-1} \binom{k-1}{y} \binom{N-k}{n-1-y} \quad [\because \text{when } x = 1, y = 0, \text{ and when } x = n, y = n-1]$$

$$= \frac{k}{\binom{N}{n}} \binom{N-1}{n-1} \qquad \left[\because \sum_{j=0}^{m} \binom{a}{j} \binom{b}{m-j} = \binom{a+b}{m}\right]$$

Now $\mu = \dfrac{k}{\binom{N}{n}}\binom{N-1}{n-1}$ can be re-written as

$$\mu = \frac{k}{\left(\dfrac{N!}{n!(N-n)!}\right)} \cdot \frac{(N-1)!}{(n-1)!((N-1)-(n-1))!}$$

$$= \frac{kn!(N-n)!}{N!} \cdot \frac{(N-1)!}{(n-1)!(N-n)!}$$

$$= \frac{kn(n-1)!(N-n)!}{N(N-1)!} \cdot \frac{(N-1)!}{(n-1)!(N-n)!} = \frac{nk}{N} = np \text{ where } p = \frac{k}{N}.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 140

Derivation of the Mean and Variance of the Hypergeometric Distribution

The variance of the Hypergeometric distribution are as follows:

$$\text{Variance} = n\frac{K}{N}\frac{\left(N-K\right)}{N}\frac{N-n}{N-1}$$

Or --- in other words --- the mean and variance of the hypergeometric probability distribution are

$$\mu = np \text{ and } \sigma^2 = npq\left(\frac{N-n}{N-1}\right),$$

$$\text{where } p = \frac{k}{N} \text{ and } q = \frac{N-k}{N}.$$

Derivation of the Variance:

Let X have the hypergeometric probability distribution given by

$$h(x;N,n,k)=\frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}},$$

for x such that, $0 \le x \le n$ and $0 \le x \le k$.

By definition, the variance, σ^2, is given by

$$\sigma^2 = E\left[X-\mu\right]^2 = E\left(X^2\right)-\mu^2$$

$$Now,\ \ E\left(X^2\right) = E\left[\,X\left(X-1\right)+X\,\right]$$

$$= E\left(X\right)+E\left[\,X\left(X-1\right)\right]$$

$$= \sum_{x=0}^{n} x.h\left(x;N,n,k\right)+\sum_{x=0}^{n} x\left(x-1\right)h\left(x;N,n,k\right)$$

$$\sigma^2 = E\left[X-\mu\right]^2 = E\left(X^2\right) - \mu^2$$

Now, $E\left(X^2\right)$

$$= \frac{nk}{N} + \frac{\sum_{x=0}^{n} x(x-1)\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} = \frac{nk}{N} + 0 + 0 + \frac{\sum_{x=2}^{n} x(x-1)\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$\text{Now } E\left(X^2\right) = \frac{nk}{N} + 0 + 0 + \frac{\sum_{x=2}^{n} x(x-1)\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \frac{nk}{N} + \frac{\sum_{x=2}^{n} x(x-1)\left(\frac{k(k-1)(k-2)!}{x(x-1)(x-2)!(k-x-2+2)!}\right)\binom{N-k}{n-x}}{\binom{N}{n}}$$

$$= \frac{nk}{N} + \frac{k(k-1)\sum_{x=2}^{n}\left(\frac{(k-2)!}{(x-2)!((k-2)-(x-2))!}\right)\binom{N-k}{n-x}}{\binom{N}{n}}$$

We have $E\left(X^2\right) = \frac{nk}{N} + \frac{k(k-1)\sum_{x=2}^{n}\binom{k-2}{x-2}\binom{N-k}{n-x}}{\binom{N}{n}}$

Let $y = x-2$ implying that $x = y+2$, then

$$E\left(X^2\right) = \frac{nk}{N} + \frac{k(k-1)\sum_{y=0}^{n-2}\binom{k-2}{y}\binom{N-k}{n-2-y}}{\binom{N}{n}} \quad \left[\because \text{when } x=2, y=0, \text{ and when } x=n, y=n-2\right]$$

Now

$$E\left(X^2\right)=\frac{nk}{N}+\frac{k\left(k-1\right)}{\binom{N}{n}}\sum_{y=0}^{n-2}\binom{k-2}{y}\binom{N-k}{n-2-y}$$

$$=\frac{nk}{N}+\frac{k\left(k-1\right)}{\binom{N}{n}}\binom{N-2}{n-2} \qquad \left[\because \sum_{j=0}^{m}\binom{a}{j}\binom{b}{m-j}=\binom{a+b}{m}\right]$$

$$\therefore E\left(X^2\right)=\frac{nk}{N}+\frac{k\left(k-1\right)}{\binom{N}{n}}\binom{N-2}{n-2}$$

$$=\frac{nk}{N}+\frac{k\left(k-1\right)\left(N-2\right)!}{\left(\frac{N!}{n!\left(N-n\right)!}\right)\left(n-2\right)!\left(N-n\right)!}=\frac{nk}{N}+\frac{k\left(k-1\right)\left(N-2\right)!n!\left(N-n\right)!}{N!\left(n-2\right)!\left(N-n\right)!}$$

$$=\frac{nk}{N}+\frac{k\left(k-1\right)\left(N-2\right)!n\left(n-1\right)\left(n-2\right)!\left(N-n\right)!}{N\left(N-1\right)\left(N-2\right)!\left(n-2\right)!\left(N-n\right)!}=\frac{nk}{N}+\frac{k\left(k-1\right)n\left(n-1\right)}{N\left(N-1\right)}$$

$$\therefore \sigma^2 = \mathrm{Var}(X) = E(X^2) - \mu^2 = \frac{nk}{N} + \frac{k(k-1)n(n-1)}{N(N-1)} - \left(\frac{nk}{N}\right)^2$$

$$= \frac{nkN(N-1)}{N^2(N-1)} + \frac{(k^2-k)(n^2-n)N}{N^2(N-1)} - \frac{(N-1)n^2k^2}{(N-1)N^2}$$

$$= \frac{nkN(N-1) + (n^2k^2 - kn^2 - nk^2 + nk)N - (N-1)n^2k^2}{N^2(N-1)}$$

$$= \frac{nkN^2 - nkN + n^2k^2N - kn^2N - nk^2N + nkN - n^2k^2N + n^2k^2}{N^2(N-1)}$$

$$= \frac{nkN^2 - kn^2N - nk^2N + n^2k^2}{N^2(N-1)} = \frac{nk(N^2 - nN - kN + nk)}{N^2(N-1)}$$

Now

$$\frac{nk\left(N^2 - nN - kN + nk\right)}{N^2\left(N-1\right)} = \frac{nk\left(N-k\right)}{N.N} \cdot \frac{(N-n)}{N-1}$$

So

$$\text{Var}(X) = npq \cdot \frac{N-n}{N-1}$$

where $p = \frac{k}{N}$ and $q = \frac{N-k}{N}$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 141

Poisson Distribution
(PMF and shape of the distribution)

Poisson Distribution
(PMF and shape of the distribution)

Consider the function $p(x)$ defined by

$$p(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!} & x = 0,1,2,\ldots,\ \lambda > 0 \\ 0 & \textit{elsewhere}, \end{cases} \qquad (1)$$

where $m > 0$.

Since $\lambda > 0$,

therefore $p(x) \geq 0$

and

$$\sum_{x} p(x) = \sum_{x=0}^{\infty} \frac{\lambda^{x} e^{-\lambda}}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!}$$

Recall that the series

$$1+m+\frac{m^2}{2!}+\frac{m^3}{3!}+\cdots=\sum_{x=0}^{\infty}\frac{m^x}{x!}$$

converges, for all values of m, to e^m.

Therefore

$$\sum_{x} p(x) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{x}}{x!} = e^{-\lambda} e^{\lambda} = e^{-\lambda+\lambda} = e^{0} = 1$$

That is , $p(x)$ **satisfies** the conditions of being a pmf of a discrete type of random variable.

A random variable that has a pmf of the form $p(x)$ given by eq.(1) is said to have a **Poisson distribution** with parameter λ,

and any such $p(x)$ is called a Poisson pmf with parameter m.

An interesting property:

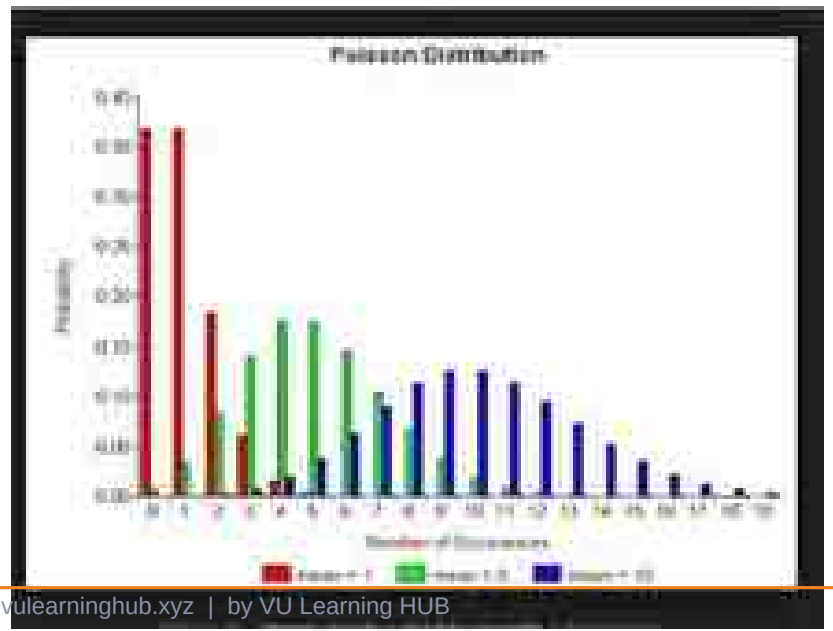
It is easy to prove that the **mean** and **variance** of the Poisson distribution given by (1) are **both equal to λ.**

Shape of the distribution

for Lambda=1, 5 & 10

i.e. expected number of occurrences = 1, 5 & 10

https://www.google.com.pk/search?q=poisson+distribution&rlz=1C1CHBD_enPK785PK785&source=lnms&tbm=isch&sa=X&ved=0ahUKEwj8LTtjNbeAhUUSo8KHZYbDWAQ_AUIDigB&biw=1280&bih=657#imgrc=qHHl3PwzE_SjaM:



When do we apply this formula?

Poisson distribution arrives in two situations:

1. It appears as the limiting form of the binomial distribution under certain conditions.
2. When we apply the Poisson distribution is that real-life situation where we are dealing with Poisson process

Poisson Process

- Whenever a particular event of interest occur randomly over a time-scale, we say that we are dealing with a Poisson process.

For example, for the Virtual University of Pakistan there is central telephone exchange, and we don't know that when the telephone operator will receive the next call, we from the past experience know that on the average 3 calls are received per minute.

But we do not know that in any one particular span of 1 min exactly how many calls were received? So on

Other than this information, the timings of incoming calls seem to be totally random. Thus, we conclude that the Poisson process might be a good model for the incoming calls at this telephone exchange.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 142

Poisson Process

The Poisson Process

The Poisson process is one of the most **widely**-used **counting** processes.

It is usually used in scenarios where we are counting the occurrences of certain events that appear to happen at a certain **rate**, but completely at random (i.e. the events that are happening without any particular structure).

Examples:

When events occur randomly over a time scale:

- From historical data, we know that earthquakes occur in a certain area with a **rate** of 22 per month.

  (Other than this information, the **timings** of earthquakes seem to be **completely random.**)

- From the record of the past six months, we know that, on a particular web server, requests for individual documents occur at the rate of 3 per hour;

  (Other than this information, the timings of the requests seem to be **totally random**.)

A process in which events occur randomly either over a time scale or over a distance scale.

In other words,

a process in which events occur randomly in a specified duration of time or in a specified portion of one-dimensional, two-dimensional or three-dimensional space.

Examples:

- From the record of the past six months, we know that, on a particular web server, requests for individual documents occur at the rate of 3 per hour;

  (Other than this information, the timings of the requests seem to be **totally random**.)

Examples:

When events occur randomly in two-dimensional space:

- *From past experience, it is known that, in a certain urban area, car accidents occur at the rate of 3 per square mile;* (Other than this information, the **locations** of the car accidents seem to be **completely random.**)

- *At a certain textile factory, it is known that flaws in a particular fabric produced in this factory occur at the rate of 2 per square meter;* (Other than this information, the **locations** of the flaws seem to be **completely random.**)

MuLearningHub.xyz | By MuLearning HUB

The important point to be noted is that, if we are dealing with a Poisson process with occurrence rate λ (i.e. λ occurrences in one unit of time),

and

X represents the number of occurrences in t units of time,

then

the probability of having x occurrences in t units of time is given by

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda t}(\lambda t)^x}{x!} & x = 0, 1, 2, \ldots, \ \lambda > 0 \\ 0 & elsewhere, \end{cases} \qquad (1)$$

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 143

Application of Poisson Process

EXAMPLE

If it is known that, on the average, **two** minor accidents occur at a busy crossing --- per day, what is the probability that, **at the most two** minor accidents will occur during an entire week ?

Solution:

This is a Poisson process with Occurrence Rate = λ = 2 per day.

Also, one week = 7 days implies t=7

Hence

So that $\lambda t = 2 \times 7 = 14$

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda t}(\lambda t)^x}{x!} = \dfrac{e^{-14}(14)^x}{x!} & x = 0,1,2,\ldots \\ 0 \quad elsewhere, \end{cases} \qquad (1)$$

Now, we require X<2.

So

$$P(X \le 2) = P(X = 0 \text{ or } X = 1 \text{ or } X = 2)$$

$$= \frac{e^{-14}(14)^0}{0!} + \frac{e^{-14}(14)^1}{1!} + \frac{e^{-14}(14)^2}{2!}$$

$$= e^{-14} + 14e^{-14} + 98e^{-14} = 113e^{-14}$$

$$= 113 / 1202592.9591 = 0.00009396$$

We can say that the probability that at the most 2 accidents will occur at this particular crossing in a week, this probability is almost 0, which is “Next to Impossible!”

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 144

Mean, Variance and Coefficient of Variation of the Poisson Distribution (explained through an example)

The Poisson pmf is given by

$$p(x) = \begin{cases} \dfrac{\lambda^x e^{-\lambda}}{x!} & x = 0, 1, 2, \ldots \\ 0 & \textit{elsewhere.} \end{cases}$$

It is easy to prove that, for a Poisson distribution, **mean = variance =** $\lambda > 0$

On this account, a Poisson pmf is frequently written as

$$p(x) = \begin{cases} \dfrac{\mu^x e^{-\mu}}{x!} & x = 0, 1, 2, \ldots \\ 0 & \textit{elsewhere.} \end{cases}$$

In other words, the parameter λ **or** μ in a Poisson pmf represents the **mean as well as the variance of the distribution.**

Hence, the Coefficient of Variation of the Poisson distribution is given by

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{\lambda}}{\lambda} \times 100\% = \frac{1}{\sqrt{\lambda}} \times 100\%$$

Example 1:

Suppose that X has a Poisson distribution with $\lambda = 4.$

Then the pmf of X is

$$p(x) = \begin{cases} \dfrac{0.5^x e^{-0.5}}{x!} & x = 0, 1, 2, \ldots \\ 0 & elsewhere. \end{cases}$$

vulearninghub.xyz by VU Learning HUB

The mean of this distribution is 4.

&

The variance of this distribution is 4.

Therefore,

The **coefficient of variation** of this distribution is

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{4}}{4} \times 100\% = \frac{1}{2} \times 100\% = 50\%$$

Example 2:

Suppose that Y has a Poisson distribution with $\lambda = 9.$

Then the pmf of Y is

$$p(y) = \begin{cases} \dfrac{2^y e^{-2}}{y!} & y = 0,1,2,\ldots \\ 0 & \textit{elsewhere.} \end{cases}$$

The mean of this distribution is 3.

&

The variance of this distribution is 3.

Therefore,

The **coefficient of variation** of this distribution is

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{9}}{9} \times 100\% = \frac{1}{\sqrt{9}} \times 100\% = 33.33\%$$

Example 3:

Suppose that Z has a Poisson distribution with $\lambda = 16.$

Then the pmf of Z is

$$p(z) = \begin{cases} \dfrac{5^z e^{-5}}{z!} & z = 0, 1, 2, \ldots \\ 0 & elsewhere. \end{cases}$$

The mean of this distribution is 16.

&

The variance of this distribution is 16.

Therefore,

The **coefficient of variation** of this distribution is

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{\lambda}}{\lambda} \times 100\% = \frac{1}{\sqrt{16}} \times 100\%$$

$$= \frac{1}{4} \times 100\% = 25\%$$

It appears that, as λ increases, the coefficient of variation of the Poisson distribution decreases.

Stated a little more rigorously:

For a Poisson distribution with parameter λ, as λ tends to infinity, the coefficient of variation of the Poisson distribution also tends to 0.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 145

Derivation of the Mean of the Poisson distribution

If the random variable X has a Poisson distribution with parameter μ,

$$p(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

then its mean and variance are given by $E(X) = \mu$ and $\mathrm{Var}(X) = \mu$.

Proof:

By definition,

$$\text{Mean} = E(X) = \sum_{x=0}^{\infty} x.p(x;\mu) \quad \text{where} \quad p(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

$$= 0.\frac{\mu^{0}}{0!}e^{-\mu} + 1.\frac{\mu^{1}}{1!}e^{-\mu} + 2.\frac{\mu^{2}}{2!}e^{-\mu} + 3.\frac{\mu^{3}}{3!}e^{-\mu} + 4.\frac{\mu^{4}}{4!}e^{-\mu} + \ldots$$

$$= 0 + 1.\frac{\mu^{1}}{1!}e^{-\mu} + 2.\frac{\mu^{2}}{2!}e^{-\mu} + 3.\frac{\mu^{3}}{3!}e^{-\mu} + 4.\frac{\mu^{4}}{4!}e^{-\mu} + \ldots$$

vulearninghub.xyz | by VU Learning HUB

$$\text{Mean} = E(X) = \sum_{x=0}^{\infty} x.p(x;\mu) \quad \text{where} \quad p(x;\mu) = \frac{e^{-\mu}\mu^{x}}{x!}$$

$$= 0 + 1.\mu e^{-\mu} + 2.\frac{\mu^2}{1\times 2}e^{-\mu} + 3.\frac{\mu^3}{1\times 2\times 3}e^{-\mu} + 4.\frac{\mu^4}{1\times 2\times 3\times 4!}e^{-\mu} + \ldots$$

$$= \mu e^{-\mu} + \frac{\mu^2}{1}e^{-\mu} + \frac{\mu^3}{1\times 2}e^{-\mu} + \frac{\mu^4}{1\times 2\times 3!}e^{-\mu} + \ldots$$

$$= \mu e^{-\mu}\left[1 + \mu + \frac{\mu^3}{2!} + \frac{\mu^4}{3!} + \ldots\right]$$

But we know that the series

$$1+m+\frac{m^3}{2!}+\frac{m^4}{3!}+\ldots=\sum_{x=0}^{\infty}\frac{m^x}{x!}$$

Converges, for all values of m, e^{μ}

Therefore

$$\text{Mean} = \mu e^{-\mu}\left[1+\mu+\frac{\mu^3}{2!}+\frac{\mu^4}{3!}+\ldots\right]$$

$$= \mu e^{-\mu}.e^{\mu} = \mu$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 146

Derivation of the Variance of the Poisson distribution

If the random variable X has a Poisson distribution with parameter μ, then its mean and variance are given by $E(X)= \mu$ and $\text{Var}(X)= \mu$.

Proof:

By definition

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

$$E(X^2) = E[X(X-1) + X] = E(X) + E[X(X-1)]$$

$$= \sum_{x=0}^{\infty} x.\frac{e^{-\mu}\mu^x}{x!} + \sum_{x=0}^{\infty} x(x-1)e^{-\mu}.\frac{\mu^x}{x!}$$

$$E\left(X^2\right)$$

$$= \sum_{x=0}^{\infty} x.\frac{e^{-\mu}\mu^x}{x!} + \sum_{x=0}^{\infty} x(x-1)e^{-\mu}.\frac{\mu^x}{x!}$$

$$= \mu + e^{-\mu}\sum_{x=0}^{\infty} x(x-1).\frac{\mu^x}{x(x-1)(x-2)!}$$

$$= \mu + \mu^2 e^{-\mu}\sum_{x=2}^{\infty}\frac{\mu^x\mu^{-2}}{(x-2)!} = \mu + \mu^2 e^{-\mu}\sum_{x=2}^{\infty}\frac{\mu^{x-2}}{(x-2)!}$$

$\left(x \text{ starts at } 2, \text{as the first two terms in the summation are zero.}\right)$

Let $y = x - 2,$ then

$$E\left(X^2\right) = \mu + \mu^2 e^{-\mu} \sum_{y=0}^{\infty} \frac{\mu^y}{y!} \qquad (y = 0,1,2,\ldots)$$

$$= \mu + \mu^2 e^{-\mu} e^{\mu} = \mu + \mu^2 \quad \begin{pmatrix} \text{when } x \to 2, y \to 0, \\ \text{and } x \to \infty, y \to \infty, \end{pmatrix}$$

Hence $Var\left(X\right) = \mu + \mu^2 - \mu^2 = \mu$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

Topic No. 147

MGF of Poisson Distribution

MGF of Poisson Distribution

The mgf of a Poisson distribution is given by

$$M(t)=\sum_{x} e^{tx} p(x)=\sum_{x=0}^{\infty} e^{tx}\,\frac{m^{x}e^{-m}}{x!}=e^{-m}\sum_{x=0}^{\infty}\frac{\left(me^{t}\right)^{x}}{x!}=e^{-m}e^{me^{t}}=e^{m\left(e^{t}-1\right)}$$

for all real values of t.

Since $$M'(t)=e^{m\left(e^{t}-1\right)}\left(me^{t}\right)$$

vulearninghub.xyz | by VU Learning HUB

Since $M'(t) = e^{m(e^t - 1)}\left(me^t\right)$

and $M''(t) = e^{m(e^t - 1)}\left(me^t\right) + e^{m(e^t - 1)}\left(me^t\right)^2,$

then

$$\mu = M'(0) = m$$

and

$\sigma^2 = M''(0) - \mu^2 = m + m^2 - m^2 = m.$

That is, a Poisson distribution has $\mu = \sigma^2 = m > 0.$

By definition, the mgf of a discrete distribution is given by

$$M(t) = \sum_{x} e^{tx} p(x)$$

for all real values of t.

Hence, the mgf of a Poisson distribution is given by

$$M(t) = \sum_x e^{tx} p(x) = \sum_{x=0}^{\infty} e^{tx} \frac{m^x e^{-m}}{x!} = e^{-m} \sum_{x=0}^{\infty} \frac{\left(me^t\right)^x}{x!} = e^{-m} e^{me^t} = e^{m\left(e^t - 1\right)}$$

for all real values of t.

Since $M'(t) = e^{m\left(e^t - 1\right)}\left(me^t\right)$

$put\, t = 0$

$M'(0) = e^{m\left(e^0 - 1\right)}\left(me^0\right) = e^{m(0)}\left(m.1\right) = 1.\left(m\right) = m$

and $M''(t) = e^{m\left(e^t - 1\right)}\left(me^t\right) + e^{m\left(e^t - 1\right)}\left(me^t\right)^2,$

Therefore

$$\mu = M'(0) = m$$

$$M''(t) = e^{m\left(e^t - 1\right)}\left(me^t\right) + e^{m\left(e^t - 1\right)}\left(me^t\right)^2$$

$$put\, t = 0$$

$$M''(0) = e^{m\left(e^0 - 1\right)}\left(me^0\right) + e^{m\left(e^0 - 1\right)}\left(me^0\right)^2$$

$$= e^{m(0)}\left(m.1\right) + e^{m(0)}\left(m.1\right)^2 = 1.\left(m\right) + 1.\left(m\right)^2$$

Therefore

$$M''(0) = m + m^2.$$

vulearninghub.xyz by VU Learning HUB

Hence

$$\sigma^2 = M''(0) - \mu^2 = \left(m + m^2\right) - m^2$$

$$= m + m^2 - m^2 = m.$$

That is, a Poisson distribution has $\mu = \sigma^2 = m > 0$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 148

Derivation of

Poisson Approximation to the Binomial distribution

To derive an approximation formula to the binomial distribution $b(x\,;\,n\,,p)$ when $n \rightarrow \infty$, $p \rightarrow 0$, and the product np remains **constant** we proceed :

The binomial distribution $b(x\,;\,n\,,p)$ may be written as

$$b\left(x;n,p\right)=\binom{n}{x}p^{x}q^{n-x},\qquad \text{for } x=0,1,\ldots,n.$$

$$b\left(x;n,p\right)=\frac{n!}{x!(n-x)!}p^{x}q^{n-x}$$

$$=\frac{n\left(n-1\right)\left(n-2\right)\ldots\left(n-x+1\right)(n-x)!}{x!(n-x)!}p^{x}q^{n-x}$$

$$=\frac{n\left(n-1\right)\left(n-2\right)\ldots\left(n-x+1\right)}{x!}p^{x}q^{n-x}\qquad \ldots(1)$$

Now, let $np = \mu$.

Then $p = \frac{\mu}{n}$ and $q = 1 - p = 1 - \frac{\mu}{n}$.

Substituting in (1), we get

$$b(x; n, p) = \frac{n(n-1)(n-2)\ldots(n-x+1)}{x!}\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}$$

Now

$$b(x;n,p)=\frac{n(n-1)(n-2)...(n-x+1)}{x!}\left(\frac{\mu}{n}\right)^{x}\left(1-\frac{\mu}{n}\right)^{n-x}$$

can be re-written as:

$$b(x;n,p)=\frac{\mu^{x}}{x!}.\frac{n(n-1)(n-2)...(n-(x-1))}{n^{x}}\left(1-\frac{\mu}{n}\right)^{n}\left(1-\frac{\mu}{n}\right)^{-x}$$

$$=\frac{\mu^{x}}{x!}.\left[\frac{n}{n}.\left(\frac{n-1}{n}\right)\left(\frac{n-2}{n}\right)...\left(\frac{n-(x-1)}{n}\right)\right]\left(1-\frac{\mu}{n}\right)^{n}\left(1-\frac{\mu}{n}\right)^{-x}$$

$$=\frac{\mu^{x}}{x!}.\left[1.\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{x-1}{n}\right)\right]\left(1-\frac{\mu}{n}\right)^{-x}\left(1-\frac{\mu}{n}\right)^{n}$$

Letting $n \rightarrow \infty$ and $p \rightarrow 0$ such that $np = \mu$ remain **constant,** we observe that each of the terms

$$\left(1-\frac{1}{n}\right),\left(1-\frac{2}{n}\right),\ldots,\left(1-\frac{x-1}{n}\right) and \left(1-\frac{\mu}{n}\right)^{-x}$$

approaches unity.

Let us now focus on the term $\left(1-\frac{\mu}{n}\right)^{n}$

which may be written as $\left[\left(1-\frac{\mu}{n}\right)^{n/\mu}\right]^{\mu}$.

Letting $\frac{n}{\mu}=k,$

the expression becomes $\left[\left(1-\frac{1}{k}\right)^{k}\right]^{\mu}$.

Now, as n increases indefinitely, so does k.

So,

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n} = \lim_{n\to\infty}\left[\left(1-\frac{\mu}{n}\right)^{n/\mu}\right]^{\mu} = \lim_{k\to\infty}\left[\left(1-\frac{1}{k}\right)^{k}\right]^{\mu}$$

$$= \lim_{k\to\infty}\left[\left(1-\frac{1}{k}\right)^{k}\left(1-\frac{1}{k}\right)^{k}....\left(1-\frac{1}{k}\right)^{k}\right] \ (\mu \text{ times})$$

Now, we know that

limit of the product = product of the limits

so that

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}=\lim_{k\to\infty}\left[\left(1-\frac{1}{k}\right)^{k}\left(1-\frac{1}{k}\right)^{k}....\left(1-\frac{1}{k}\right)^{k}\right]\ (\mu \text{ times})$$

$$=\left[\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}....\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}\right]\ (\mu \text{ times})$$

Now, it is well-known that $\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}=e^{-1}$ where $e=2.71828.$

Therefore

$$\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}=\left[\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}....\lim_{k\to\infty}\left(1-\frac{1}{k}\right)^{k}\right]\ (\mu \text{ times})$$

i.e. $\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}=\left[e^{-1}e^{-1}....e^{-1}\right]\ (\mu \text{ times})$

or $\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n}=e^{-\mu}.$

Thus the limiting value of $P(X=x)$ is given by the expression

$$\lim_{n\to\infty} b\left(x;n,p\right)=\frac{\mu^{x}}{x!}.1.1...1.e^{-\mu}=\frac{\mu^{x}e^{-\mu}}{x!}$$

$$\text{for } x=0,1,...,\infty$$

In other words, if X is a binomial r.v. such that

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \text{ then}$$

$$\lim_{\substack{n\to\infty \\ np=\mu}} P(X = x) = \frac{\mu^x e^{-\mu}}{x!}, \qquad \text{for } x = 0,1,2,\ldots,\infty$$

It is denoted by $p(x;\mu)$ and the r.v. X having p.d. $p(x;\mu)$ is said to have a Poisson distribution with parameter μ.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 149

An Example of the Poisson Approximation to the Binomial distribution

- To derive an approximation formula to the binomial distribution $b(x;n, p)$ where $n \rightarrow \infty$, $p \rightarrow 0$, and the product np remains **constant**, we proceed:

Example:

Suppose 8% of the tires manufactured at a particular plant are defective. To illustrate the use of the Poisson approximation for the binomial, the probability of obtaining exactly one defective tire from a sample of 20 using equation (5.18) is calculated as follows:

$$P(X=1) \equiv \frac{e^{-(20)(.08)}\left[(20)(.08)\right]^1}{1!} = \frac{e^{-1.6}\left[1.6\right]^1}{1!} = 0.3230$$

vulearninghub.xyz [by VU Learning HUB

Had the true distribution, the binomial, been used instead of the approximation,

$$P(X=1) \equiv \binom{20}{1}(.08)^{1}(.92)^{19} = 0.3282$$

To illustrate the use of the Poisson approximation for the binomial, the probability of obtaining exactly one defective tire from a sample of 20 using the formula is calculated as follows:

$$P(X=1) \equiv \frac{e^{-(20)(.08)}\left[(20)(.08)\right]^1}{1!} = \frac{e^{-1.6}\left[1.6\right]^1}{1!} = 0.3230$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

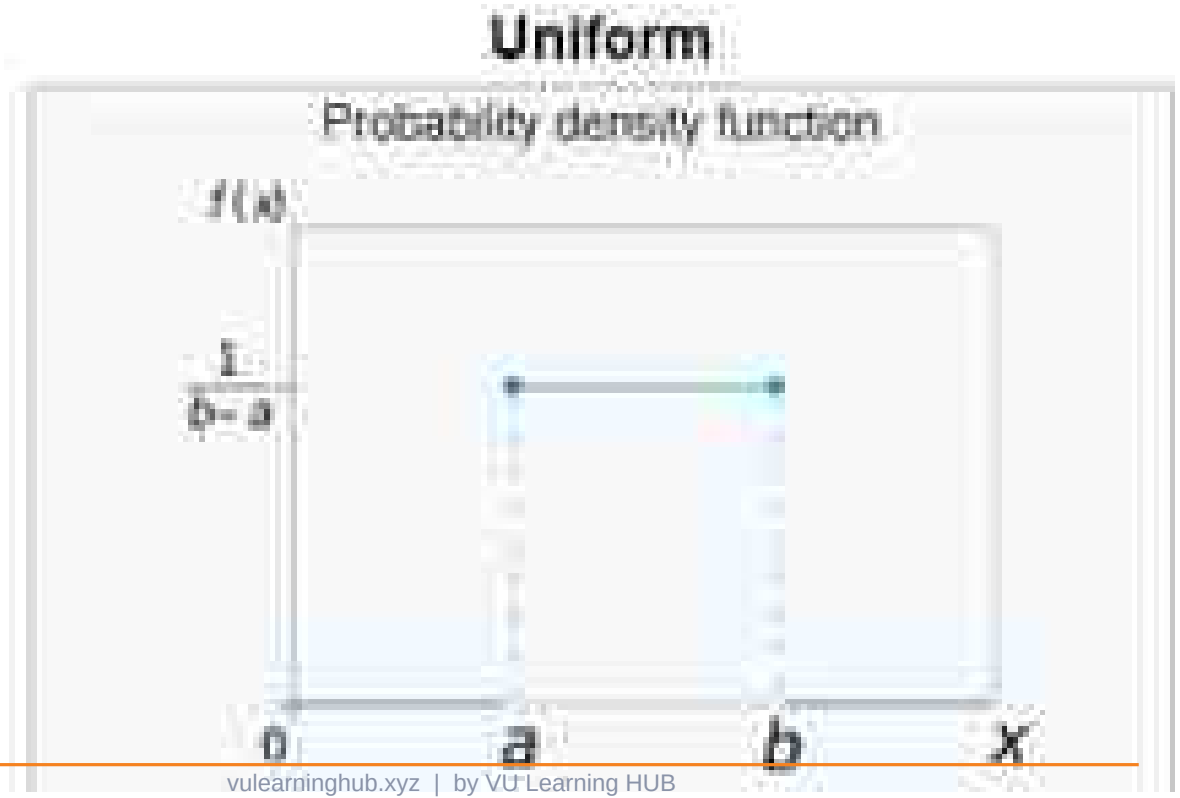
Topic No. 150

Continuous Uniform Distribution (Rectangular Distribution) (PDF, CDF and Shape of the distribution)

Probability Distribution Function

The probability density function of the continuous uniform distribution is:

$$f(x) = \begin{cases} \dfrac{1}{b-a} & for\ \ a \le x \le b, \\ 0 & for\ \ x < a\ \ or\ \ x > b \end{cases}$$


Uniform
Probability density function
f(x)
1
b− a
0
a
b
x

Cumulative Distribution Function

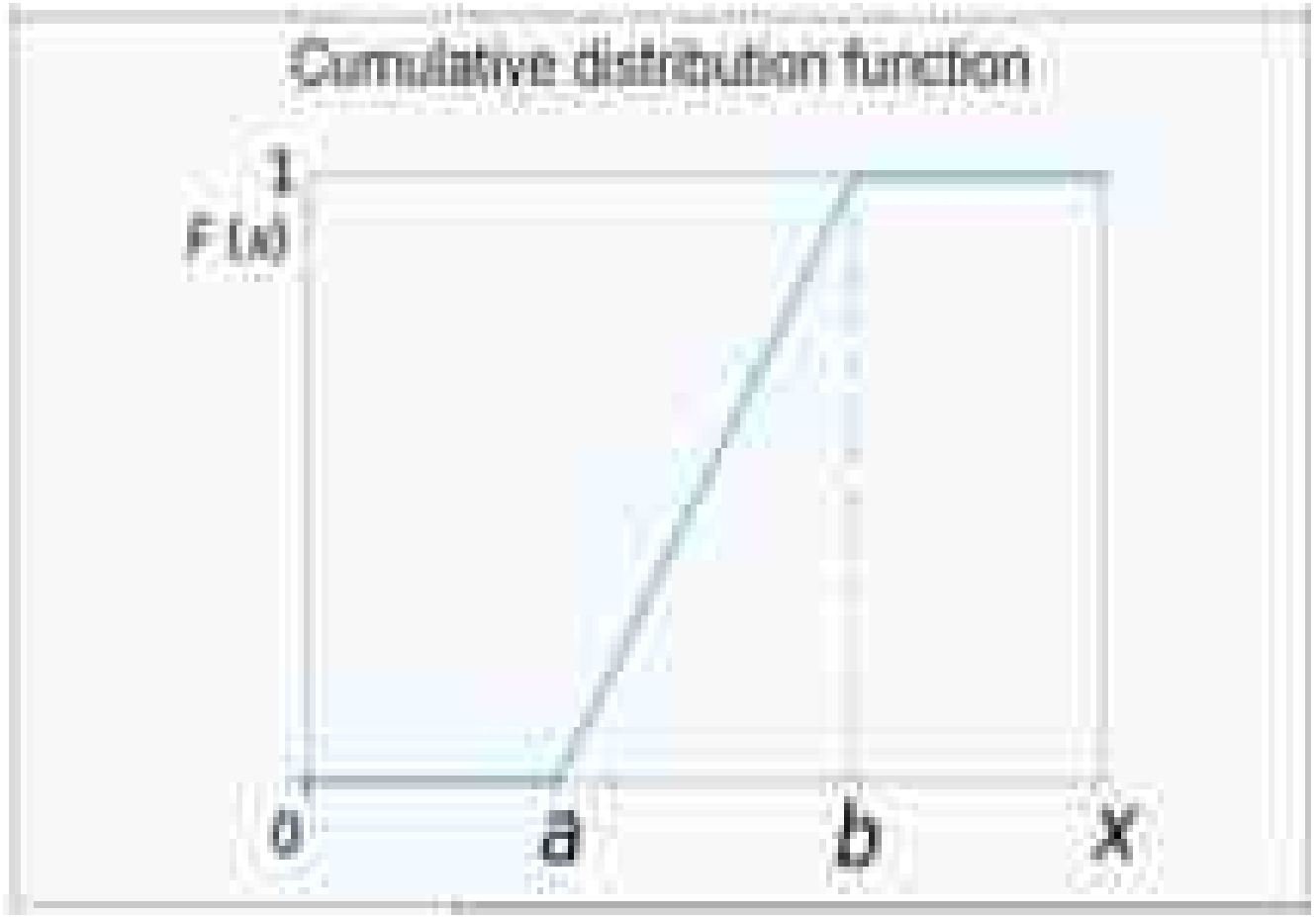
The Cumulative Distribution function of the continuous uniform distribution is:

$$F(x) = \begin{cases} 0 & for\ \ x < a \\ \dfrac{x-a}{b-a} & for\ \ a \le x \le b, \\ 1 & for\ \ x > b \end{cases}$$

Derivation of the expression for the interval $a \le x \le b$:

$$F(x) = \int_a^x f(x)\,dx = \int_a^x \frac{1}{b-a}\,dx = \frac{1}{b-a}\int_a^x 1\,dx$$

$$= \frac{1}{b-a}\left[x\right]_a^x = \frac{1}{b-a}\left[x-a\right] = \frac{x-a}{b-a}$$


Cumulative distribution function
1
F (x)
0
a
b
x

Why is it a straight line between a and b ?

$$F(x) = \frac{x-a}{b-a} = \frac{x}{b-a} - \frac{a}{b-a}$$

or

$$F(x) = \left(\frac{1}{b-a}\right)x + \left(-\frac{a}{b-a}\right)$$

which is of the form $y = mx + c$.

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 151

Derivation of the Mean and Variance of the Continuous Uniform distribution

Expectation and Variance

If $X \sim U(a\ ,\ b)$, then:

$$E(X) = \frac{(a+b)}{2}$$

$$\text{Var}(X) = \frac{(b-a)^2}{12}$$

Proof of Expectation:

The probability density function of the continuous uniform distribution is given by

$$f(x)=\begin{cases}\dfrac{1}{b-a} & for\ \ a\le x\le b,\\ 0 & for\ \ x<a\ \ or\ \ x>b\end{cases}$$

Therefore

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx = \int_{a}^{b} x\left(\frac{1}{b-a}\right)dx = \frac{1}{b-a}\int_{a}^{b} xdx$$

$$= \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)}$$

$$= \frac{b+a}{2} = \frac{a+b}{2}$$

Proof of Variance:

By definition,

$$\text{Var}(X) = E(X^2) - [E(X)]^2.$$

Now,

$$E(X^2) = \int_a^b x^2 \left(\frac{1}{b-a}\right) dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$$

$$= \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(a^2+ab+b^2)}{3(b-a)} = \frac{a^2+ab+b^2}{3}$$

Therefore,

$$\text{Var}(X) = \frac{a^2 + ab + b^2}{3} - \left(\frac{a+b}{2}\right)^2$$

$$= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} = \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab - 3b^2}{12}$$

$$= \frac{a^2 - 2ab + b^2}{12} = \frac{(b-a)^2}{12}.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 152

Application

of the Uniform Distribution

(explained through an example)

Example:

Suppose in a quiz there are 30 participants.

A question is given to all 30 participants and the time allowed to answer it is 25 seconds.

i) Find the probability that a participant will respond within 6 seconds.

ii) How many of the 30 participants can be expected to respond within 6 seconds?

vulearninghub.xyz | by VU Learning HUB

Solution:

Regarding **time taken to answer the quiz question** as the variable of interest X, we have:

Interval of probability distribution = [0 seconds, 25 seconds]

i.e. $0 \leq x \leq 25$

and,

assuming that X is **uniformly** distributed,

$f(x) = 1/(25-0) = 1/25$

i) Since we are interested in the probability of participants' responding within 6 seconds, therefore we need to compute $P(X < 6) = F(6)$.

Now, for $a \le x \le b$, the Cumulative Distribution Function of the continuous uniform distribution is $F(x) = \frac{x-a}{b-a}$

So, here $F(6) = \frac{6-a}{b-a} = \frac{6-0}{25-0} = \frac{6}{25}$.

vulearninghub.xyz | by VU Learning HUB

ii) There are 30 participants in the quiz

Hence the number of participants likely to answer it in 6 seconds = $(6/25) \times 30 \approx 7$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 153

Exponential Distribution
(PDF, CDF and shape of the distribution)

The Probability density function (pdf) of an exponential distribution is

$$f\left(x;\lambda\right)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\ \ \lambda>0\\ 0 & x<0.\end{cases}$$

So it is obvious that is a **single-parameter** distribution.

Let $\theta = \frac{1}{\lambda}$

then,

mean $= \theta$

and

$$f\left(x;\theta\right) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}} & x \geq 0,\ \theta > 0 \\ 0 & x < 0. \end{cases}$$

vdlearninghub.xyz | by Vdl Learning HUB

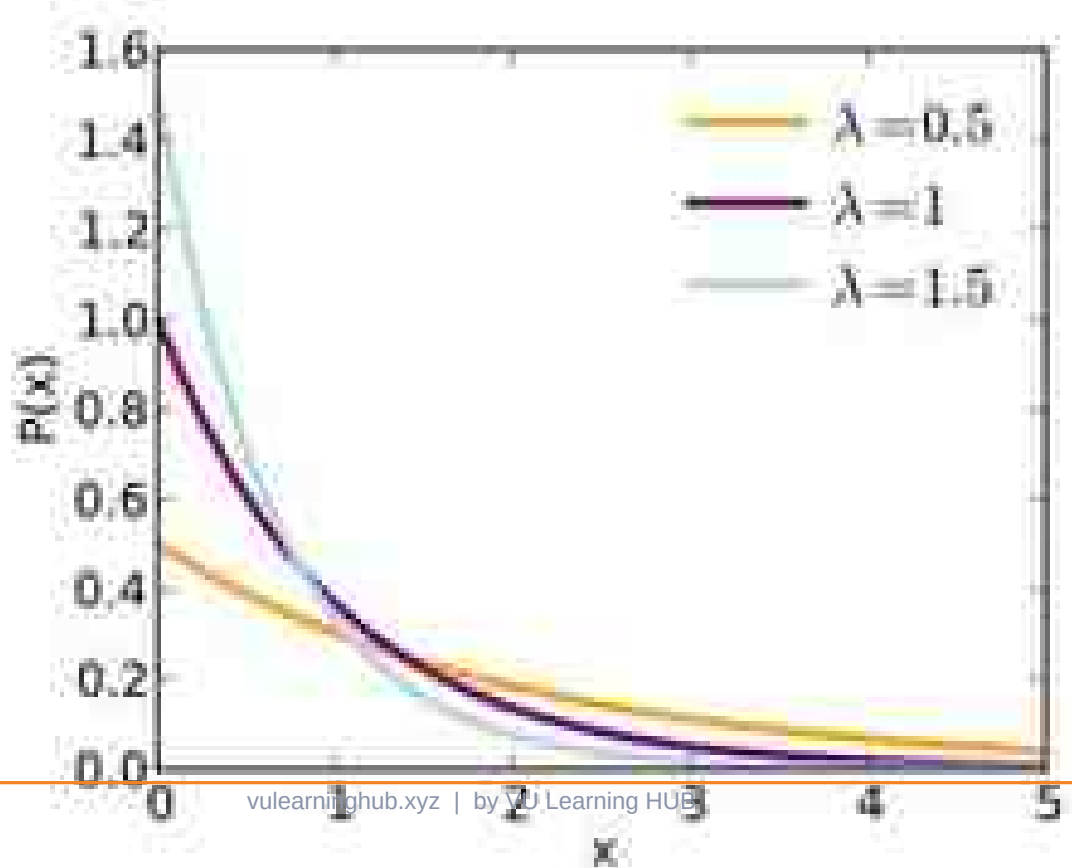
So it is obvious that is a **single-parameter** distribution.

First and foremost, let us consider the shape of the distribution:

PDF of the **Exponential Distribution**

with mean= θ = 0.5, 1, 1.5

- $f(x) = \left(\frac{1}{0.5}\right) e^{-\left(\frac{x}{0.5}\right)}, \quad 0 < x < \infty$ → orange
- $f(x) = e^{-(x)}, \quad 0 < x < \infty$ → purple
- $f(x) = \left(\frac{1}{1.5}\right) e^{-\left(\frac{x}{1.5}\right)}, \quad 0 < x < \infty$ → skyblue


1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
P(x)
0
1
2
3
4
5
x
λ=0.5
λ=1
λ=1.5

Next: CDF

The Cumulative distribution function (cdf) of an exponential distribution is

$$F(x;\lambda)=\begin{cases}1-e^{-\lambda x} & x\geq 0,\ \ \lambda>0\\ 0 & x<0.\end{cases}$$

The Cumulative distribution function (cdf) of an exponential distribution in terms of θ

$$F\left(x;\theta\right)=\begin{cases}1-e^{-\frac{x}{\theta}} & x\geq 0,\ \ \theta>0\\ 0 & x<0.\end{cases}$$

Proof:

$$F(x) = \int_{-\infty}^{x} f(x)\,dx = \int_{0}^{x} \lambda e^{-\lambda x}\,dx = -\int_{0}^{x} (-\lambda) e^{-\lambda x}\,dx$$

$$= -\left[e^{-\lambda x} \right]_0^x = -\left[e^{-\lambda x} - e^0 \right] = -\left[e^{-\lambda x} - 1 \right]$$

or

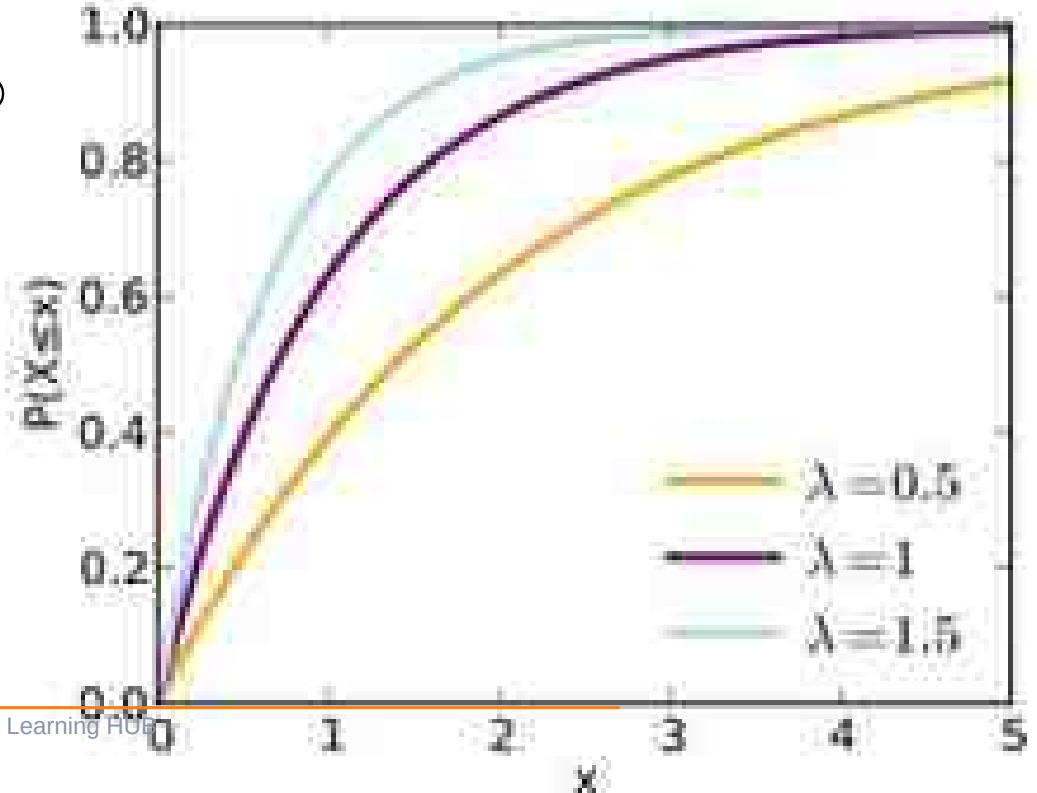
$$F(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

Now, let us consider the shape of the CDF:

CDF of the **Exponential Distribution** with mean=0.5, 1, 1.5

$$F(x) = 1 - e^{-x}, \quad 0 < x < \infty$$

- $F(x) = 1 - e^{-\frac{x}{0.5}}, \quad 0 < x < \infty$
- $F(x) = 1 - e^{-x}, \quad 0 < x < \infty$
- $F(x) = 1 - e^{-\frac{x}{1.5}}, \quad 0 < x < \infty$



Important Note:

- If we are dealing with Poisson process in which events are occurring randomly over a time scale then the **waiting time** between successive events is distributed according to the exponential distribution.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 154

Derivation of the Mean of the Exponential Distribution

A very interesting property:

- The hazard rate of this distribution is constant.

- The **mean and standard deviation** of the exponential distribution are equal.

1. Derivation of the **mean:**

The pdf of the Exponential distribution is given by

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0, \lambda > 0.$$

Therefore,

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_{0}^{\infty} x.\lambda e^{-\lambda x} dx$$

First of all we will apply the formula of integration by parts:

$$\int uvdx = u\int vdx - \int\left(\frac{du}{dx}\int vdx\right)dx$$

Here, we let $u = x$ and $v = \lambda e^{-\lambda x} = -(-\lambda e^{-\lambda x})$

so that $\int vdx = \int -(-\lambda e^{-\lambda x})dx = -\int(-\lambda e^{-\lambda x})dx = -\left(e^{-\lambda x}\right)$

and $\frac{du}{dx} = \frac{d}{dx}(x) = 1.$

Hence, we have

$$\int uvdx = u\int vdx - \int\left(\frac{du}{dx}\int vdx\right)dx$$

$$= x\left(-e^{-\lambda x}\right) - \int\left(1.\left(-e^{-\lambda x}\right)\right)dx = -xe^{-\lambda x} - \frac{1}{\lambda}\int\left(\lambda\left(-e^{-\lambda x}\right)\right)dx$$

$$= -xe^{-\lambda x} - \frac{1}{\lambda}\int\left(-\lambda e^{-\lambda x}\right)dx = -xe^{-\lambda x} - \frac{1}{\lambda}\left(e^{-\lambda x}\right)$$

Now, in order to find the mean, we will apply the limits 0 to ∞ to the expression we have obtained:

$$E(X) = \int_0^{\infty} x.\lambda e^{-\lambda x} dx = \left[-xe^{-\lambda x} \right]_0^{\infty} - \frac{1}{\lambda} \left[e^{-\lambda x} \right]_0^{\infty}$$

$$= -\left[xe^{-\lambda x} \right]_0^{\infty} - \frac{1}{\lambda} \left[\frac{1}{e^{\lambda x}} \right]_0^{\infty}$$

In order to apply the upper limit to the first expression, we need to utilize the L'Hospital Rule.

We have $xe^{-\lambda x} = \frac{x}{e^{\lambda x}}$.

Now, $\underset{x\to\infty}{\text{Lim}} \frac{x}{e^{\lambda x}} = \underset{x\to\infty}{\text{Lim}} \frac{1}{\lambda e^{\lambda x}} = \frac{1}{\lambda e^{\lambda(\infty)}} = \frac{1}{\lambda e^{(\infty)}} = \frac{1}{\lambda(\infty)} = 0.$

On the other hand, $\underset{x\to 0}{\text{Lim}} \frac{x}{e^{\lambda x}} = \frac{0}{e^{\lambda(0)}} = \frac{0}{e^{0}} = \frac{0}{1} = 0.$

Hence,

$$E(X) = -\left[\frac{x}{e^{\lambda x}}\right]_0^{\infty} - \frac{1}{\lambda}\left[\frac{1}{e^{\lambda x}}\right]_0^{\infty} = -(0-0) - \frac{1}{\lambda}\left[\frac{1}{e^{\lambda x}}\right]_0^{\infty}$$

$$= -0 - \frac{1}{\lambda}\left[\frac{1}{\infty} - \frac{1}{e^0}\right] = -\frac{1}{\lambda}\left[\frac{1}{\infty} - \frac{1}{e^0}\right] = -\frac{1}{\lambda}\left[0 - \frac{1}{1}\right]$$

$$= -\frac{1}{\lambda}[-1] = \frac{1}{\lambda}.$$

Thus, $E(X)$ *or* $\mu = \frac{1}{\lambda}$.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 155

Derivation of the Variance of the Exponential Distribution

A very interesting property:

The **mean and standard deviation** of the exponential distribution are equal.

We know that, for the exponential distribution given by

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0, \lambda > 0$$

the mean is

$$E(X) = \frac{1}{\lambda}$$

1. Derivation of the **Variance:**

According to the short-cut formula

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2,$$

Now

$$E\left(X^2\right)=\int_0^{\infty} x^2.f\left(x\right)dx=\int_0^{\infty} x^2.\lambda e^{-\lambda x}dx$$

First of all we will apply the formula of integration by parts:

$$\int uvdx = u\int vdx - \int\left(\frac{du}{dx}\int vdx\right)dx$$

Here, we let $u = x^2$ and $v = \lambda e^{-\lambda x} = -(-\lambda e^{-\lambda x})$

so that $\int vdx = \int -(-\lambda e^{-\lambda x})dx = -\left(e^{-\lambda x}\right)$

and $\frac{du}{dx} = \frac{d}{dx}\left(x^2\right) = 2x.$

Hence, we have

$$\int x^2.\lambda e^{-\lambda x}dx = u\int vdx - \int\left(\frac{du}{dx}\int vdx\right)dx$$

$$= x^2(-e^{-\lambda x}) - \int 2x.\left(-e^{-\lambda x}\right)dx = -x^2e^{-\lambda x} - 2\int\left(-xe^{-\lambda x}\right)dx$$

$$= -x^2e^{-\lambda x} - \frac{2}{\lambda}\int\left(\lambda\left(-xe^{-\lambda x}\right)\right)dx$$

$$= -x^2e^{-\lambda x} - \frac{2}{\lambda}\int x\left(-\lambda e^{-\lambda x}\right)dx$$

Hence, we have

$$\int x^2.\lambda e^{-\lambda x} dx = -x^2 e^{-\lambda x} + \frac{2}{\lambda}\int x\left(\lambda e^{-\lambda x}\right) dx$$

implying that

$$E\left(X^2\right) = \left[-x^2 e^{-\lambda x}\right]_0^{\infty} + \frac{2}{\lambda}E\left(X\right)$$

or

$$E\left(X^2\right) = -\left[\frac{x^2}{e^{\lambda x}}\right]_0^{\infty} + \frac{2}{\lambda}E\left(X\right)$$

In order to apply the upper limit to $\left[\frac{x^2}{e^{\lambda x}}\right]$, we need to utilize the L'Hospital Rule.

We have $\frac{x^2}{e^{\lambda x}}$

Now, $\lim_{x\to\infty}\frac{x^2}{e^{\lambda x}} = \lim_{x\to\infty}\frac{2x}{\lambda e^{\lambda x}} = \lim_{x\to\infty}\frac{2}{\lambda^2 e^{\lambda x}} = \frac{2}{\lambda^2 e^{\lambda(\infty)}} = \frac{2}{\infty} = 0.$

On the other hand, $\lim_{x\to 0}\frac{x^2}{e^{\lambda x}} = \frac{(0)^2}{e^{\lambda(0)}} = \frac{0}{e^0} = \frac{0}{1} = 0.$

Muzaamilhub.xyz | by VQ Learning HUB

Hence, we have

$$E\left(X^2\right) = -\left[\frac{x^2}{e^{\lambda x}}\right]_0^{\infty} + \frac{2}{\lambda}E\left(X\right)$$

$$= -\left[0-0\right] + \frac{2}{\lambda}E\left(X\right)$$

$$= -0 + \frac{2}{\lambda}E\left(X\right) = \frac{2}{\lambda}E\left(X\right) = \frac{2}{\lambda}\left(\frac{1}{\lambda}\right) = \frac{2}{\lambda^2}$$

Hence,

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{2-1}{\lambda^2} = \frac{1}{\lambda^2}$$

implying that

$$S.D.(X) = \sigma = \frac{1}{\lambda}.$$

Hence we have the interesting property that the **mean and standard deviation** of the exponential distribution are equal.

This, in turn, means that **the Coefficient of Variation of the Exponential distribution is always equal to 1.**

Explanation:

The coefficient of variation is defined as $CV = \frac{\sigma}{\mu}$.

Hence, for the exponential distribution given by

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

the coefficient of variation is

$$CV = \frac{\sigma}{\mu} = \frac{1/\lambda}{1/\lambda} = \frac{\lambda}{\lambda} = 1.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 156

Application of the Exponential Distribution (explained through an example)

Example:

The duration of long-distance telephone calls is found to be exponentially distributed with a mean of 3 minutes. What is the probability that a call will last

(i) more than 3 minutes, (ii) more than 5 minutes?

Let X be the exponential r.v. with parameter θ.

Then the mean, i.e. $\frac{1}{\theta} = 3$, so that $\theta = \frac{1}{3}$.

and the pdf can be written as

$$f(x) = \int_{3}^{\infty} \left(\frac{1}{\theta} \right) e^{-x/\theta} dx$$

Now (i) the probability that a call will last more than 3 minutes, is given by

$$P(X > 3) = \int_3^{\infty} \left(\frac{1}{3}\right) e^{-x/3} dx = \left[-e^{-x/3}\right]_3^{\infty} = e^{-1} = 0.3679.$$

And (ii) the probability that a call will last more than 5 minutes, is given by

$$P(X > 5) = \int_5^{\infty} \left(\frac{1}{3}\right) e^{-x/3} dx = \left[-e^{-x/3}\right]_5^{\infty} = e^{-1.7} = 0.1827.$$

We say that, the probability of having call duration longer than 5 minutes its probability is almost half of the probability of getting a call duration greater than three minutes.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 157

MGF of the Exponential Distribution

We know that the probability density function (pdf) of an exponential distribution is

$$f\left(x;\lambda\right)=\begin{cases}\lambda e^{-\lambda x} & x\geq 0,\ \ \lambda>0\\ 0 & x<0.\end{cases}$$

where λ is the only parameter of the distribution.

So to find the MGF, we proceed as follows:

$$M_0(t) = E\left[e^{tX}\right] = \int_0^\infty e^{tx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{-(\lambda-t)x}dx$$

$$= \lambda\left[\frac{e^{-(\lambda-t)x}}{-(\lambda-t)}\right]_0^\infty = \lambda\left[\frac{-e^{-(\lambda-t)x}}{\lambda-t}\right]_0^\infty$$

Now, if $t < \lambda$ then $\lambda - t > 0$

and we have

$$M_0(t) = \lambda\left[\frac{-e^{-(\lambda-t)\infty}}{\lambda-t} - \frac{-e^{-(\lambda-t)0}}{\lambda-t}\right] = \lambda\left[\frac{-e^{-\infty}}{\lambda-t} - \frac{-e^{0}}{\lambda-t}\right]$$

$$= \lambda\left[\frac{0}{\lambda-t} - \frac{-1}{\lambda-t}\right] = \lambda\left[0 + \frac{1}{\lambda-t}\right] = \frac{\lambda}{\lambda-t}$$

vulearninghub.xyz | by VU Learning HUB

Hence, the MGF of the Exponential Distribution is given by

$$M_0(t) = \frac{\lambda}{\lambda - t} \quad \text{for} \quad t < \lambda.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 158

Gamma distribution
(PDF, CDF and shape of the distribution)

Formal definition:

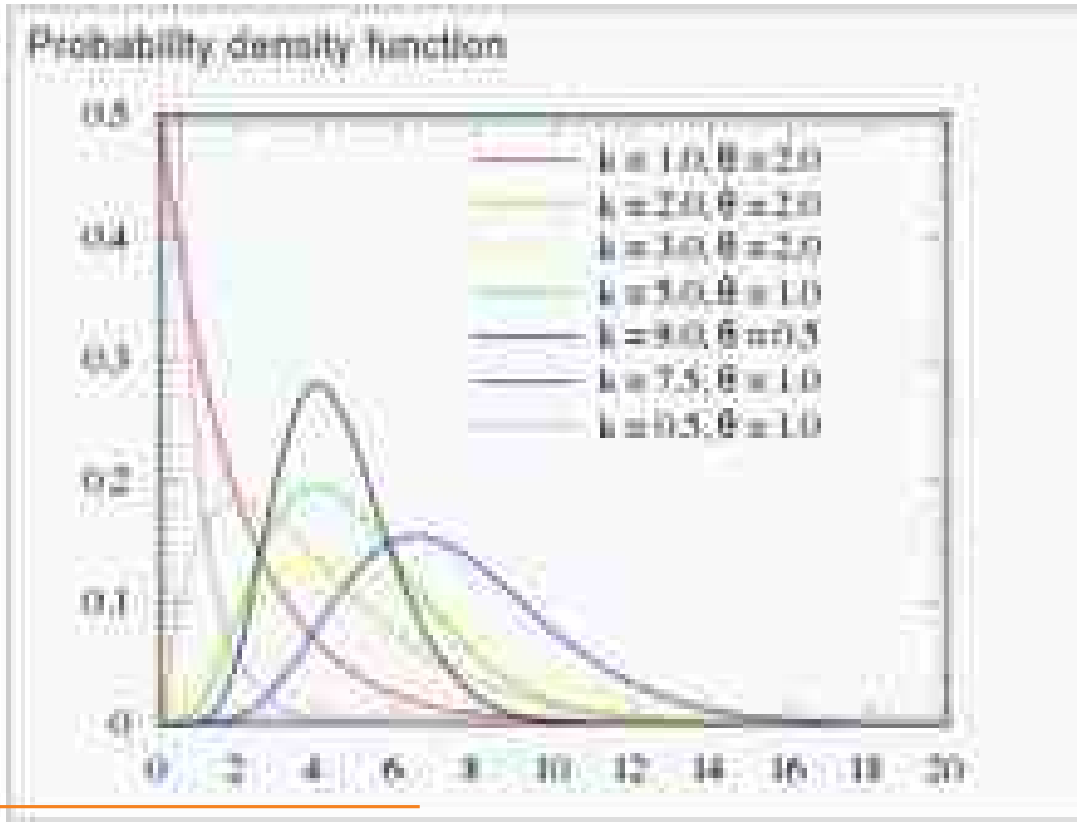
Gamma distribution

$$f(x)=\begin{cases}\dfrac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta} & 0<x<\infty,\ \alpha>0,\ \beta>0\\ 0 & \text{elsewhere}\end{cases}$$

is said to be the pdf of a gamma distribution with shape parameter α and scale parameter β and we write this as X has a $\Gamma(\alpha, \beta)$ distribution.

Shape of the distribution

- From the graph it is obvious that, for any fixed value of the scale parameter β, **as α increases**, the shape of the distribution tends to normality.



First we introduce the **gamma (Γ) function**:

In calculus, the integral

$$\int_0^\infty y^{\alpha-1}e^{-y}dy$$

exists for $\alpha > 0$ and that the value of the integral is a positive number.

The integral is called the **gamma function**, and we write

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Properties of the Gamma Function

- If α =1, clearly (easy to verify through integration)

$$\Gamma(1) = \int_0^\infty e^{-y} dy = 1$$

- If $\alpha > 1$, an integration by parts show that

$$\Gamma(\alpha) = (\alpha - 1) \int_0^\infty y^{\alpha-2} e^{-y} dy = (\alpha - 1)\Gamma(\alpha - 1).$$

- Accordingly, if α is a positive **integer** greater than 1,

$$\Gamma(\alpha) = (\alpha - 1)(\alpha - 2)\cdots(3)(2)(1)\Gamma(1) = (\alpha - 1)!.$$

- Since $\Gamma(1) = 1$, this suggests we take **0! =1**.

- It is important to note that **the value of a gamma function is always positive**.

In the integral that defines $\Gamma(\alpha)$, let us introduce a new variable by writing $y = \frac{x}{\beta}$, where $\beta > 0$

Now

$$y = \frac{x}{\beta} \Rightarrow x = y\beta \Rightarrow \frac{dx}{dy} = \beta \Rightarrow \frac{dx}{\beta} = \text{dy}$$

Also, limits remain unchanged.

Hence, we have:

$$\Gamma(\alpha) = \int_0^{\infty} \left(\frac{x}{\beta}\right)^{\alpha-1} e^{-x/\beta} \left(\frac{1}{\beta}\right) dx,$$

Hence

$$1 = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} \frac{1}{\beta^{\alpha-1+1}} x^{\alpha-1} e^{-x/\beta} dx = \frac{1}{\Gamma(\alpha)} \int_0^{\infty} \frac{1}{\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx$$

$$= \int_0^{\infty} \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx$$

Now

$$\int_0^{\infty} \frac{1}{\Gamma(\alpha)} \cdot \frac{1}{\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx = 1$$

implies that

$$\frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty$$

is **eligible** to be called a **probability density function.**

CDF of the Gamma Distribution

The CDF of the distribution is given by

$$F(x)=\begin{cases}\dfrac{1}{\Gamma(\alpha)}\gamma\left(\alpha,\dfrac{x}{\beta}\right) & \alpha>0,\beta>0\\ 0 & \text{elsewhere}\end{cases}$$

where this $\gamma\left(\alpha,\dfrac{x}{\beta}\right)$ is the incomplete gamma function.

Incomplete gamma function

In mathematics, the **upper** and **lower incomplete gamma functions** are types of special functions which arise as solutions to various mathematical problems such as certain integrals.

The upper incomplete gamma function is defined as:

$$\Gamma(\alpha, y) = \int_{y}^{\infty} t^{\alpha-1} e^{-t} dt,$$

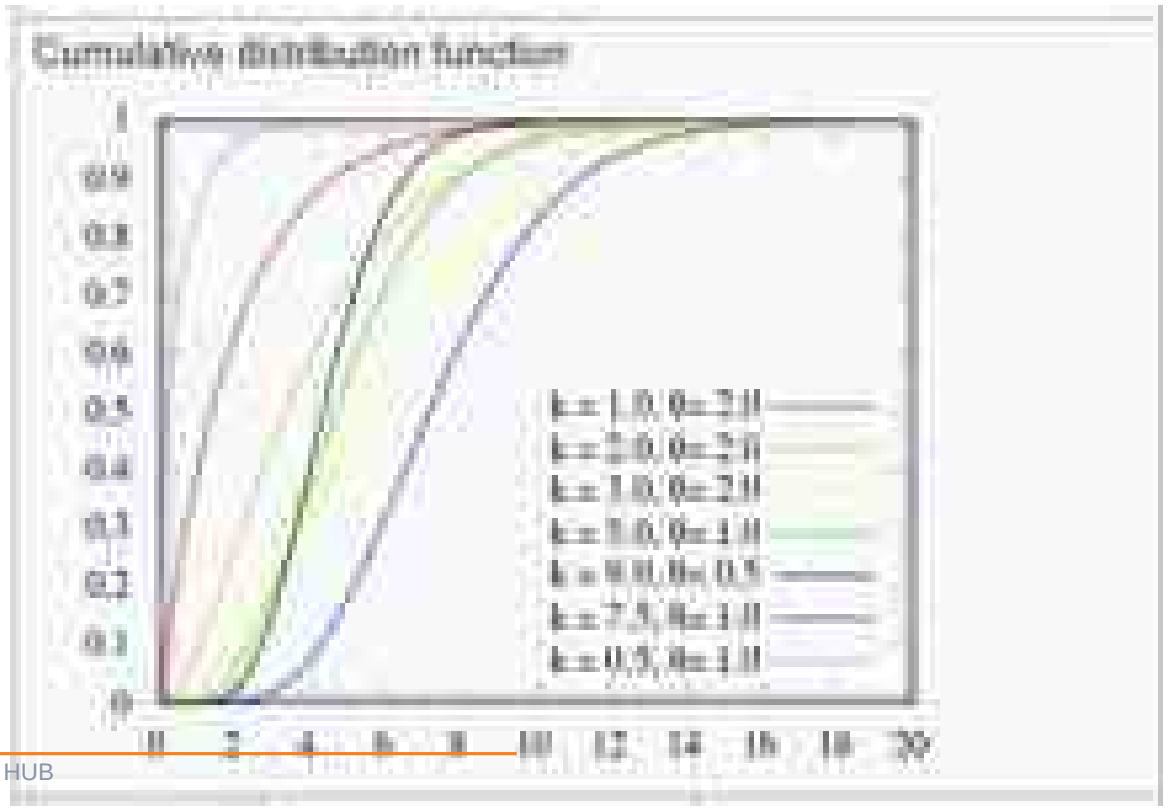
whereas the lower incomplete gamma function is defined as:

$$\gamma(\alpha, y) = \int_{0}^{y} t^{\alpha-1} e^{-t} dt.$$

It is interesting to note that $\gamma(\alpha, y) + \Gamma(\alpha, y) = \Gamma(\alpha).$

Shape of the CDF:

We can see that for smaller values of the shape parameter α, i.e. α =0.5 and 1.0, the shape of the CDF appears to be parabolic whereas, for larger values of α i.e. α =2.0, 3.0,5.0 etc , the CDF is S-shaped.



Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 159

Derivation of Mean and Variance of the Gamma distribution

Mean:

$$\mu_X = E\left(X\right) = \alpha\beta$$

Variance:

$$\sigma_X^2 = \alpha\beta^2$$

Derivation of the Mean

- If X has a gamma distribution with parameters α and β, then the mean of X is given by

$$\mu_X = E(X) = \alpha\beta$$

Proof:

$$\int_0^\infty x.\left(\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\right)dx=\int_0^\infty\left(\frac{x^{\alpha}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\right)dx=\frac{\alpha}{\alpha}\int_0^\infty\left(\frac{x^{\alpha}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\right)dx$$

$$=\alpha\int_0^\infty\left(\frac{x^{\alpha}e^{-x/\beta}}{\beta^\alpha\alpha\Gamma(\alpha)}\right)dx=\alpha\int_0^\infty\left(\frac{x^{\alpha}e^{-x/\beta}}{\beta^\alpha\,\Gamma(\alpha+1)}\right)dx$$

$\because$ we know that, $\Gamma(\alpha+1)\ =\alpha\Gamma(\alpha).$

Hence,

$$E(X) = \alpha \int_0^{\infty} \left(\frac{x^{\alpha} e^{-x/\beta}}{\beta^{\alpha} \Gamma(\alpha+1)} \right) dx = \alpha\beta \int_0^{\infty} \left(\frac{x^{\alpha} e^{-x/\beta}}{\beta^{\alpha+1} \Gamma(\alpha+1)} \right) dx$$

$$= \alpha\beta \int_0^{\infty} f(x; \alpha+1, \beta) dx = \alpha\beta(1) = \alpha\beta$$

$\therefore$ The total area under the curve of a probability density function is always equal to 1.

Derivation of the Variance

If X has a gamma distribution with parameters α and β,

The variance of X is

$$\sigma_X^2 = E\left(X^2\right) - \left[E\left(X\right)\right]^2$$

$$E\left(X^2\right)=\int_0^{\infty} x^2.\left(\frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}\right)dx=\alpha(\alpha+1)\beta^2$$

The variance of X is

$$\sigma_X^2 = E\left(X^2\right) - \left[E\left(X\right)\right]^2 = \alpha\left(\alpha+1\right)\beta^2 - \left(\alpha\beta\right)^2$$

$$= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 160

Example of computation of probabilities for a gamma-distributed random variable

Example:

Let X have a gamma distribution with pdf

$$f(x) = \frac{1}{\beta^2} x e^{-x/\beta}, \quad 0 < x < \infty,$$

zero elsewhere. If $x = 2$ is the unique mode of the distribution, find the parameter β and $P(X < 9.46)$.

Solution:

The pdf of the gamma distribution is given by

$$f(x)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-1}e^{-x/\beta},\quad 0<x<\infty,\alpha>0,\beta>0$$

Letting $\alpha = 2$, we have

$$f(x) = \frac{1}{\Gamma(2)\beta^2} x^{2-1} e^{-x/\beta} = \frac{1}{1!\beta^2} x e^{-x/\beta}$$

$$= \frac{1}{\beta^2} x e^{-x/\beta}, \qquad 0 < x < \infty$$

Hence we can see that the given density function is the pdf of the gamma distribution having α =2.

vulearninghub.xyz | by VU Learning HUB

- In general, it is easy to prove that, for $\alpha \geq 1$, the mode of the Gamma distribution is given by $(\alpha - 1)\beta$. .
- Here we have been given the information that the mode=2. Therefore,

$$Mode = (2-1)\beta = 2 \Rightarrow \beta = \frac{2}{2-1} = 2.$$

Hence,

$$f(x) = \frac{1}{\beta^2} x e^{-x/\beta} = \frac{1}{2^2} x e^{-x/2} = \frac{1}{4} x e^{-x/2}, \qquad 0 < x < \infty$$

Now to find $P(X < 9.46)$.

$$P(X < 9.46) = \int_0^{9.46} \frac{1}{4} x e^{-x/2} dx = \frac{1}{4} \int_0^{9.46} x e^{-x/2} dx$$

$$P(X < 9.46) = \frac{1}{4}\left(-2e^{-x/2}x - 4e^{-x/2}\right)_0^{9.46}$$

$$= \frac{1}{4}\left[\left(-2e^{-9.46/2}(9.46) - 4e^{-9.46/2}\right) - \left(-2e^{-0/2}.0 - 4e^{-0/2}\right)\right]$$

$$= \frac{1}{4}\left[\left(0.16700 - 0.003531\right) - \left(-0 - 4e^{0}\right)\right]$$

$$= \frac{1}{4}\left[\left(-0.20230\right) - \left(-4\right)\right] = \frac{1}{4}\left[\left(-0.20230\right) + \left(4\right)\right] = \frac{3.79770}{4} = 0.9494.$$

In other words, the probability that $X < 9.46$ is 94.94% almost 95 %.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 161

MGF of Gamma Distribution

The moment-generating function of the gamma distribution is

$$M(t) = (1 - \beta t)^{-\alpha} \quad for\ t < \frac{1}{\beta}$$

In order to derive the mgf, we begin as follows

$$M(t) = E\left(e^{tX}\right) = \int_0^{\infty} e^{tx} \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx$$

PROOF:

- First let's combine the two exponential terms and move the gamma fraction out of the integral:

$$M(t) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{tx - x/\beta} dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha-1} e^{-x(1-\beta t)/\beta} dx$$

vulearninghub.xyz | by VU Learning HUB

Now we're ready to do a substitution in the integral.

$$\text{Let } u = \frac{x(1-\beta t)}{\beta} \Rightarrow x = \frac{\beta u}{1-\beta t}$$

$$\text{That means we have, } \quad du = \frac{1-\beta t}{\beta}\,dx \Rightarrow dx = \frac{\beta}{1-\beta t}\,du$$

Also, the limits remain unchanged.

vulearninghub.xyz | by VU Learning HUB

Now substitute those in the integral :

$$M(t)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\int_{0}^{\infty}x^{\alpha-1}e^{-x(1-\beta t)/\beta}dx$$

$$M(t)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\int_{0}^{\infty}\left(\frac{\beta u}{1-\beta t}\right)^{\alpha-1}\left(e^{-u}\right)\frac{\beta}{1-\beta t}du$$

$$M(t)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\int_{0}^{\infty}(1-\beta t)^{-\alpha+1}(1-\beta t)^{-1}(\beta u)^{\alpha-1}\beta e^{-u}du$$

vulearninghub.xyz | by VU Learning HUB

$$M(t)=\frac{1}{\Gamma(\alpha)\beta^{\alpha}}\int_{0}^{\infty}(1-\beta t)^{-\alpha+1}(1-\beta t)^{-1}(\beta u)^{\alpha-1}\beta e^{-u}du$$

can be re-written as

$$M(t)=\frac{(1-\beta t)^{-\alpha}}{\Gamma(\alpha)\beta^{\alpha}}\int_{0}^{\infty}(\beta u)^{\alpha-1}\beta e^{-u}du$$

$$M(t) = \frac{(1-\beta t)^{-\alpha}}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} (\beta u)^{\alpha-1} \beta e^{-u} du$$

can be re-written as

$$M(t) = \frac{(1-\beta t)^{-\alpha}}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} u^{\alpha-1} \beta^{\alpha} e^{-u} du = \frac{(1-\beta t)^{-\alpha}}{\Gamma(\alpha)} \int_0^{\infty} u^{\alpha-1} e^{-u} du$$

The integral is now the gamma function: $\Gamma(\alpha) = \int_0^\infty u^{\alpha-1} e^{-u} du.$

Making that substitution:

$$M(t) = \frac{(1-\beta t)^{-\alpha}}{\Gamma(\alpha)} \Gamma(\alpha)$$

Now, cancelling out the terms and we have the required **moment-generating function**:

$$M(t) = (1 - \beta t)^{-\alpha} \quad for\ t < \frac{1}{\beta}$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 162

Beta distribution
(PDF, CDF and shape of the distribution)

The **probability density function** of the β distribution is

$$f\left(x;\alpha,\beta\right)=\frac{1}{B\left(\alpha,\beta\right)}x^{\alpha-1}\left(1-x\right)^{\beta-1},\qquad 0\le x\le 1;\alpha,\beta>0$$

where,

$$B\left(\alpha,\beta\right)=\frac{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}{\Gamma\left(\alpha+\beta\right)}$$

and α and β are the **shape parameters** respectively.

vulearninghub.xyz - by VuLearningHUB

About the Beta function B

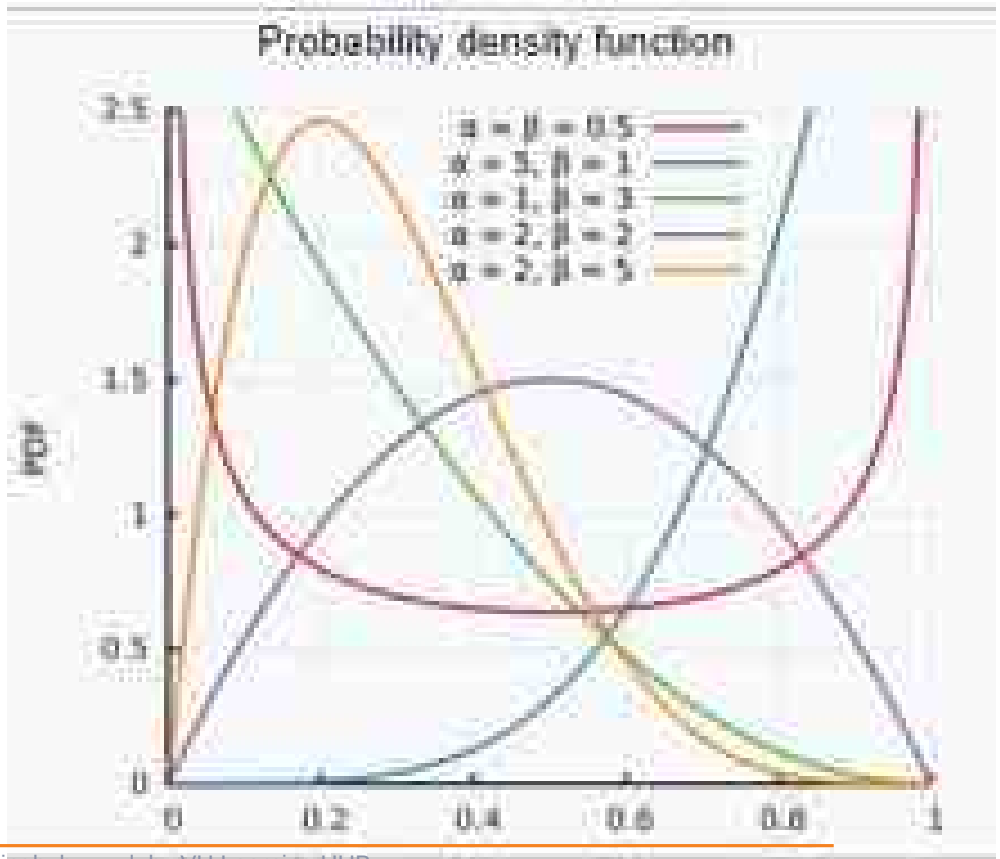
- The Beta function B in the denominator plays the role of a "**normalizing constant**" which assures that the total area under the density curve equals 1.

- The Beta function is equal to a **ratio** of Gamma functions:

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- Keeping in mind that for integers, $\Gamma(k) = (k-1)!$, one can do some checking and get an idea of what the shape might be.

Shape of the distribution:



Cumulative distribution Function

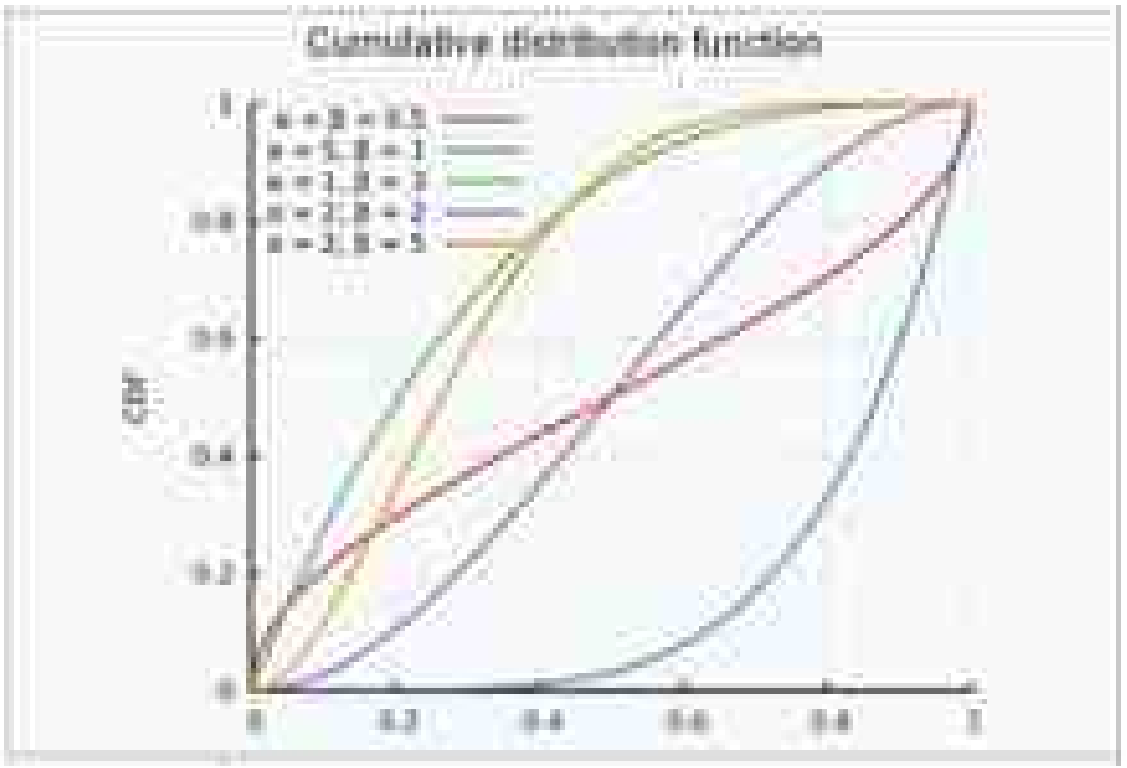
- The formula for the cumulative distribution function of the beta distribution is also called the **incomplete β function ratio** (commonly denoted by I_x) and is defined as

$$F(x) = I_x(\alpha, \beta) = \frac{\int_0^x t^{\alpha-1}(1-t)^{\beta-1}}{B(\alpha, \beta)}, \qquad 0 \le x \le 1; \alpha, \beta > 0$$

where B is the beta function defined above.

vulearninghub.xyz | by VU Learning HUB

Shape of the CDF:



Some Properties:

- **Mean** $=\mu$

$$= E(x) = \alpha/\alpha+\beta$$

- **Variance(*x*)**

$$= \alpha\beta / (\alpha+\beta)^2 (\alpha+\beta+1)$$

Mode:

- If $\alpha > 1$and $\beta > 1$, the peak of the density is in the interior of [0,1] and mode of the Beta distribution is

$$\textbf{mode} = \alpha - 1/\alpha + \beta - 2$$

If α or $\beta < 1$, the mode may be at an edge.

Relation to Uniform Distribution

Proposition 1:

A Beta distribution with parameters α and β is a uniform distribution on the interval .

Proof: When α=1 and β=1 , we have,

$$\frac{1}{B(\alpha,\beta)}x^{\alpha-1}(1-x)^{\beta-1}=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}$$

$$=\frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)}x^{1-1}(1-x)^{1-1}=\frac{\Gamma(2)}{\Gamma(1)\Gamma(1)}x^{0}(1-x)^{0}$$

$$=\frac{(2-1)!}{0!0!}1.(1)=1. \qquad \because \Gamma(n)=(n-1)!$$

- Therefore, the probability density function of a Beta distribution with parameters α and β can be written as

$$f_X\left(x;\alpha=1,\beta=1\right)=1, \qquad if, x\in[0,1]$$
$$=0, \qquad f, x\notin[0,1]$$

which is the probability density function of a **uniform distribution** of X on the **interval** [0,1].

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 163

Derivation of the Mean and Variance of the Beta distribution

- As we know that the mean of a distribution of a continuous random variable is given by

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

$$\mu = E(X) = \int_0^1 x.\frac{1}{B(\alpha,\beta)}.x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$\mu = E(X) = \int_0^1 x.\frac{1}{B(\alpha,\beta)}.x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \int_0^1 \frac{1}{B(\alpha,\beta)}.x^{\alpha-1+1}(1-x)^{\beta-1}\,dx = \int_0^1 \frac{1}{B(\alpha,\beta)}.x^{\alpha}(1-x)^{\beta-1}\,dx$$

$$= \frac{1}{B(\alpha,\beta)}\int_0^1 x^{\alpha}(1-x)^{\beta-1}\,dx$$

Now we know that the function is given by

$$B(\alpha,\beta)=\int_0^1 t^{\alpha-1}(1-t)^{\beta-1}\,dt$$

Hence,

$$\mu=\frac{1}{B(\alpha,\beta)}\int_0^1 x^{(\alpha+1)-1}(1-x)^{\beta-1}\,dx=\frac{1}{B(\alpha,\beta)}.B(\alpha+1,\beta)$$

vulearninghub.xyz | by VU Learning HUB

But we know that,

$$B(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

Hence

$$\mu=\frac{B(\alpha+1,\beta)}{B(\alpha,\beta)}=\frac{\dfrac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+1+\beta)}}{\left(\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}\right)}=\frac{\alpha\Gamma(\alpha)\Gamma(\beta)}{(\alpha+\beta)\Gamma(\alpha+\beta)}.\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$$

$$=\frac{\alpha}{(\alpha+\beta)\Gamma(\alpha+\beta)}.\Gamma(\alpha+\beta)=\frac{\alpha}{(\alpha+\beta)}.$$

- The variance of this distribution is $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ respectively.

They are computed as below:

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

$$E(X^2) = \int_0^1 x^2 f(x)\,dx$$

$$= \int_0^1 x^2 \frac{1}{B(\alpha,\beta)}.x^{\alpha-1}(1-x)^{\beta-1}\,dx$$

$$= \int_0^1 \frac{1}{B(\alpha,\beta)}.x^{(\alpha+2)-1}(1-x)^{\beta-1}\,dx = \frac{1}{B(\alpha,\beta)}\int_0^1 x^{(\alpha+2)-1}(1-x)^{\beta-1}\,dx$$

$$= \frac{B(\alpha+2,n)}{B(\alpha,\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

vulearninghub.xyz | by VU Learning HUB

Therefore,

$$\sigma^2 = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \left(\frac{\alpha}{\alpha+\beta}\right)^2 = \frac{\alpha+\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 164

Example of Application of the Beta Distribution

Important Note

It is important to note that actually there are two types of Beta distribution:

- Beta distribution of the first kind
- Beta distribution of the first kind is defined on the interval from 0 to 1

Important Note

It is important to note that actually there are two types of Beta distribution:

- Beta distribution of the second kind.
- Beta distribution of the second kind is defined on the interval from 0 to infinity.

I will explain this to you with the help of an example for the beta distribution of the first kind.

It is both interesting and important to note that the Beta distribution of the first kind (also called “Beta distribution”) is applicable in those situations where the **random variable** X goes from 0 to 1.

This may happen in a number of situations but the one that might come to mind first of all is the case when X represents the **proportions** of successes in a large sample.

Question:

Suppose that the proportion of defective DVDs in a large shipment follow a Beta distribution with $\alpha = 2$ and $\beta = 5$.

What is the probability that the shipment has 20% to 30% defective DVDs ?

The **largeness** of the sample will facilitate the assumption that our random variable X is **continuous**.

Question:

Suppose that the proportion of defective DVDs in a large shipment follow a Beta distribution with $\alpha = 2$ and $\beta = 5$.

What is the probability that the shipment has 20% to 30% defective DVDs ?

Solution:

Let X represent the proportion of defective DVDs in this large shipment.

(Here **defectiveness** denotes **‘Success’.)**

Then, obviously, X will go from **0 to 1.**

The probability density function of the Beta distribution is given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, 0 \leq x \leq 1, \alpha > 0, \beta > 0$$

In this example, $\alpha = 2$ and $\beta = 5$.

Hence, our PDF is

$$f(x) = \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)}, 0 \leq x \leq 1$$

We need to compute

$$P(20\% \leq x \leq 30\%) = P(0.20 \leq x \leq 0.30)$$

which is given by

$$P(0.2 \leq x \leq 0.3) = \int_{0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)} dx$$

Solving the beta function

$$B(a,b)=\frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

So,

$$B(2,5)=\frac{\Gamma(2)\Gamma(5)}{\Gamma(2+5)}=\frac{\Gamma(2)\Gamma(5)}{\Gamma(7)}=\frac{(2-1)!(5-1)!}{(7-1)!}$$

$$=\frac{(1)!(4)!}{(6)!}=0.03333$$

We need to compute

$$P(20\% \leq x \leq 30\%) = P(0.20 \leq x \leq 0.30)$$

which is given by

$$P(0.2 \leq x \leq 0.3) = \int_{0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)} dx = \int_{0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{0.033333} dx = 0.2352$$

Interpretation:

- Hence we can say that the probability that the shipment has 20% to 30% defective DVDs is 23.52% i.e. a little less than 25%.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 165

Normal Distribution
(PDF, CDF and Shape of the distribution)

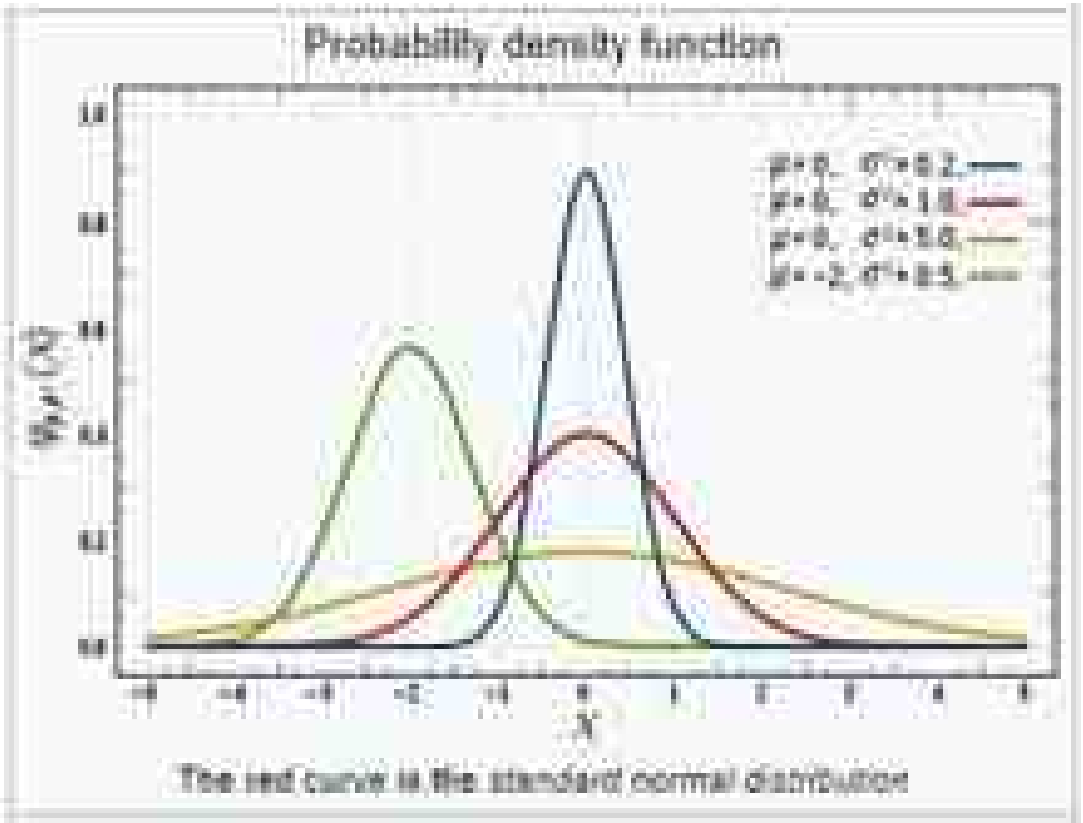
Definition:

A random variable X is said to have a normal distribution if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for} \quad -\infty < x < \infty.$$

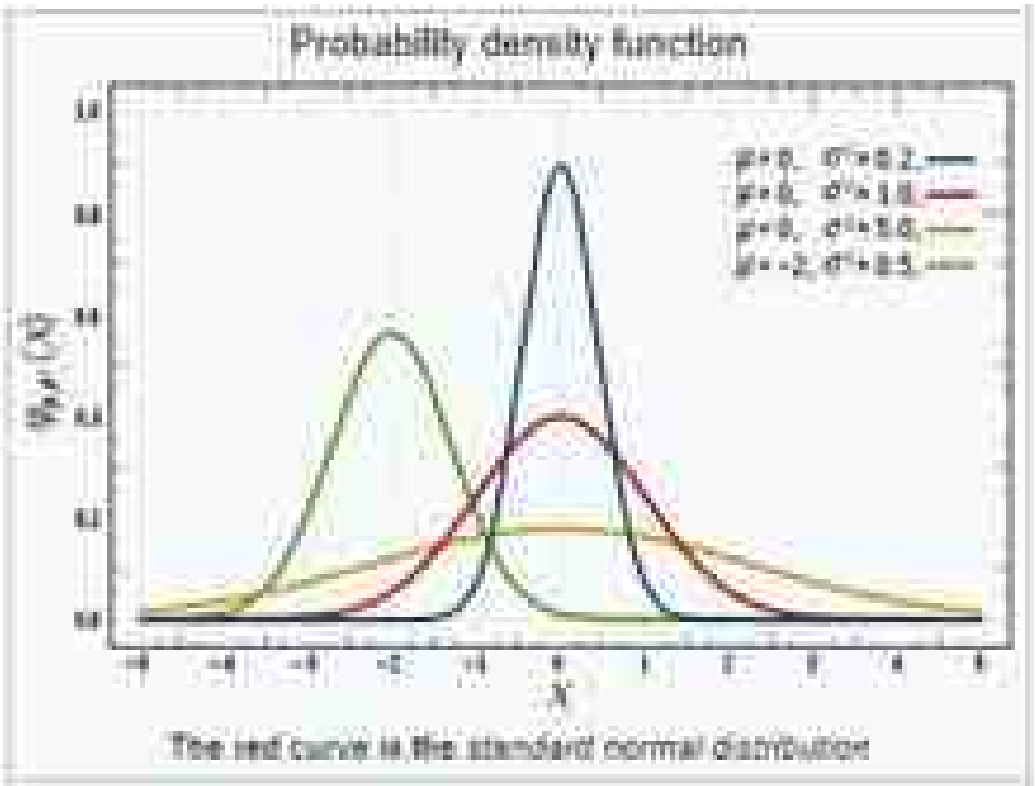
- μ is the mean or expectation of the distribution (and also its median and mode),
- $\sigma > 0$ is the standard deviation, and
- σ^2 is the variance.

Shape of the distribution:



The graph points to the following **two basic properties** of the normal distribution:

1. The normal curve is **asymptotic** to the x-axis,



2. The maximum ordinate

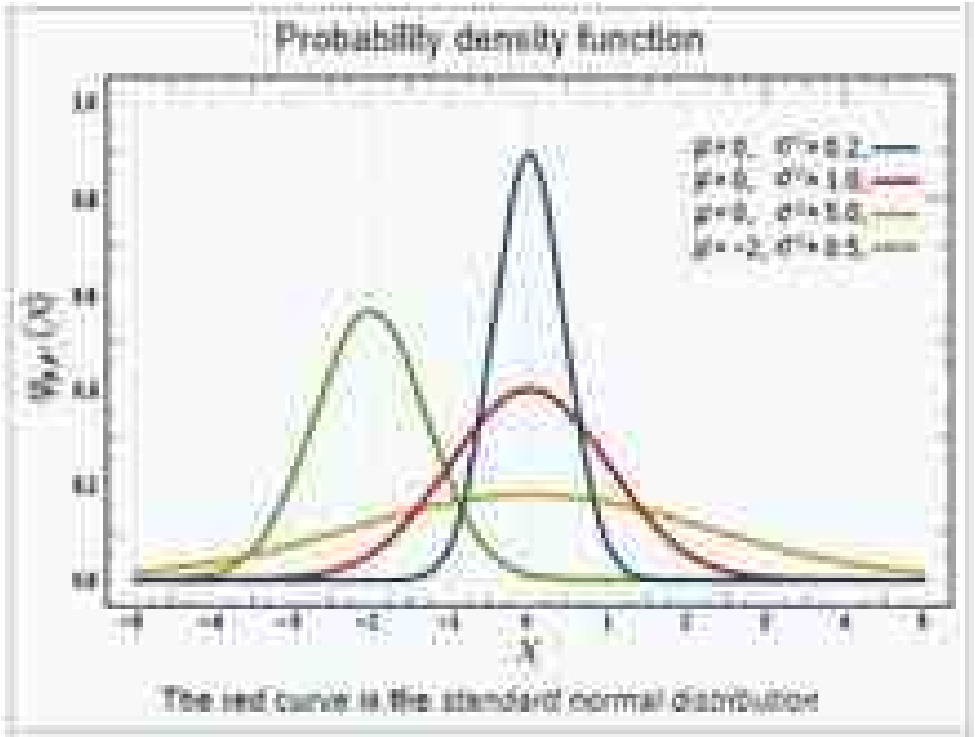
(i.e. the **modal ordinate**)

of the normal distribution

at $x=\mu$ is equal to

$1/(\sigma\sqrt{2\pi})$

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for} \quad -\infty < x < \infty.$$



Cumulative Distribution Function

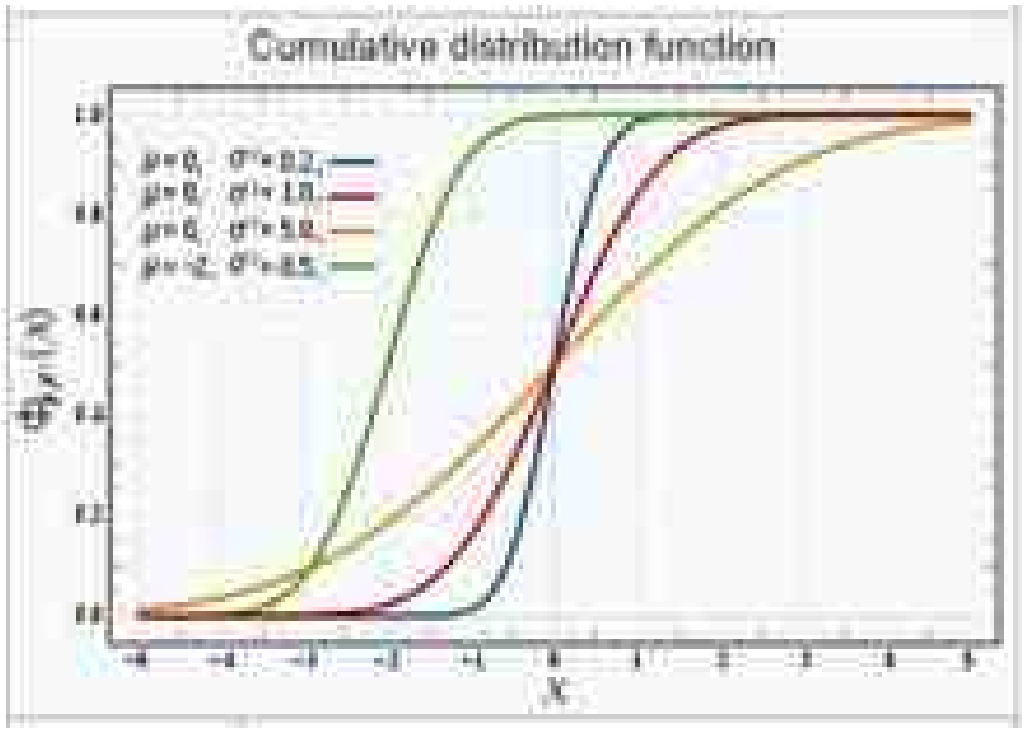
The formula for the **cumulative distribution function** of the normal distribution is

where $$F(x) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{x-\mu}{\sigma\sqrt{2}}\right)\right] \quad \text{for } -\infty < x < \infty.$$

erf(z) is the “error function”

defined by $$\operatorname{erf}(z) \equiv \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2}\, dt.$$

The shape of the **cumulative distribution function** of the normal distribution is **S-shaped**:



Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 166

Standard Normal Distribution

- The standard normal distribution is a **special** case of the normal distribution .

- It is the distribution that occurs when a normal random variable has a **mean** of zero and a **standard deviation** of one.

In other words, if the random variable Z follows a normal distribution with parameters (0, 1), i.e. $Z \sim N(0, 1)$ then we say Z is a standard normal random variable.

The density of standard normal distribution is therefore,

$$f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{z^2}{2}\right)}, \quad -\infty < z < \infty$$

Derivation of the Mean of the Standard Normal Distribution

The mean of the standard normal random variable Z is given by

$$E(Z) = \int_{-\infty}^{\infty} z f_Z(z)\, dz$$

We know that, for any **odd function** g i.e. a function for which

$$g(-x) = -g(x) \ \forall x \in \mathbb{R}$$

the integral $\int_{-\infty}^{\infty} g(x)\,dx$ is equal to zero.

Here, it is easy to see that the function

$$zf_Z(z) = z\frac{1}{\sqrt{2\pi}}e^{\left(-\frac{z^2}{2}\right)} = \frac{z}{\sqrt{2\pi}}e^{\left(-\frac{z^2}{2}\right)}$$

is an **odd function**.

Hence, the mean $E(Z) = \int_{-\infty}^{\infty} zf_Z(z)dz = 0.$

Next: **Derivation of the Variance of the Standard Normal Distribution**

Using the short-cut formula of the variance, we have

$$\text{Var}(Z) = E(Z^2) - \left[E(Z)\right]^2$$

Now we know that,

for the standard normal distribution, $E(Z) = 0.$

So

$$\mathrm{Var}(Z) = E\left(Z^2\right) - [0]^2 = E\left(Z^2\right) = \int_{-\infty}^{\infty} z^2 \frac{e^{\left(-\frac{z^2}{2}\right)}}{\sqrt{2\pi}} dz$$

Now, we know that, for any **even function** g i.e. a function for which $g(-x) = g(x)\ \forall x \in \mathbb{R}$

the integral $\int_{-\infty}^{\infty} g(x)dx$ is equal to $2\int_{0}^{\infty} g(x)dx$.

Now $z^2 \frac{e^{\left(-\frac{z^2}{2}\right)}}{\sqrt{2\pi}}$ is an even function of z.

Hence

$$\text{Var}(Z) = E\left(Z^2\right) = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{\left(-\frac{z^2}{2}\right)} dz.$$

Let $w = \frac{z^2}{2}$, then $2w = z^2 \Rightarrow \sqrt{2w} = z \Rightarrow (2w)^{1/2} = z$

$\Rightarrow (2)^{1/2}(w)^{1/2} = z \Rightarrow \sqrt{2}(w)^{1/2} = z.$

Therefore

$$\frac{dz}{dw} = \sqrt{2}\,\frac{1}{2}(w)^{(1/2)-1} \Rightarrow \frac{dz}{dw} = \frac{1}{\sqrt{2}}(w)^{-1/2}$$

$$\Rightarrow \frac{dz}{dw} = \frac{1}{\sqrt{2w}} \Rightarrow dz = \frac{1}{\sqrt{2w}}\,dw.$$

And

$$\text{Var}(Z) = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} z^2 e^{\left(-\frac{z^2}{2}\right)} dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (2w) e^{(-w)} \frac{1}{\sqrt{2w}} dw$$

$$= \frac{2(2)}{\sqrt{2}\sqrt{2\pi}} \int_0^{\infty} (w) e^{(-w)} \frac{1}{\sqrt{w}} dw = \frac{2(2)}{\sqrt{2}\sqrt{2}\sqrt{\pi}} \int_0^{\infty} (w) e^{(-w)} \frac{1}{\sqrt{w}} dw$$

$$= \frac{\sqrt{2}\sqrt{2}}{\sqrt{\pi}} \int_0^{\infty} \sqrt{w} e^{(-w)} dw = \frac{2}{\sqrt{\pi}} \int_0^{\infty} w^{1/2} e^{(-w)} dw = \frac{2}{\sqrt{\pi}} \int_0^{\infty} w^{(3/2)-1} e^{(-w)} dw$$

$$= \frac{2}{\sqrt{\pi}} \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} . \frac{1}{2} \Gamma\left(\frac{1}{2}\right) = \frac{2}{\sqrt{\pi}} . \frac{1}{2} . \sqrt{\pi} \quad \left[\sin ce \, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right]$$

$$= 1.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 167

The Process of Standardization,
Area Table of the Standard Normal Distribution
and
Computation of probabilities pertaining to a normal random variable
(explained through an example)

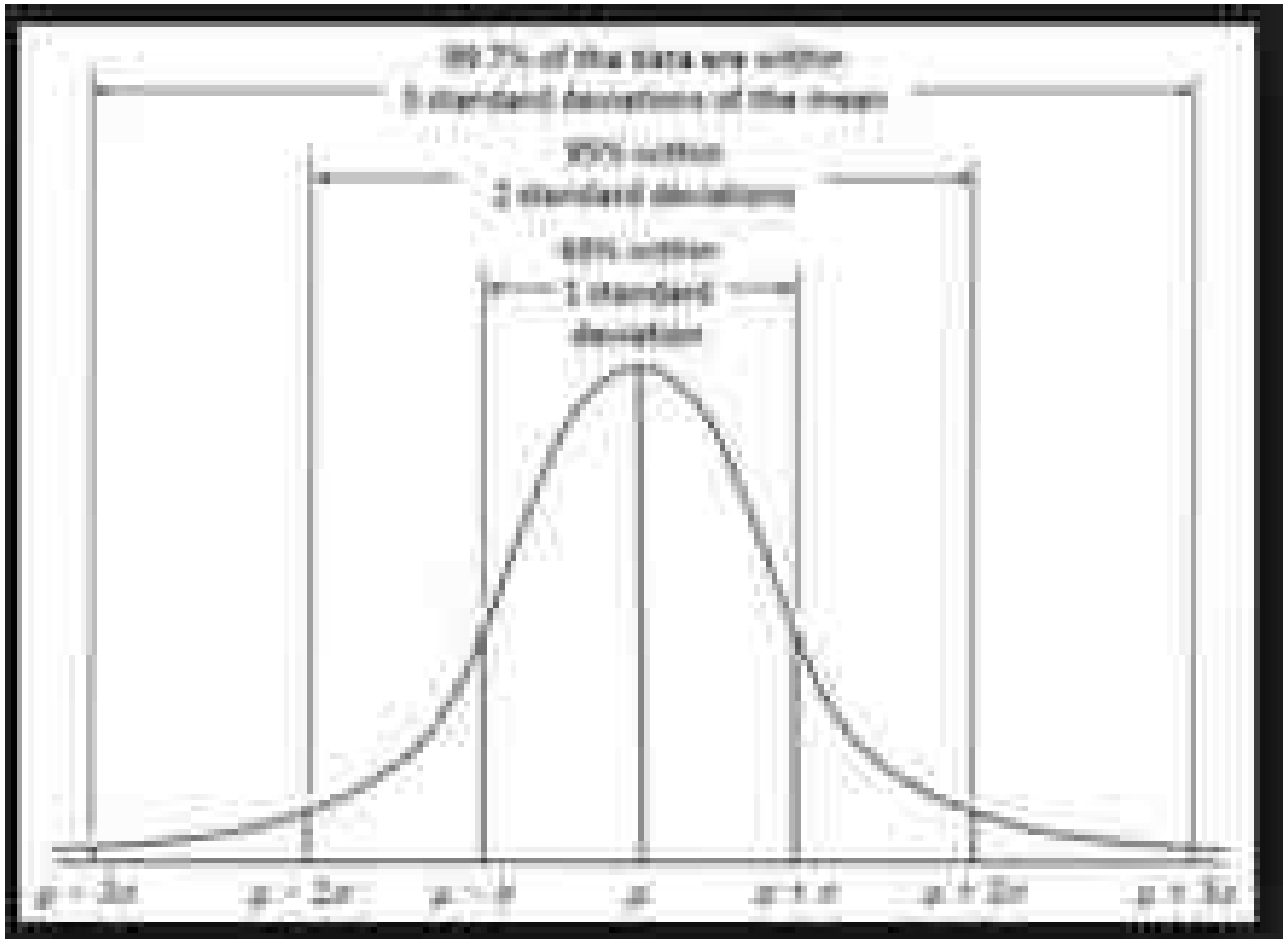
Normal Distribution

A random variable X is said to have a normal distribution if its pdf is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty$$

where μ is the mean of the distribution and $\sigma > 0$ is the standard deviation.

We say that X has the $N(\mu, \sigma^2)$ distribution.

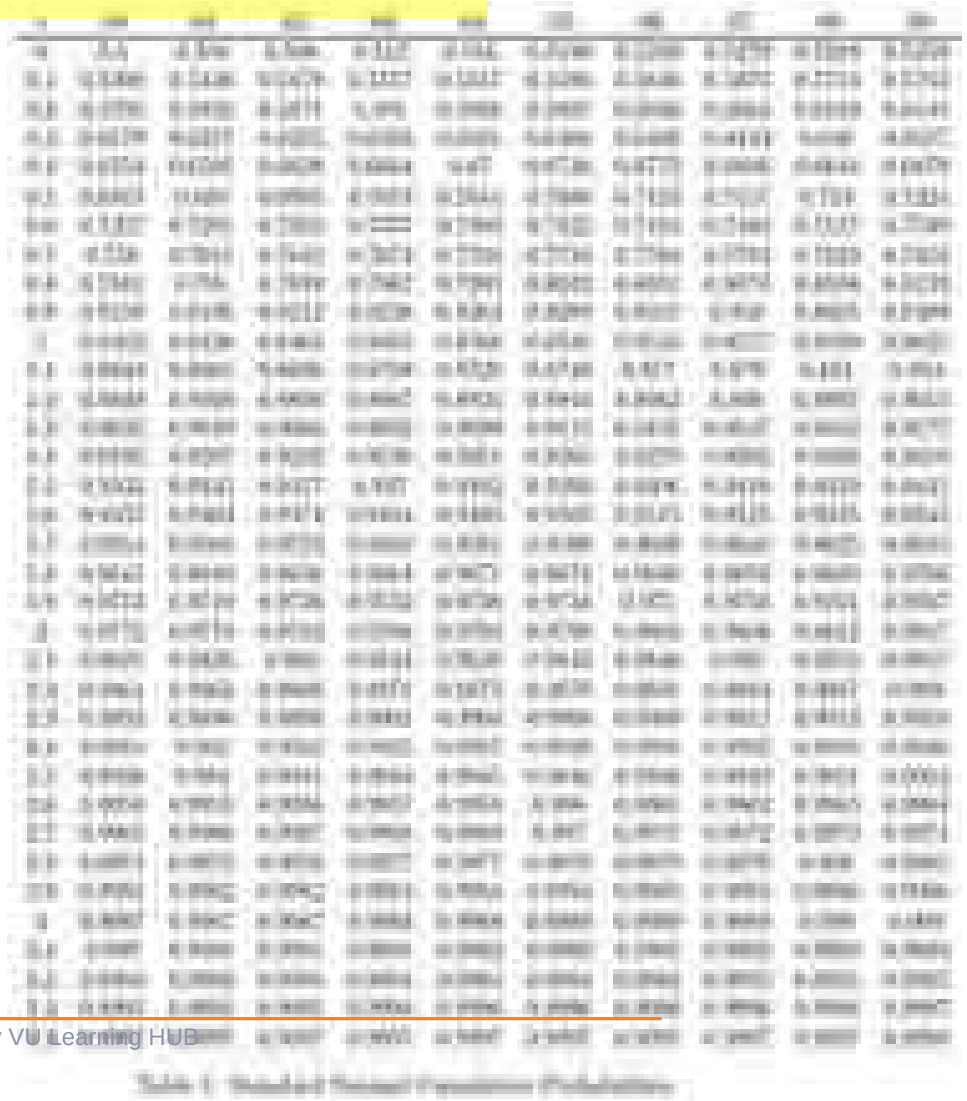


1. For this purpose, we let $Z = \frac{X - \mu}{\sigma}$.

Then, it is easy to prove that $Z \sim N(0,1)$.

2. Area Table of the Standard Normal Distribution

Table I



The Process of Standardization

Let X be $N(\mu,\sigma)$ and suppose that we are interested in computing the probability $P\left(a < X < b\right)$.

In order to be able to do so, we will need to convert our $N(\mu,\sigma^2)$ random variable X into the standard normal random variable Z~$N(0,1)$.

3. The Computation of Probabilities

Example:

Let X be $N(2, 25)$ and suppose that we are interested in computing the probability

$$P(0 < X < 10) = F(10) - F(0).$$

Applying the standardization formula $Z = \frac{X-\mu}{\sigma} = \frac{X-2}{5}$,

we obtain

$$P(0 < X < 10) = F(10) - F(0)$$

$$= \Phi\left(\frac{10-2}{5}\right) - \Phi\left(\frac{0-2}{5}\right)$$

$$= \Phi\left(\frac{8}{5}\right) - \Phi\left(\frac{-2}{5}\right) = \Phi(1.60) - \Phi(-0.40).$$

vulearninghub.xyz | by VU Learning HUB

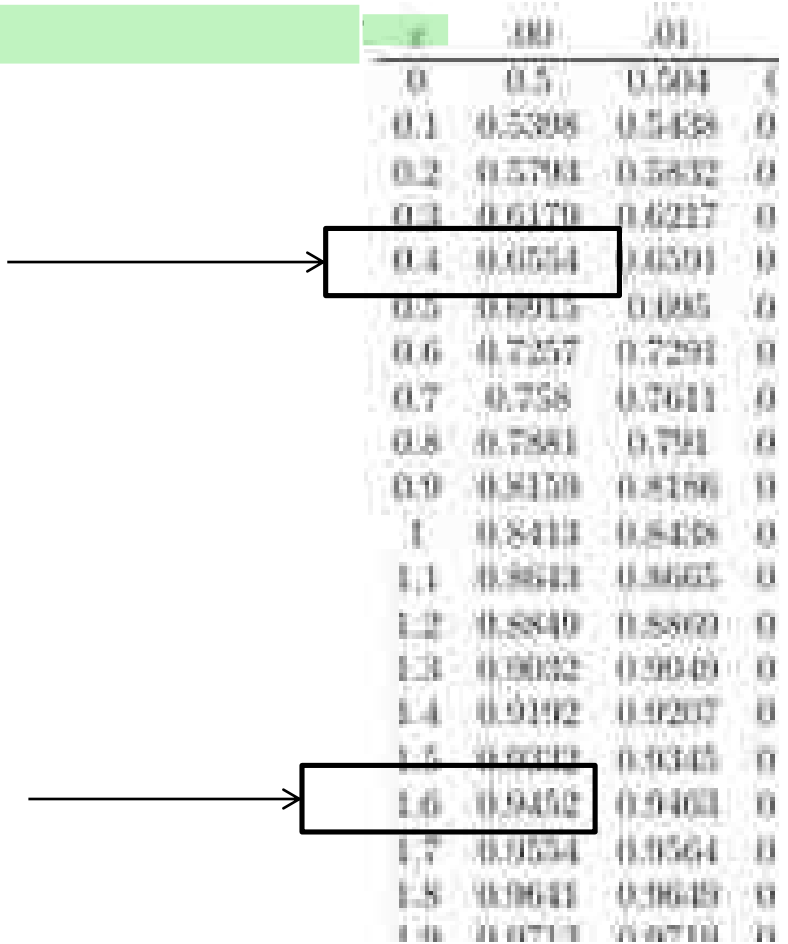
The **standard normal table**, gives us values of the distribution function of a standard normal variable,

i.e. it gives the values of $\Phi(z)$ for various values of z.

Area Table of the Standard Normal Distribution

Table I

Area Table of the Standard Normal Distribution



$$P(0 < X < 1.60) = \Phi(1.60) - \Phi(-0.00)$$

So

$$P(0 < X < 10) = \Phi\left(\frac{10-2}{5}\right) - \Phi\left(\frac{0-2}{5}\right)$$

$$= \Phi(1.60) - \Phi(-0.40) = 0.945 - (1 - 0.655) = 0.600 = 60\%.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 168

Computation of Some Specific Percentiles of a Normal Distribution (explained through an example)

Computing Percentiles

The **standard** normal distribution can also be used for computing **percentiles**.

For example, the median is the 50th percentile (P_{50}), the first quartile is the 25th percentile (P_{25}), and the third quartile is the 75th percentile (P_{75}).

- In some instances it may be of interest to compute other percentiles, for example the 5th or 95th.

Let us Recall the Process of Standardization

Let X be $N(\mu,\sigma)$ and suppose that we are interested in computing the probability $P\left(a < X < b\right)$.

In order to be able to do so, we will need to convert our $N(\mu,\sigma^2)$ random variable X into the standard normal random variable Z~$N(0,1)$.

For this purpose, we let $Z = \frac{X - \mu}{\sigma}$.

Then, it is easy to prove that $Z \sim N(0,1)$.

The **formula** below is used to compute percentiles of a normal distribution is obtained by “inverting” the standardization formula given above.

$$Z = \frac{X - \mu}{\sigma} \Rightarrow X - \mu = \sigma Z \Rightarrow X = \mu + \sigma Z$$

where, μ is the mean and σ is the standard deviation of the variable X,

and Z is the value from the standard normal distribution for the desired percentile.

Example:

The BMI for men aged 60 is normally distributed with mean equal to 29 with a standard deviation of 6 whereas the BMI for women aged 60 is normally distributed with mean 28 and standard deviation 7.

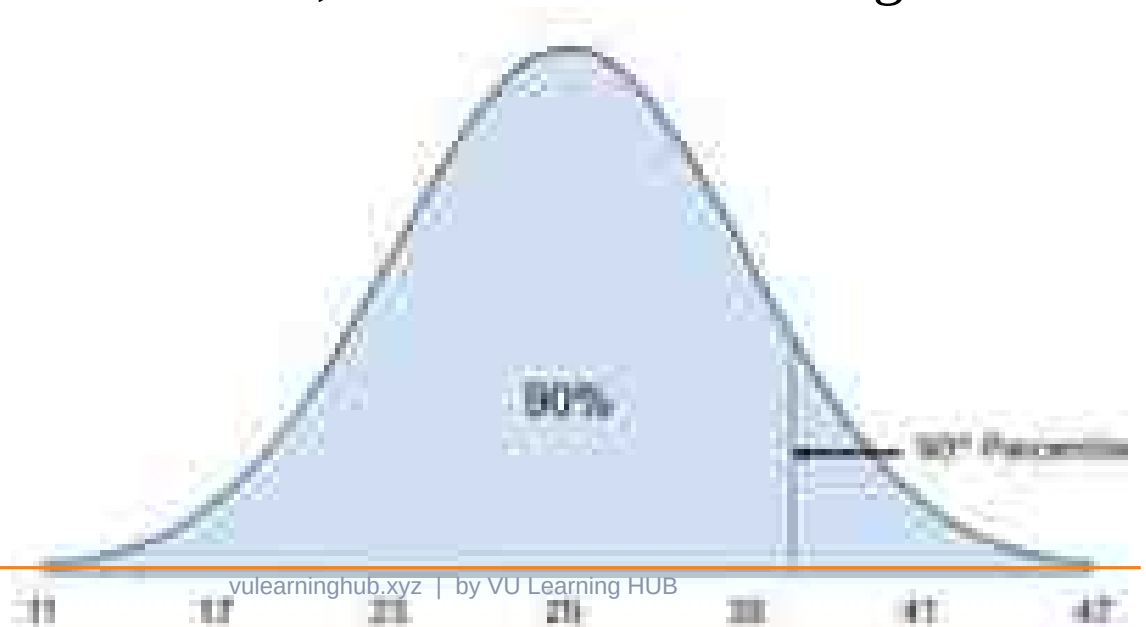
What is the 90th percentile of BMI for men and what does it mean ?

First of all, let us see what is meant by BMI:

- **Body Mass Index (BMI)** is a person's weight in kilograms divided by the square of height in meters.

- A high BMI can be an indicator of **high body fatness.**

- The 90th percentile is the BMI that holds 90% of the BMIs below it and 10% above it, as illustrated in the figure below.



- To compute the 90th percentile, we use the formula $X = \mu + \sigma Z$, and we will use the **standard normal distribution area table**, except that we will work in the **opposite** direction.

- Previously we started with a particular "*X*-value" and used the table to find the probability.

- However, in this case we want to start with a 90% probability and find the value of "*X*" that represents it.

- So we begin by going into the **interior** of the standard normal distribution area table to find the area under the curve closest to 0.90 i.e. 0.9000, and from this we can determine the corresponding *Z*-score.

- When we go to the table, we find that the value 0.9000 is not there exactly, however, the values 0.8997 and 0.9015 are there and correspond to Z values of 1.28 and 1.29 respectively.

- Since 0.8997 is much closer to 0.9000 than 0.9015, therefore we choose the z-score corresponding to 0.8997.

- The z-score is 1.28 (i.e., 89.97% of the area under the standard normal curve is below 1.28.

- Now that we have found $z=1.28$, we can use the equation

  $x = \mu + \sigma z = 29+6(1.28)$

  because we already know that the mean and standard deviation are 29 and 6, respectively.

So, using z =1.28 the 90th percentile of BMI for men comes out to be:

$$x = 29 + 6(1.28) = 29 + 7.68 = 36.68$$

Interpretation:

Ninety percent of the BMIs in men aged 60 are 36.68 or less. Ten percent of the BMIs in men aged 60 are 36.68 or higher.

REMARK:

The exact Z value holding 90% of the values below it is 1.282 which can be determined from a table of standard normal probabilities with **more** precision.

Using Z =1.282 the 90th percentile of BMI for men is:

$$X = 29 + 6(1.282) = 29 + 7.69 = 36.69$$

--- very close.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 169

The function *f*(x) defining the normal distribution is a proper PDF

The function *f(x)* defining the normal distribution is a proper pdf

i.e.

$$f(x) \geq 0$$

and

the total area under the normal curve is unity.

Recall:

A random variable X is said to have a normal distribution if its pdf is

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty < x < \infty.$$

where μ is the mean or expectation of the distribution and $\sigma > 0$ is the standard deviation.

(i) **Clearly** $f(x)$ is always non-negative.

VTLearningHub.xyz / by VTLearning HUB

(ii) The total probability (i.e. the total area under the curve) is

$$\text{Area} = \int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\,dx$$

Let, $z = \frac{x-\mu}{\sigma} => \sigma z = x - \mu => x = \mu + \sigma z$

so that $\frac{dx}{dz} = 0 + \sigma = \sigma => dx = \sigma dz.$

Limits remain unchanged.

Substituting these values, we get

$$\text{Area} = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \sigma dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz$$

$$= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{0} e^{-z^2/2} dz + \int_{0}^{\infty} e^{-z^2/2} dz \right]$$

The function $\int_{-\infty}^{0} e^{-z^2/2} dz$ being an even function of z,

can be by letting $w = -z$, written as $\int_{-\infty}^{0} e^{-z^2/2} dz = \int_{0}^{\infty} e^{-w^2/2} dw.$

Then,

$$\text{Area} = 2.\frac{1}{\sqrt{2\pi}} \int_{0}^{\infty} e^{-z^2/2} dz = \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} e^{-z^2/2} dz$$

Let $v = \frac{1}{2} z^2$ so that $\frac{dv}{dz} = \frac{1}{2} 2z = z \Rightarrow dv = zdz$

implying that

$$dz = \frac{dv}{z} = \frac{dv}{\sqrt{2v}}$$

Limits remain unchanged.

Then,

$$\text{Area}=\sqrt{\frac{2}{\pi}}\int_0^{\infty} e^{-v}\frac{dv}{\sqrt{2v}}=\frac{1}{\sqrt{\pi}}\int_0^{\infty} e^{-v}v^{-1/2}dv=\frac{1}{\sqrt{\pi}}\int_0^{\infty} e^{-v}v^{(1/2)-1}dv$$

$$=\frac{1}{\sqrt{\pi}}\Gamma\left(\frac{1}{2}\right)=\frac{1}{\sqrt{\pi}}.\sqrt{\pi}=1.$$

Thus the total area (probability) under the normal curve is **unity.**

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 170

Derivation of the mean and a variance of the Normal Distribution

We know that the pdf of the normal distribution with the parameters μ and σ respectively is written as

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} \qquad -\infty \leq x \leq \infty$$

The **mean** the normal distribution is μ

Proof:

By definition,

$$\mu = E(X) = \int_{-\infty}^{\infty} xf\left(x\right)dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x.e^{-[(x-\mu)^2/2\sigma^2]}dx$$

$$\text{Let,} \quad z = \frac{x-\mu}{\sigma}. \text{Then,} \; x = \mu + z\sigma \; \text{and} \, dx = \sigma dx.$$

$$\{\text{Limits: when } x = \infty, z = \infty; when \; x = -\infty, z = -\infty\}$$

vulearninghub.xyz | by VU Learning HUB

$$\text{Therefore, } E(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + z\sigma) e^{-z^2/2} dz$$

$$= \frac{\mu}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz + \frac{\sigma}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z e^{-z^2/2} dz$$

The first integral presents μ times the area under a normal curve with zero mean and unit variance and hence is equal to μ. The integral being an odd function, equals zero.

Thus, E(X)= μ i.e. μ is the mean of the normal distribution.

vulearninghub.xyz | by VU Learning Hub

And, now of find the variance:

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

$$= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 e^{-[\{x-\mu)^2/2\sigma^2} dx$$

on putting $z = \frac{x - \mu}{\sigma}$

$$Var(X) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \sigma^2 . e^{-z^2/2} dz = \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z.z\, e^{-z^2/2} dz,$$

To integrate by parts, we use the formula

$\int_{-\infty}^{\infty} u dv = uv - \int_{-\infty}^{\infty} v du$ and make the

substitution $dv = ze^{-z^2/2}dz$ and $u = z$

so that $v = -e^{-z^2/2}$ and $du = dz$.

Then,

$$Var(X) = \frac{\sigma^2}{\sqrt{2\pi}}\left[-z.e^{-z^2/2}\right]_{-\infty}^{\infty} + \frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-z^2/2}dz$$

in first part applying L'Hospital rule,

$$= 0 + \sigma^2$$

Hence, $E(X) = \mu$ and $Var(X) = \sigma^2$.

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

Topic No. 171

The Median and the Mode of the Normal distribution are each equal to μ, the mean of the distribution

(i) The **median** , *a*, is obtained as the solution of

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

m– the median of the distribution

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{M} e^{-[(x-\mu)^2/2\sigma^2]} dx = \frac{1}{2}$$

Putting $z = \frac{x-\mu}{\sigma} \Rightarrow x = \mu + \sigma z \Rightarrow \frac{dx}{dz} = 0 + \sigma = \sigma \Rightarrow dx = \sigma dz$

Limits

$x \to -\infty, \;\; z \to -\infty$

$x \to M, \;\; z \to \frac{M-\mu}{\sigma}$

Hence, the median is the solution of the equation

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{M-\mu}{\sigma}} e^{-[(x-\mu)^2/2\sigma^2]} dx = \frac{1}{2}$$

or

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\frac{M-\mu}{\sigma}} e^{-z^2/2} \sigma dz = \frac{1}{2} \Rightarrow \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{M-\mu}{\sigma}} e^{-z^2/2} dz$$

vulearninghub.xyz | by VU Learning HUB

But we know from the symmetry of the standard normal distribution that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0} e^{-z^2/2}dz = \frac{1}{\sqrt{2\pi}}\int_{0}^{\infty} e^{-z^2/2}dz = \frac{1}{2},$$

$$\therefore \frac{M-\mu}{\sigma} = 0 \; or \; M = \mu,$$

i.e. μ is the median of the distribution.

ii) By definition, for any pdf $f(x)$, the **mode**, if any, is that value of x for which

$$f'(x) = 0$$

and

$$f''(x) < 0.$$

Now,

$$f'(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-[(x-\mu)^2/2\sigma^2]} \left(\frac{x-\mu}{\sigma^2}\right)\left(-\frac{2}{2}\right)$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}} (x-\mu) e^{-[(x-\mu)^2/2\sigma^2]}$$

Equating $f'(x) = 0$, we see that $x = \mu$.

Now,

Again differentiating, we obtain

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-(x-\mu)^2/2\sigma^2} - e^{-(x-\mu)^2/2\sigma^2}\ \frac{(x-\mu)^2}{\sigma^2} \right]$$

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \left[1 - e^{-(x-\mu)^2/2\sigma^2} \frac{(x-\mu)^2}{\sigma^2} \right]$$

Substituting $x = \mu$ in $f''(x)$, we see that $f''(x) < 0$.

Thus,

$x = \mu$ is the mode of the normal distribution.

Hence the median and mode are both equal to μ.

Since,

mean = median = mode,

the normal distribution is symmetrical and unimodal.

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 172

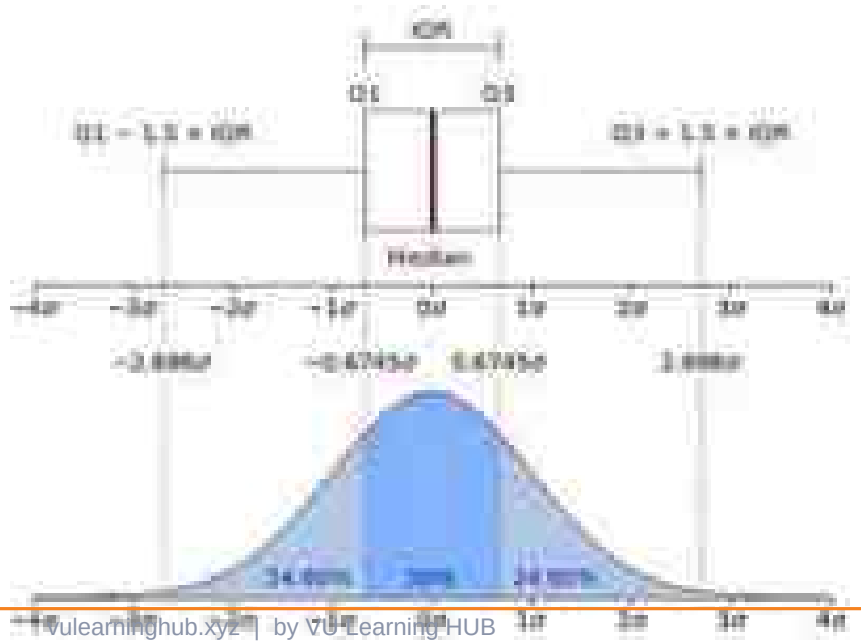
The Quartile Deviation of the Normal distribution is approximately (2/3)rd of its standard deviation

Explanation:

For any distribution, the **Semi-Inter-Quartile Range** or **Quartile Deviation**, Q, is given by

$$Q = \frac{Q_3 - Q_1}{2}$$

Inter-Quartile Range (IQR)



Now, for a **Symmetric** distribution,

$$Q_3 - Q_2 = Q_2 - Q_1 \Rightarrow Q_3 = 2Q_2 - Q_1$$

$$\Rightarrow Q = \frac{2Q_2 - Q_1 - Q_1}{2} = \frac{2Q_2 - 2Q_1}{2} = Q_2 - Q_1$$

and

$$Q = Q_3 - Q_2$$

But, for a **Symmetric** distribution,

$$Q_2 = \mu$$

implying that, for a symmetric distribution,

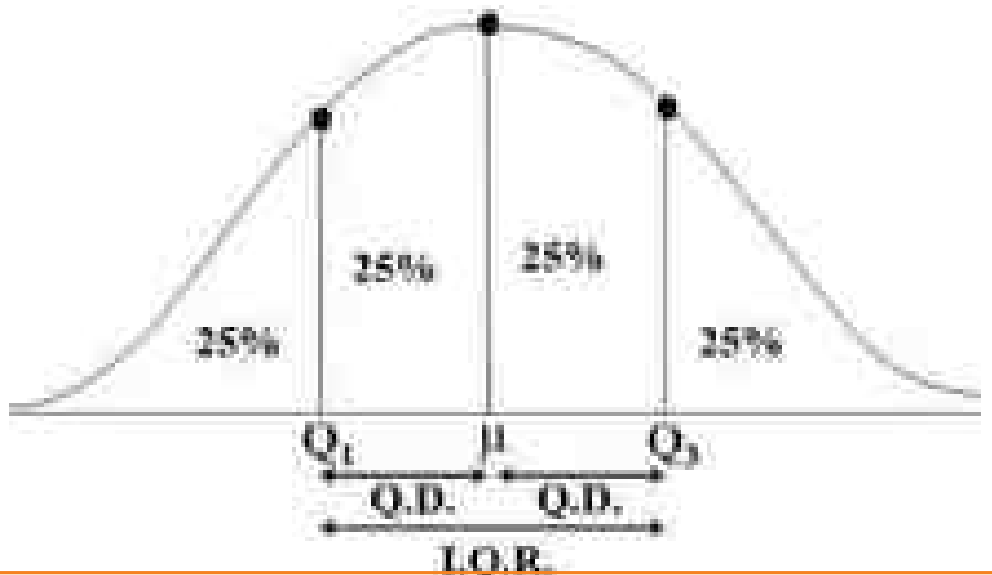
we can write

$$Q = \mu - Q_1$$

and also

$$Q = Q_3 - \mu$$

For any **symmetric** distribution,



As such, for the normal distribution, the following equation holds:

$$\int_{Q_1}^{Q_3} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2} \Rightarrow \int_{\mu-Q}^{\mu+Q} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

which can be re-written as

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-Q}^{\mu+Q} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

vulearninghub.xyz | by VU Learning HUB

We have

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{\mu-Q}^{\mu+Q} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = \frac{1}{2}$$

Putting, $z = \frac{x-\mu}{\sigma} \Rightarrow \sigma z = x-\mu \Rightarrow x = \mu + \sigma z$

$$\Rightarrow \frac{dx}{dz} = 0 + \sigma = \sigma \Rightarrow \sigma dz = dx$$

Change of limits:

$$\text{As } x \to \mu + Q \Rightarrow z \to \frac{\mu + Q - \mu}{\sigma} = \frac{Q}{\sigma}$$

$$\text{and as } x \to \mu - Q \Rightarrow z \to \frac{\mu - Q - \mu}{\sigma} = -\frac{Q}{\sigma}$$

So

substituting the values, we obtain

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-Q/\sigma}^{Q/\sigma} e^{-\frac{1}{2}(z)^2} .\sigma dz = \frac{1}{2}$$

Now, $e^{-\frac{1}{2}(z)^2}.\sigma$ is an even function

So

$$\frac{1}{\sigma\sqrt{2\pi}}\int_{-Q/\sigma}^{Q/\sigma} e^{-\frac{1}{2}(z)^2}.\sigma dz = \frac{1}{2}$$

can be re-written as

$$\frac{1}{\sigma\sqrt{2\pi}}2\int_{0}^{Q/\sigma} e^{-\frac{1}{2}(z)^2}.\sigma dz = \frac{1}{2} \Rightarrow \int_{0}^{Q/\sigma} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \frac{1}{4}$$

Now, from the area table of the standard normal distribution, we find that

$$\int_{0}^{0.6745} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = 0.2500$$

Hence, we can write

$$Q/\sigma = 0.6745$$

or

$$Q = 0.6745\ \sigma \qquad \ldots(1)$$

Therefore, using eq.(1), we obtain the values of the quartiles as follows:

$$Q_1 = \mu - 0.6745\sigma \quad \text{and} \quad Q_3 = \mu + 0.6745\sigma$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

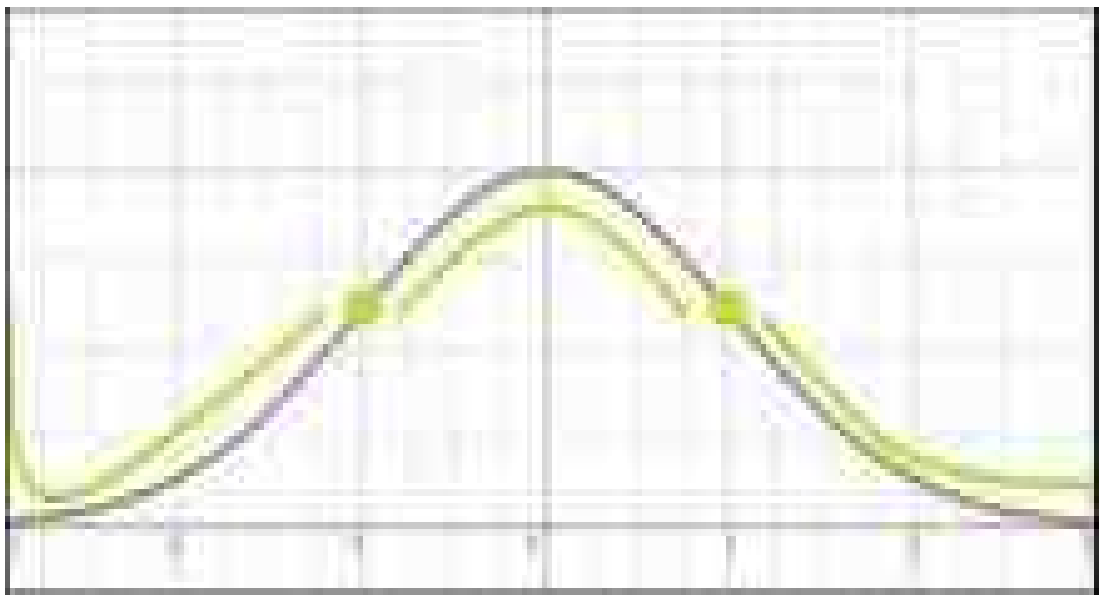
Topic No. 173

The normal curve has

Points of Inflection

which are equidistant from the mean

Definition: By a ‘point of inflection’ we mean a point at which the **concavity** of a curve changes,



From calculus, we know that such a point is obtained by solving the equation

$$f''(x) = 0$$

Taking the first derivative of the PDF of the normal distribution, we have

$$f'(x) = \frac{d}{dx}\left(\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\right) = \frac{1}{\sigma\sqrt{2\pi}} \frac{d}{dx}\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\right)$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \left\{ e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \frac{d}{dx}\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)\right\}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \left[\left(-\frac{1}{2}\right).2.\left(\frac{x-\mu}{\sigma}\right)\left(\frac{1}{\sigma}\right)\right]$$

$$= \frac{1}{\sigma\sqrt{2\pi}} . e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \left[-\left(\frac{x-\mu}{\sigma^2}\right)\right] = -\frac{1}{\sigma^2\sqrt{2\pi}}(x-\mu) . e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}$$

vulearninghub.xyz | by VU Learning HUB

- To find the **points of inflection**, we take the **second** derivative:

$$f''(x) = \frac{d}{dx}\left(-\frac{1}{\sigma^3\sqrt{2\pi}}(x-\mu).e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \right) = -\frac{1}{\sigma^3\sqrt{2\pi}}\left[\frac{d}{dx}\left((x-\mu).e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \right) \right]$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \left(\frac{d}{dx}(x-\mu) \right) + (x-\mu)\left(\frac{d}{dx}\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \right) \right) \right]$$

or

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}(1) + (x-\mu)\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\frac{d}{dx}\left(-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right)\right)\right]$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} - (x-\mu)\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\left(\frac{1}{2}.2\left[\frac{x-\mu}{\sigma}\right]\left[\frac{d}{dx}\left[\frac{x-\mu}{\sigma}\right]\right]\right)\right)\right]$$

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} - (x-\mu)\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\left(\frac{1}{2}.2\left[\frac{x-\mu}{\sigma}\right]\left[\frac{1}{\sigma}\right]\right)\right)\right]$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} - (x-\mu)\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}.\left(\frac{x-\mu}{\sigma^2}\right)\right)\right]$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}}\left[e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} - \frac{(x-\mu)^2}{\sigma^2}\left(e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\right)\right]$$

$$= -\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

so that, finally, we have:

$$f''(x) = -\frac{1}{\sigma^3\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \left[1 - \frac{(x-\mu)^2}{\sigma^2}\right]$$

Now, equating the second derivative to zero, we have

$$-\frac{1}{\sigma^3\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]=0$$

$$\Rightarrow\left[1-\frac{(x-\mu)^2}{\sigma^2}\right]=0\times\left(-\frac{\sigma^3\sqrt{2\pi}}{e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2}}\right)\Rightarrow 1-\frac{(x-\mu)^2}{\sigma^2}=0$$

$$\Rightarrow\frac{(x-\mu)^2}{\sigma^2}=1\Rightarrow(x-\mu)^2=\sigma^2$$

or $\quad x-\mu=\pm\sigma\Rightarrow x=\mu\pm\sigma$

or $\quad x=\mu+\sigma,\quad x=\mu-\sigma.$

At these two points, the value of the function $f(x)$ is

$$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\mu\pm\sigma-\mu}{\sigma}\right]^2}=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left[\frac{\pm\sigma}{\sigma}\right]^2}=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}[\pm1]^2}$$

$$=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}}=\frac{1}{\sigma\sqrt{2\pi}e^{\frac{1}{2}}}=\frac{1}{\sigma\sqrt{2\pi e}}.$$

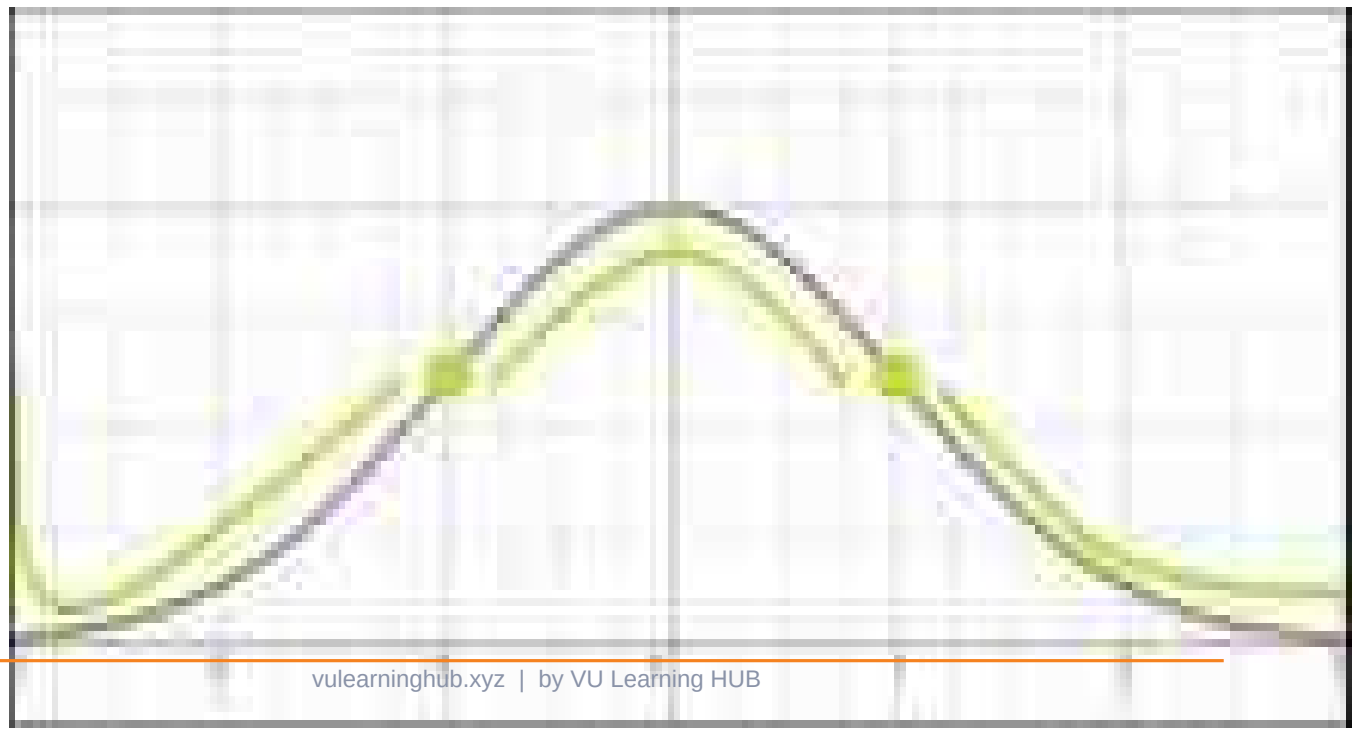
vulearninghub.xyz | by VU Learning HUB

Hence, the two points of inflection of normal curve are

$$\left[\mu-\sigma,\frac{1}{\sigma\sqrt{2\pi e}}\right] and \left[\mu+\sigma,\frac{1}{\sigma\sqrt{2\pi e}}\right].$$

In other words, the points of inflection occur on the **right** and on the **left** of the mean **at a distance equal to standard deviation** and thus the graph of the normal curve is **bell-shaped**.

The ‘bell curve’



Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 174

For the normal distribution,

the odd order moments about the mean are All zero

and

the even order moments are given by

$$\mu_{2n} = (2n-1)(2n-3)(2n-5)\ldots 5.3.1\,\sigma^{2n}$$

Theorem:

For the normal distribution, the **odd** order moments about the mean are all zero and the **even** order moments are given by

$$\mu_{2n} = (2n-1)(2n-3)\ldots 5.3.1\ \sigma^{2n}$$

Proof:

The **odd order moments** about the mean are given by

$$\mu_{2n+1} = E\left(X-\mu\right)^{2n+1} = \int_{-\infty}^{\infty}\left(x-\mu\right)^{2n+1}.\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}dx$$

where $n \in \square$.

Substituting, $z = \dfrac{x-\mu}{\sigma} \Rightarrow \sigma z = x-\mu \Rightarrow x = \mu+\sigma z$

$$\Rightarrow \frac{dx}{dz} = \sigma \Rightarrow \sigma dz = dx$$

Limits remain unchanged.

& we get,

$$\mu_{2n+1} = \int_{-\infty}^{\infty} (\sigma z)^{2n+1} \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \sigma dz = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (\sigma z)^{2n+1} e^{-\frac{1}{2}(z)^2} dz$$

$$= \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} e^{-\frac{z^2}{2}} dz = 0 \text{ as } z^{2n+1} e^{-\frac{z^2}{2}} \text{ is an odd function of } z.$$

Thus,

$$\mu_1 = \mu_3 = \mu_5 = \cdots = 0$$

$$\mu_{2n+1} = \frac{\sigma^{2n+1}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^{2n+1} e^{-\frac{z^2}{2}} dz = 0$$

as $z^{2n+1} e^{-\frac{z^2}{2}}$ is an odd function of z.

Thus,

$$\mu_1 = \mu_3 = \mu_5 = \cdots = 0$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 175

Derivation of the Variance of the Exponential Distribution

A very interesting property:

The **mean and standard deviation** of the exponential distribution are equal.

We know that, for the exponential distribution given by

$$f(x) = \lambda e^{-\lambda x}, \qquad x \geq 0, \lambda > 0$$

the mean is

$$E(X) = \frac{1}{\lambda}$$

1. Derivation of the **Variance:**

According to the short-cut formula

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2,$$

Now

$$E\left(X^2\right)=\int_0^{\infty} x^2.f\left(x\right)dx=\int_0^{\infty} x^2.\lambda e^{-\lambda x}dx$$

First of all we will apply the formula of integration by parts:

$$\int uvdx = u\int vdx - \int\left(\frac{du}{dx}\int vdx\right)dx$$

Here, we let $u = x^2$ and $v = \lambda e^{-\lambda x} = -(-\lambda e^{-\lambda x})$

so that $\int vdx = \int -(-\lambda e^{-\lambda x})dx = -\left(e^{-\lambda x}\right)$

and $\frac{du}{dx} = \frac{d}{dx}\left(x^2\right) = 2x.$

Hence, we have

$$\int x^2.\lambda e^{-\lambda x}dx = u\int vdx - \int\left(\frac{du}{dx}\int vdx\right)dx$$

$$= x^2(-e^{-\lambda x}) - \int 2x.\left(-e^{-\lambda x}\right)dx = -x^2e^{-\lambda x} - 2\int\left(-xe^{-\lambda x}\right)dx$$

$$= -x^2e^{-\lambda x} - \frac{2}{\lambda}\int\left(\lambda\left(-xe^{-\lambda x}\right)\right)dx$$

$$= -x^2e^{-\lambda x} - \frac{2}{\lambda}\int x\left(-\lambda e^{-\lambda x}\right)dx$$

Hence, we have

$$\int x^2.\lambda e^{-\lambda x}dx = -x^2e^{-\lambda x} + \frac{2}{\lambda}\int x\left(\lambda e^{-\lambda x}\right)dx$$

implying that

$$E\left(X^2\right) = \left[-x^2e^{-\lambda x}\right]_0^{\infty} + \frac{2}{\lambda}E\left(X\right)$$

or

$$E\left(X^2\right) = -\left[\frac{x^2}{e^{\lambda x}}\right]_0^{\infty} + \frac{2}{\lambda}E\left(X\right)$$

In order to apply the upper limit to $\left[\frac{x^2}{e^{\lambda x}}\right]$, we need to utilize the L'Hospital Rule.

We have $\frac{x^2}{e^{\lambda x}}$

Now, $\lim_{x\to\infty}\frac{x^2}{e^{\lambda x}} = \lim_{x\to\infty}\frac{2x}{\lambda e^{\lambda x}} = \lim_{x\to\infty}\frac{2}{\lambda^2 e^{\lambda x}} = \frac{2}{\lambda^2 e^{\lambda(\infty)}} = \frac{2}{\infty} = 0.$

On the other hand, $\lim_{x\to 0}\frac{x^2}{e^{\lambda x}} = \frac{(0)^2}{e^{\lambda(0)}} = \frac{0}{e^0} = \frac{0}{1} = 0.$

Hence, we have

$$E\left(X^2\right) = -\left[\frac{x^2}{e^{\lambda x}}\right]_0^{\infty} + \frac{2}{\lambda}E\left(X\right)$$

$$= -\left[0-0\right] + \frac{2}{\lambda}E\left(X\right)$$

$$= -0 + \frac{2}{\lambda}E\left(X\right) = \frac{2}{\lambda}E\left(X\right) = \frac{2}{\lambda}\left(\frac{1}{\lambda}\right) = \frac{2}{\lambda^2}$$

Hence,

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{2-1}{\lambda^2} = \frac{1}{\lambda^2}$$

implying that

$$S.D.(X) = \sigma = \frac{1}{\lambda}.$$

Hence we have the interesting property that the **mean and standard deviation** of the exponential distribution are equal.

This, in turn, means that **the Coefficient of Variation of the Exponential distribution is always equal to 1.**

Explanation: The coefficient of variation is defined as $CV = \frac{\sigma}{\mu}$.

Hence, for the exponential distribution given by

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

the coefficient of variation is

$$CV = \frac{\sigma}{\mu} = \frac{1/\lambda}{1/\lambda} = \frac{\lambda}{\lambda} = 1.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 176

Obtaining the first two Raw Moments of the Normal Distribution from its MGF

The first two Raw Moments of the Normal Distribution are given by

$$E(X) \ \& \ E(X^2).$$

We know that, the **mgf** of Normal Distribution is as

$$M(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, \qquad \text{for} \quad -\infty < t < \infty.$$

The first two derivatives of $M_X(t)$ are easily derived as,

$$M_X'(t) = \frac{d}{dt}\left[e^{\mu t + \frac{1}{2}\sigma^2 t^2} \right] = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \frac{d}{dt}\left[\mu t + \frac{1}{2}\sigma^2 t^2 \right]$$

$$= e^{\mu t + \frac{1}{2}\sigma^2 t^2} \left[\mu + \frac{1}{2}\sigma^2 2t \right] = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \left[\mu + \sigma^2 t \right]$$

Putting $t = 0$, we have

$$M_X'(0) = e^{\mu(0)+\frac{1}{2}\sigma^2(0^2)}\left[\mu + \sigma^2(0)\right]$$

$$= e^{0+0}(\mu + 0) = e^0(\mu) = (1)\mu = \mu.$$

But, we know that

$$M_X'(0) = E(X)$$

Therefore

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

And now taking the **second derivative** of $M_X(t)$ as follows:

$$M_X''(t)$$

$$= \frac{d}{dt}\left[M_X'(t) \right] = \frac{d}{dt}\left[e^{\mu t + \frac{1}{2}\sigma^2 t^2} \left[\mu + \sigma^2 t \right] \right] = \frac{d}{dt}\left[\mu e^{\mu t + \frac{1}{2}\sigma^2 t^2} + \sigma^2 t\, e^{\mu t + \frac{1}{2}\sigma^2 t^2} \right]$$

$$= \left[\mu e^{\mu t + \frac{1}{2}\sigma^2 t^2} \left(\mu + \frac{1}{2}\sigma^2 2t \right) \right] + \left[\sigma^2 \left(e^{\mu t + \frac{1}{2}\sigma^2 t^2} (1) + t \left(e^{\mu t + \frac{1}{2}\sigma^2 t^2} \left(\mu + \frac{1}{2}\sigma^2 2t \right) \right) \right) \right]$$

$$= \left[\mu^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} + \mu\sigma^2 t e^{\mu t + \frac{1}{2}\sigma^2 t^2} \right] + \left[\left(\sigma^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} + \sigma^2 \mu t e^{\mu t + \frac{1}{2}\sigma^2 t^2} + \sigma^4 t^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} \right) \right]$$

Putting t = 0, we have,

$$M_X''(0) = \left[\mu^2 e^{\mu(0)+\frac{1}{2}\sigma^2(0)} + \mu\sigma^2(0) e^{\mu t+\frac{1}{2}\sigma^2(0)} \right]$$

$$+\left[\left(\sigma^2 e^{\mu(0)+\frac{1}{2}\sigma^2(0)} + \sigma^2\mu(0) e^{\mu(0)+\frac{1}{2}\sigma^2(0)} + \sigma^4(0) e^{\mu(0)+\frac{1}{2}\sigma^2(0)} \right)\right]$$

$$= \left[\mu^2 e^0 + 0 \right] + \left[\sigma^2 e^0 + 0 + 0 \right]$$

$$= \mu^2(1) + \sigma^2(1) = \mu^2 + \sigma^2.$$

But, we know that

$$M_X''(0) = E\left(X^2\right)$$

Therefore

$$E\left(X^2\right) = \mu^2 + \sigma^2.$$

So, we find the **variance** as

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2$$

$$= \mu^2 + \sigma^2 - \mu^2 = \sigma^2.$$

NOTE:

All the higher moments can be derived in a similar manner.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 177

Cumulant Generating Function

of the $N(\mu,\sigma^2)$ distribution

and

its utilization for finding its cumulants

Definition of Cumulant Generating Function

For any distribution for which the mgf exists, the cumulant-generating function $K(t)$ is the natural logarithm of the moment generating function i.e.

$$K(t) = \ln E\left(e^{tX}\right)$$

For the **normal distribution** with expected value μ and variance σ^2, the cumulant generating function is

$$K(t) = \mu t + \frac{\sigma^2 t^2}{2}.$$

- The first derivative of the cumulant generating function is obtained as follows:

$$K'(t) = \frac{d}{dt}\left(\mu t + \frac{\sigma^2 t^2}{2}\right) = \left(\mu \frac{d}{dt} t + \frac{d}{dt}\frac{\sigma^2 t^2}{2}\right)$$

$$= \left(\mu(1) + \frac{\sigma^2}{2}(2t)\right) = \left(\mu + \sigma^2(t)\right) = \mu + \sigma^2 t$$

The second derivative of the cumulant generating function is obtained as follows:

$$K''(t) = \frac{d}{dt}\left(K'(t)\right) = \frac{d}{dt}\left(\mu + \sigma^2 t\right)$$

$$= \left(\frac{d}{dt}(\mu) + \sigma^2 \frac{d}{dt}(t)\right) = 0 + \sigma^2 (1) = \sigma^2.$$

The third derivative of the cumulant generating function is given by:

$$K'''(t) = \frac{d}{dt}\left(K''(t)\right) = \frac{d}{dt}\left(\sigma^2\right) = 0$$

implying that all higher derivatives will be zero.

Now, the first cumulant is obtained as follows:

$$K'(0) = \left(\mu + \sigma^2(t)\right)_{t=0} = \mu + \sigma^2(0) = \mu + 0 = \mu$$

Similarly, the second cumulant is:

$$K''(0) = \left(\sigma^2\right)_{t=0} = \sigma^2.$$

And the third cumulant is:

$$K'''(0) = (0)_{t=0} = 0.$$

vulearninghub.xyz/ by VU Learning HUB

Similarly, all the higher cumulants will be zero.

Thus, in a nutshell, the cumulants of the normal distribution with expected value μ and variance σ^2 are

$$K_1 = \mu,$$

$$K_2 = \sigma^2,$$

and

$$K_3 = K_4 = \dots = 0.$$

Virtual University of Pakistan

Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 178

For the normal distribution,

the odd order moments about the mean are All zero

and

the even order moments are given by

$$\mu_{2n} = (2n-1)(2n-3)(2n-5)\ldots 5.3.1\,\sigma^{2n}$$

Theorem:

For the normal distribution, the **odd** order moments about the mean are all zero and the **even** order moments are given by

$$\mu_{2n} = (2n-1)(2n-3)\ldots 5.3.1\ \sigma^{2n}$$

Proof:

The **even order moments** about the mean are obtained as follows:

$$\mu_{2n} = E\left(X-\mu\right)^{2n}$$

$$= \int_{-\infty}^{\infty} \left(x-\mu\right)^{2n} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

Substituting, $z = \frac{x-\mu}{\sigma} \Rightarrow \sigma z = x-\mu \Rightarrow x = \mu + \sigma z$

$$\Rightarrow \frac{dx}{dz} = \sigma \Rightarrow \sigma dz = dx,$$ Limits will remain **unchanged**.

Therefore,

$$\mu_{2n} = \int_{-\infty}^{\infty} (x-\mu)^{2n} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx$$

can be re-written as

$$\mu_{2n} = \int_{-\infty}^{\infty} (\sigma z)^{2n} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \sigma dz$$

And, Now, as we know that, for any **even function** g i.e. a function for which $g(-x) = g(x)\ \forall x \in \mathbb{R}$

the integral $\int_{-\infty}^{\infty} g(x)\,dx$ is equal to $2\int_{0}^{\infty} g(x)\,dx$.

Therefore, we have

$$\mu_{2n} = 2\int_0^{\infty} (\sigma z)^{2n} . \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2} \sigma dz = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} (\sigma z)^{2n} e^{-\frac{1}{2}(z)^2} dz$$

$$= \frac{2\sigma^{2n}}{\sqrt{2\pi}} \int_0^{\infty} \left(z^2\right)^n e^{-\frac{z^2}{2}} dz,$$

Let $y = \frac{z^2}{2} \Rightarrow 2y = z^2 \Rightarrow \sqrt{2y} = z$

Also, $y = \frac{z^2}{2} \Rightarrow \frac{dy}{dz} = \frac{2z}{2} = z \Rightarrow dy = zdz \Rightarrow \frac{dy}{z} = dz \Rightarrow \frac{dy}{\sqrt{2y}} = dz$

Limits remain unchanged.

Therefore, $\mu_{2n} = \frac{2\sigma^{2n}}{\sqrt{2\pi}} \int_0^\infty (2y)^n e^{-y} \frac{dy}{\sqrt{2y}} = \frac{2\sigma^{2n}}{\sqrt{2\pi}} \int_0^\infty (2y)^{n-\frac{1}{2}} e^{-y} dy$

$= \frac{2\sigma^{2n}}{\sqrt{2\pi}} \int_0^\infty 2^{n-\frac{1}{2}} (y)^{n-\frac{1}{2}} e^{-y} dy = \frac{\sigma^{2n}}{\sqrt{2\pi}} \int_0^\infty 2^{n+1-\frac{1}{2}} (y)^{n-1/2} e^{-y} dy$

or

$$\mu_{2n} = \frac{\sigma^{2n}}{\sqrt{2\pi}} 2^{n+\frac{1}{2}} \int_0^{\infty} y^{n-1/2} e^{-y} dy = \frac{\sigma^{2n}}{\sqrt{\pi}} .2^{-\frac{1}{2}} .2^{n+\frac{1}{2}} \int_0^{\infty} y^{n-1/2} e^{-y} dy$$

$$= \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^{\infty} y^{n-1/2-1+1} e^{-y} dy = \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^{\infty} y^{n+1/2-1} e^{-y} dy$$

$$= \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \int_0^{\infty} y^{(n+1/2)-1} e^{-y} dy$$

$$\mu_{2n} = \frac{2^n \sigma^{2n} \Gamma\left(n+1/2\right)}{\sqrt{\pi}}$$

$$= \frac{2^n \sigma^{2n}}{\sqrt{\pi}} \cdot \left(\frac{2n-1}{2}\right)\left(\frac{2n-3}{2}\right)\cdots\left(\frac{5}{2}\right)\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}$$

$$\left(\because \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}\right)$$

$$\mu_{2n} = \left(2n-1\right)\left(2n-3\right)\ldots 5.3.1\ \sigma^{2n}$$

Putting $n = 1$ *and* 2, *we get* $\mu_2 = \sigma^2$ *and* $\mu_4 = 3\sigma^4$.

Hence $\beta_1=0$ meaning that skewness is zero,

and

$$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3,$$

i.e. the normal curve has zero kurtosis.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 178

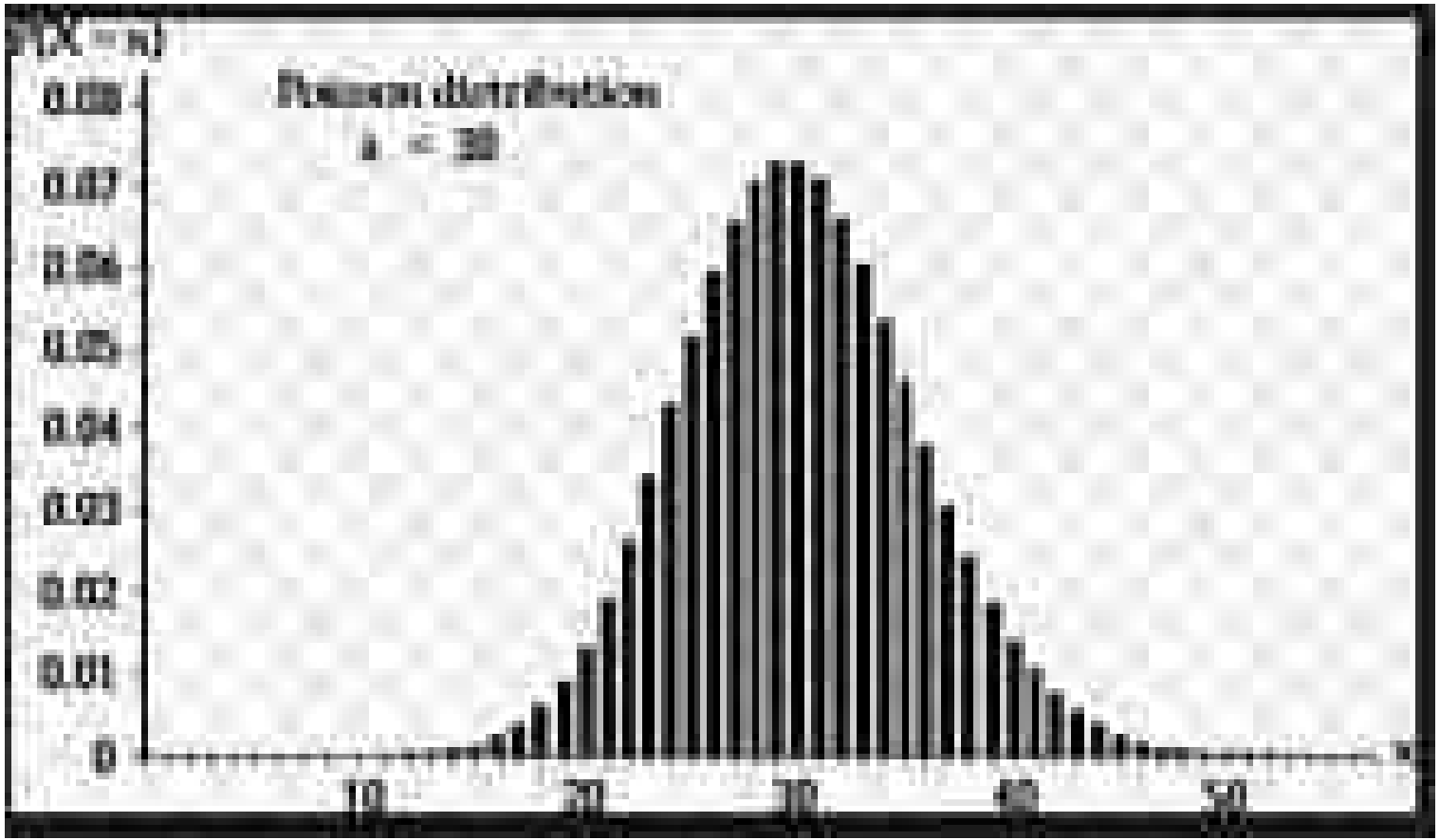
An example of

the

Normal Approximation to the Poisson distribution

Normal Approximation to the Poisson distribution

- It has been mathematically proved that, in the case of a Poisson Process leading to a Poisson distribution with mean=variance= λ, if λ tends to infinity, then the Poisson distribution tends to the Poisson distribution.


Poisson distribution
λ = 30

- In practice, for sufficiently large values of λ, (say $\lambda>1{,}000$), the **Normal**($\mu = \lambda, \sigma^2 = \lambda$) **Distribution** is an excellent **approximation to the Poisson**(λ) **Distribution**.

- If λ is greater than about 10, then the **Normal Distribution** is a good **approximation** if an appropriate continuity correction is performed.

Example:

The number of calls received by an office switch board per hour follows a Poisson distribution with parameter 25.

Find the probabilities that in one hour

(a) There are between 23 and 26 calls (inclusive),

(b) more than 30 calls, using the **normal approximation** to the Poisson distribution.

Let X be the random variable, the number of calls received in one hour.

Then, X is $P(x\ ;\ 25)$.

We require (a) $P(23 \leq X \leq 26)$ and (b) $P(X>30)$.

Using the normal approximation, X is $N(25,25)$

(a) P(23≤X≤26) becomes on continuous scale P(22.5≤X≤26.5) , X is N(25,25).

The z-values are: At, $x = 22.5,\ z = \dfrac{22.5 - 25}{5} = -0.5,$ and

at $x = 26.5,\ z = \dfrac{26.5 - 25}{5} = 0.3,$

$$\therefore \quad P\left(22.5 \le X \le 26.5\right) = P\left(-0.5 \le Z \le 0.3\right)$$

$$= P\left(0 < Z < 0.5\right) + P\left(0 < Z < 0.3\right) = 0.1915 + 0.1179$$

$$= 0.3094 = 30.94\%.$$

(b) P(X >30) becomes on continuous scale P(X >29.5)

$$\therefore \quad P\left(X > 29.5\right) = P\left(\frac{X-25}{5} > \frac{29.5-25}{5}\right)$$

$$= P\left(Z > \frac{4.5}{5}\right) = P\left(Z > 0.9\right) = 1 - P\left(Z < 0.9\right)$$

$$= 1 - 0.8159 = 0.1841 = 18.41\%.$$

vulearninghub.xyz | by VU Learning HUB

I would like to encourage you to compute the exact probabilities by using the pmf of the Poisson distribution

and then compare the exact probabilities with those obtained through the Normal Approximation to the Poisson distribution.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 180

Bivariate Normal Distribution

(PDF and Shape of the distribution & a few real-life examples)

Definition

A pair of random variables X_1 and X_2 have a bivariate normal distribution and they are referred to as **jointly normally distributed** random variables **if and only if** their joint probability density is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$,

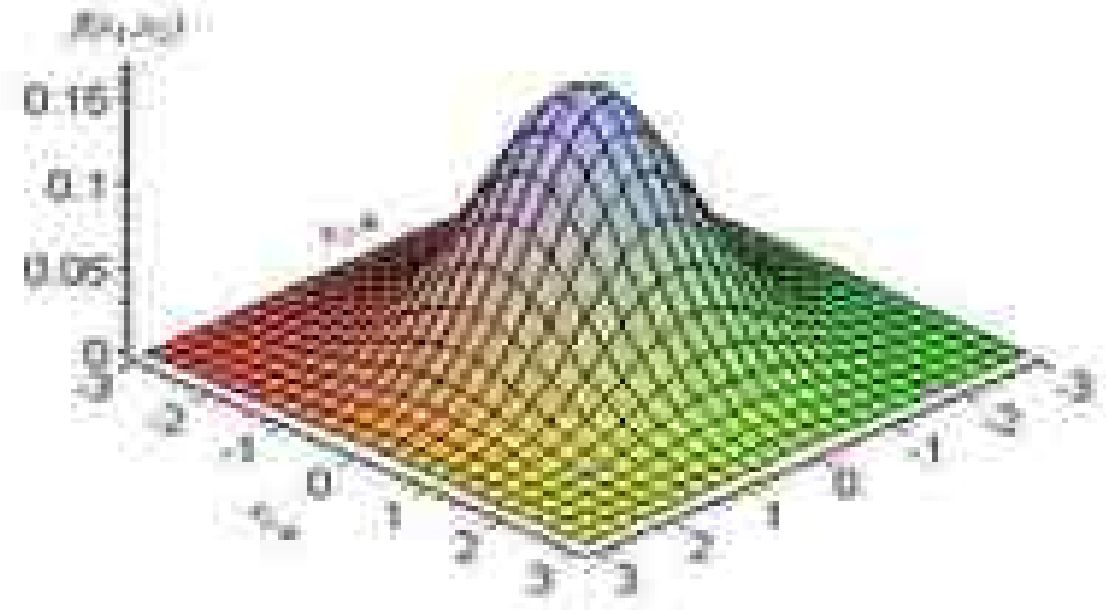
where $\mu_1 \in \square$, $\mu_2 \in \square$, $\sigma_1 > 0$, $\sigma_2 > 0$, and $-1 \le \rho \le 1$.

vulearninghub.xyz | by VU Learning HUB

This is referred to as the distribution. $N\left(\mu_x,\ \mu_y,\ \sigma_x^2, \sigma_y^2, \rho\right)$

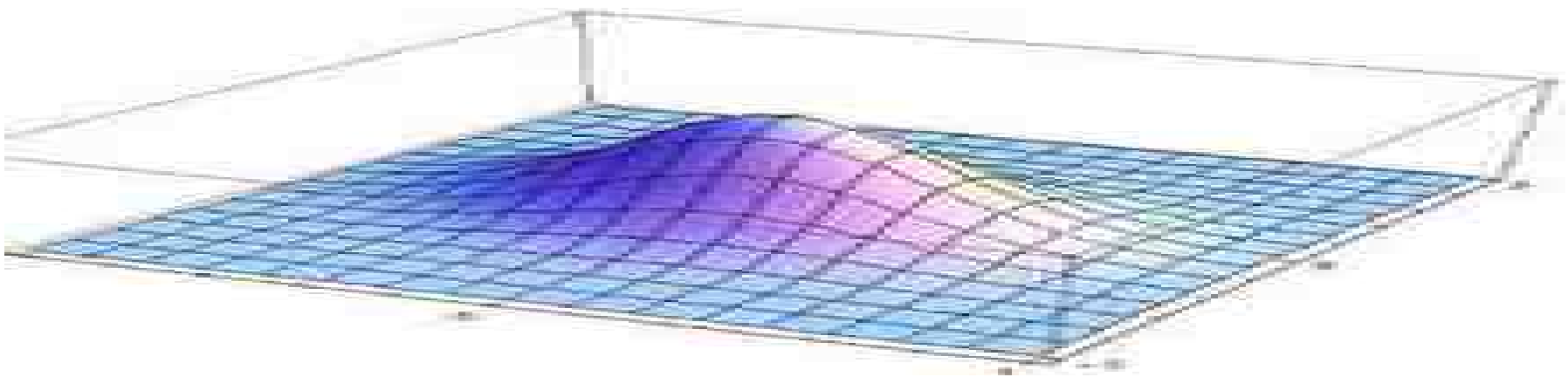
This joint p.d.f., has a bell-type shape.

Bivariate Normal Distribution with $\rho = 0$ and $\sigma_1 = \sigma_2 = \sigma$:



An exact, beautiful bell-shaped surface

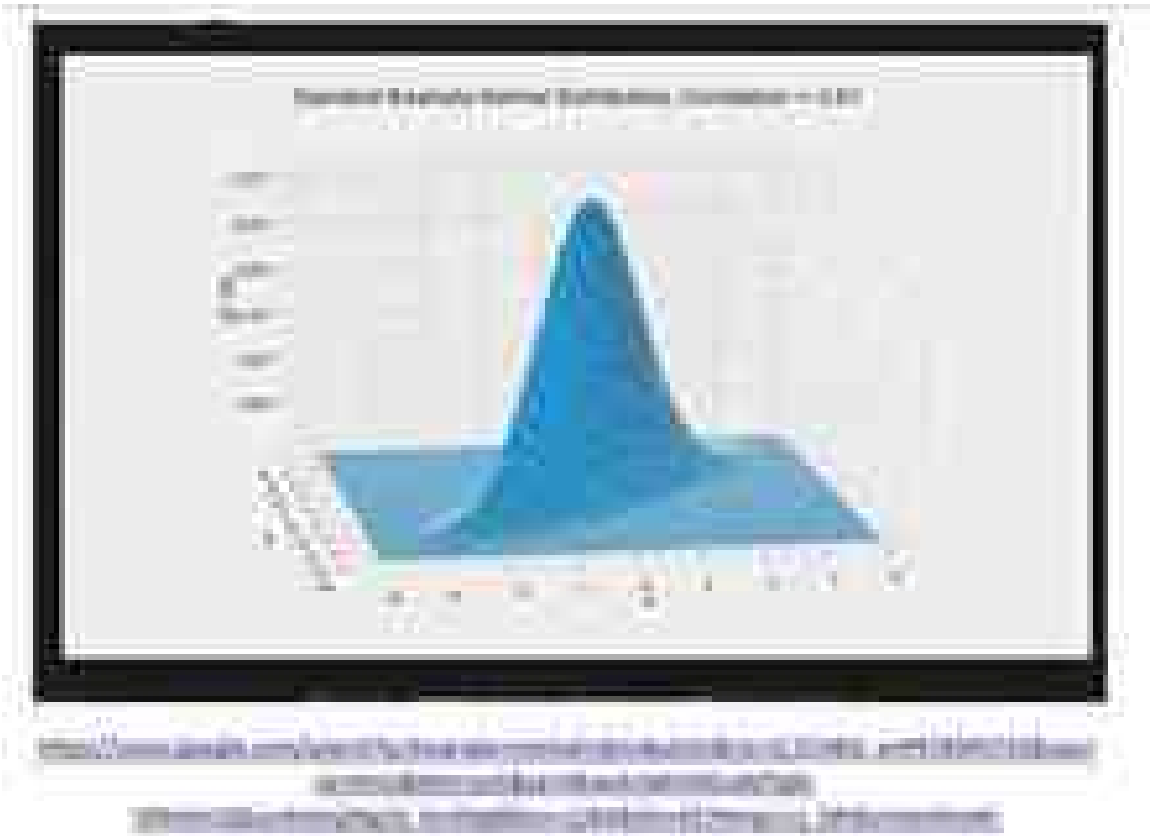
Bivariate Normal Distribution with $\rho = 0$ and $\sigma_1 \neq \sigma_2$:



Even then, it is almost a bell-type surface

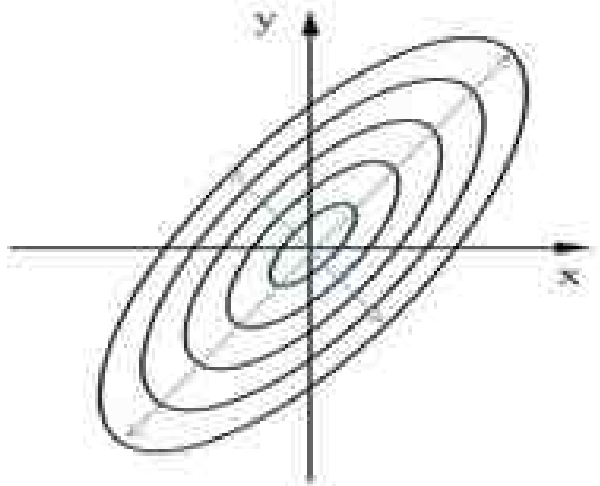
vulearninghub.xyz | by VU Learning Hub

Bivariate Normal Distribution with $\rho = 0.87$



‘Base’

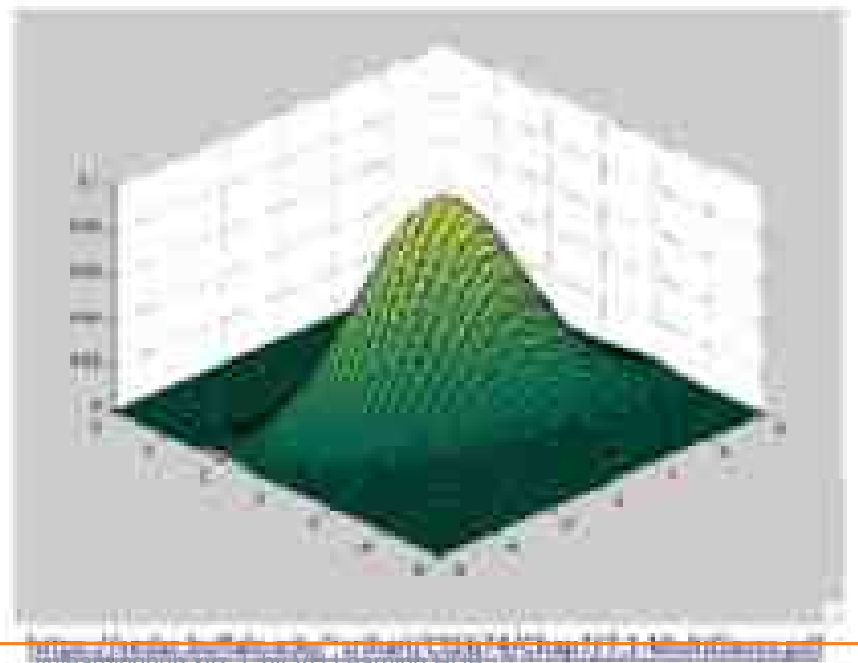
of the Bivariate Normal Distribution with $\rho = 0.87$



The diagram above needs to be ACCORDING to the bell-type shape in the previous graph (should be “thinner”).

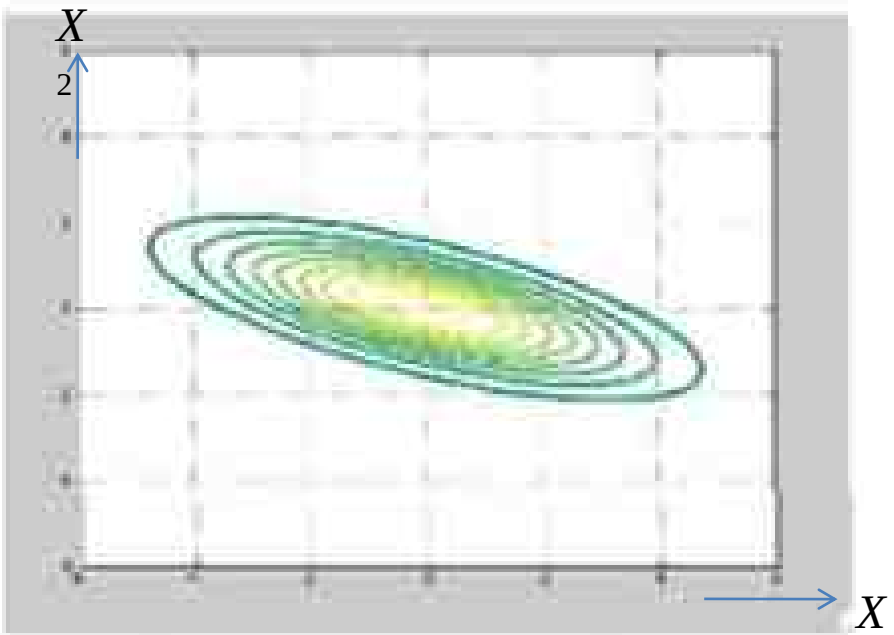
vulearninghub.xyz | by VU Learning HUB

Bivariate Normal Distribution with ρ = - 0.67



‘Base’

of the Bivariate Normal Distribution with ρ = - 0.67



Explanation

Conceive it:

as if we have a ‘bell-type’ that is placed on the FLOOR.

The concept of

the ‘X_1X_2 floor’

What is meant by the 'X_1X_2 floor'?

Consider the three-dimensional space for example the room in which you are sitting right now.

- Now focus on the floor of the room.
- One 'edge' of the floor will be called the X_1 axis
- The other 'edge' of the floor will be called the X_2 axis

And, as such, the floor will be called the 'X_1X_2 floor'

(actually it is the X_1X_2-plane)

Next:

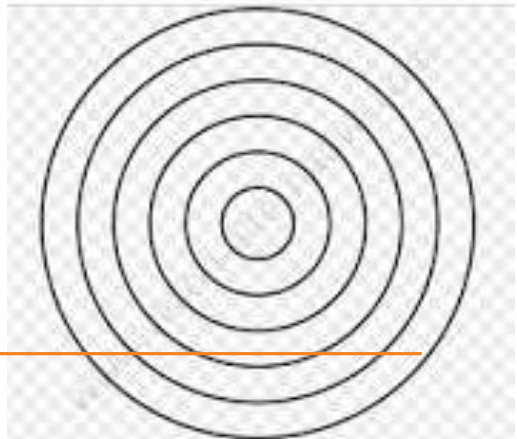
The concept of

Contours

By the term **‘contour’**, we mean the **set of points on the X_1X_2 floor** corresponding to which **the ordinates** are of **equal height.**

Explanation: Consider the bivariate normal distribution for which ρ = **0** and $\sigma_1 = \sigma_2 = \sigma$.

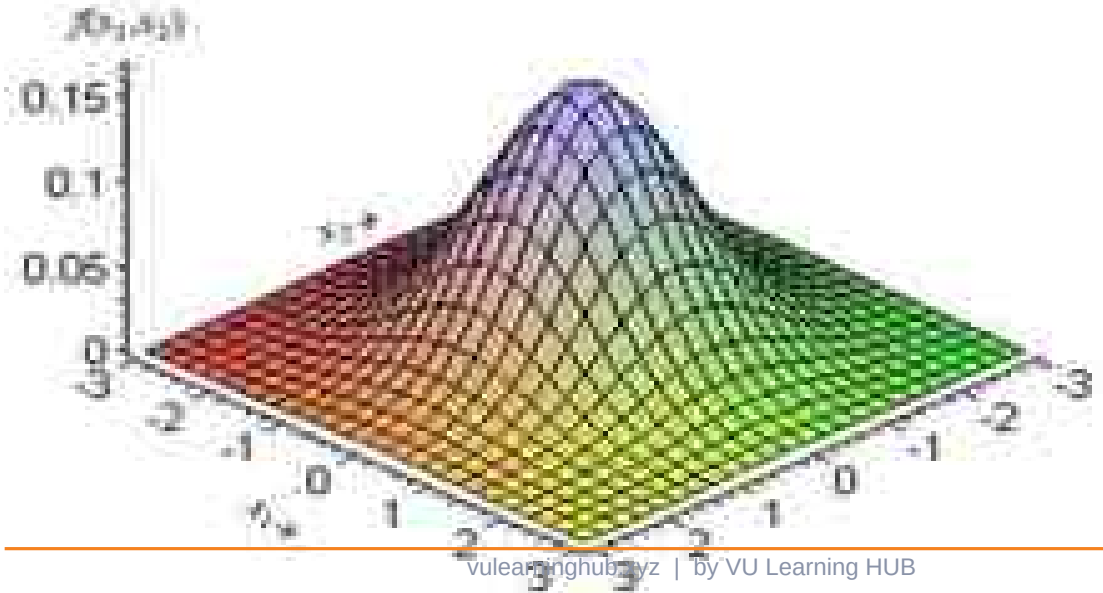
In this case the **contours** of the distribution will be **concentric circles** because, against **each** of these **circles** (on the X_1X_2 floor), the ordinates are of **equal height.**



- The ordinates against the outer circle are very short;
- The ordinates against the Second circle (inner circle) are a little taller than the ones against the outer circle;
- The ordinates against the Third (inner) circle are taller than the ones against the Second circle;
- The ordinates against the Fourth (inner) circle are taller than the ones against the Third circle;

And so on.

The **circular contour** corresponding to a lower height is **wider** than the one corresponding to a greater height.



Next:

A circle can be regarded as

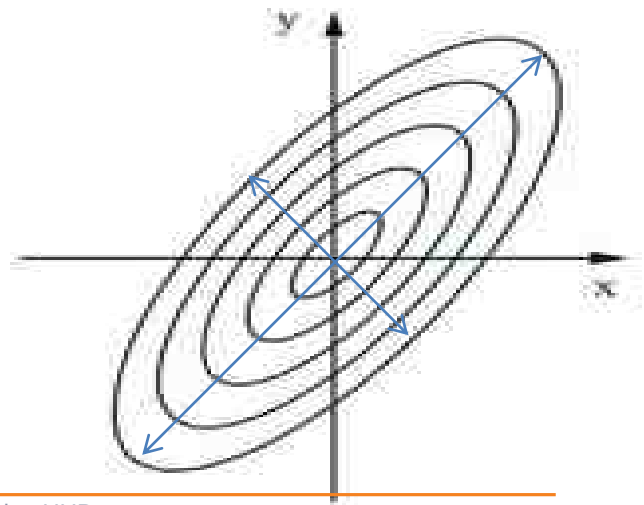
a special case

of an **ELLIPSE.**

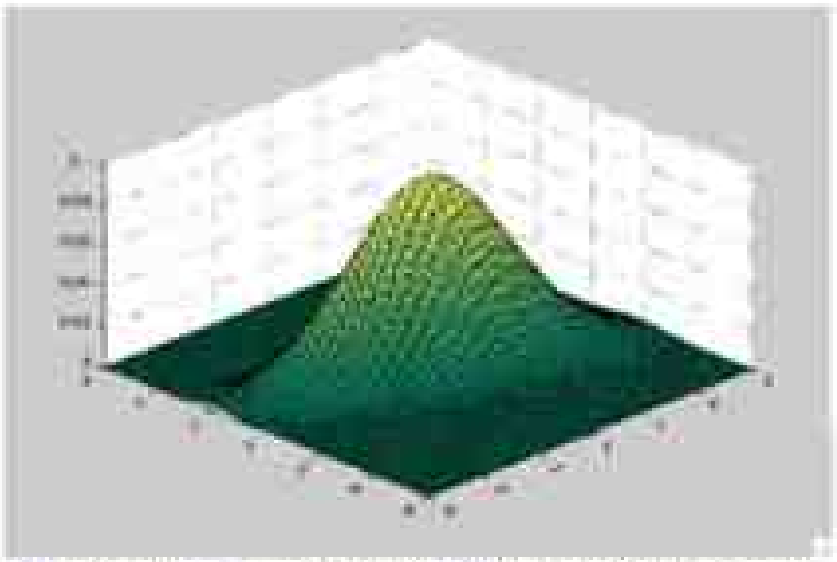
General Concept of Ellipse:

We may assume that **one** of the many diameters of the circle is being **‘elongated’**.

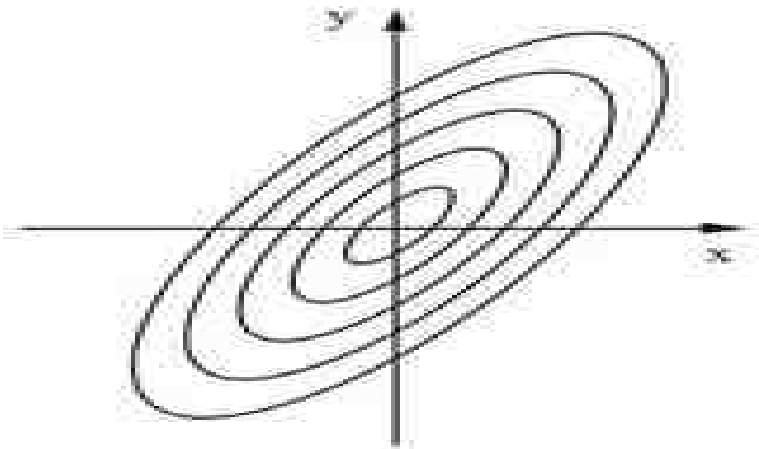
As soon as we ‘elongate’ one of the diameters, we get a **major axis** as well as a **minor axis**.

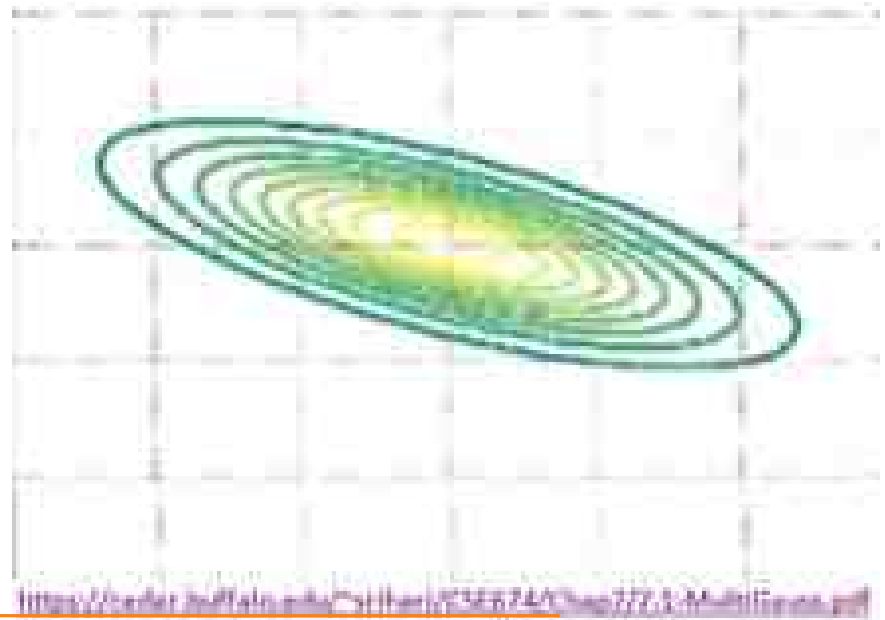


Bivariate Normal Distributions with $\rho = 0.87$ (to the left) and $\rho = -0.67$ (to the right)



Bivariate Normal Distributions with $\rho = 0.87$ (to the left) and $\rho = -0.67$ (to the right)





A few real-life examples

- Heights and weights
- Weights and Blood Pressures
- Marks in Mathematics & Marks in Statistics

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 181

Bivariate Normal Distribution

Detailed discussion of the Shape of the distribution when $\rho = 0$)

RECALL Definition

Joint probability density is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$,

where $\mu_1 \in \square$, $\mu_2 \in \square$, $\sigma_1 > 0, \sigma_2 > 0$, and $-1 \leq \rho \leq 1$.

PDF of the bivariate normal distribution for the case $\rho = 0$.

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-0^2}} e^{-\frac{1}{2(1-0^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2(0)\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1}} e^{-\frac{1}{2(1)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 0 + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

Now, since $f(x_1,x_2)$ represents the **ordinate** of the distribution against the point (x_1,x_2) on the X_1X_2-floor, therefore, for any particular **contour**, we can write:

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}=\underline{\text{constant}}=c$$

$$\text{Now } \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \underline{\text{constant}} = c$$

$$\Rightarrow e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = (2\pi\sigma_1\sigma_2).c = \text{a new constant}$$

$$\Rightarrow \ln e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \ln[(2\pi\sigma_1\sigma_2).c] = \text{another new constant} = e$$

$$\text{or } -\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right) = e$$

vulearninghub.xyz | by VU Learning HUB

Further,

$$-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=e \Rightarrow -\frac{1}{2e}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=1$$

$$\Rightarrow \frac{1}{k}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=1 \text{ where } k=-2e$$

Hence, we have

$$\frac{(x_1-\mu_1)^2}{k\sigma_1^2}+\frac{(x_2-\mu_2)^2}{k\sigma_2^2}=1 \qquad \ldots(2)$$

RECALL

the Mathematical Equation of An Ellipse

The equation of an ellipse with major axis parallel to the X_1-axis and minor axis parallel to the X_2-axis and having its center at (c_1,c_2) is as follows:

$$\frac{(x_1-c_1)^2}{a^2}+\frac{(x_2-c_2)^2}{b^2}=1 \qquad \ldots(1)$$

Next:

The Shape of the distribution

when

$\rho = 0$ **AND** $\sigma_1 = \sigma_2$

In the special case $a = b$, equation (1) reduces to

$$\frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{a^2} = 1 \Rightarrow (x_1 - c_1)^2 + (x_2 - c_2)^2 = a^2$$

and we **already** know that

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 = a^2$$

is the equation of **that particular circle** with center at the point (c_1, c_2) and with radius a.

vulearninghub.xyz | by VU Learning HUB

Hence, if we put $\sigma_1 = \sigma_2 = \sigma$ in equation (2), we obtain

$$\frac{(x_1 - \mu_1)^2}{k\sigma^2} + \frac{(x_2 - \mu_2)^2}{k\sigma^2} = 1$$

$$\Rightarrow (x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 = k\sigma^2 \qquad \ldots(3)$$

which is the equation of a circle

with center at (μ_1, μ_2) and with radius = $k\sigma$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 187

Conditional distributions in the case of Bivariate Normal Distribution

(Graphical Interpretation)

Graphical interpretation

of the concept of

conditional

distributions

associated with

the

bivariate normal distribution.

So, let us start with the pdf of the bivariate normal distribution:

We know that the **joint p.d.f.** of Bivariate Normal distribution (X, Y) is

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}$$

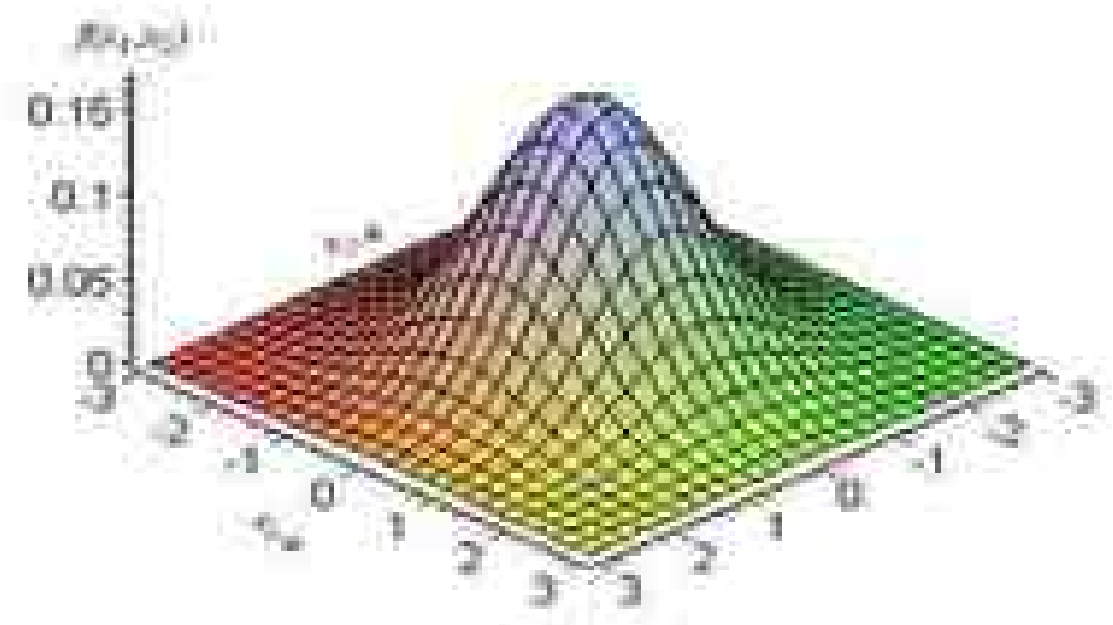
where $-\infty < x, y < \infty$ and the parameters are such that

$-\infty < \mu_x, \mu_y < \infty;$

$\sigma_x, \sigma_y > 0;$

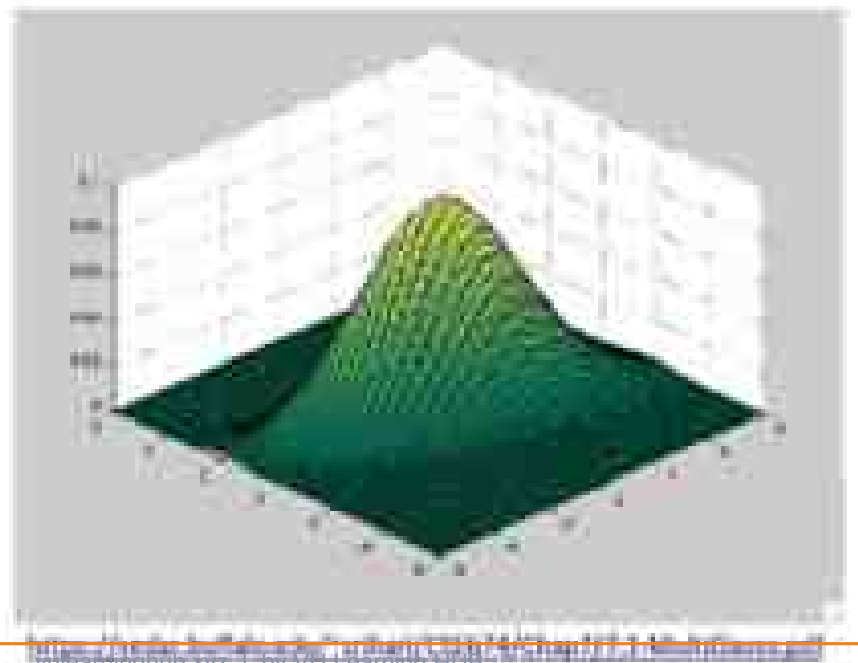
$-1 < \rho < 1.$

Bivariate Normal Distribution with $\rho = 0$ and $\sigma_1 = \sigma_2 = \sigma$:



An exact, beautiful bell-shaped surface

Bivariate Normal Distribution with ρ = - 0.67



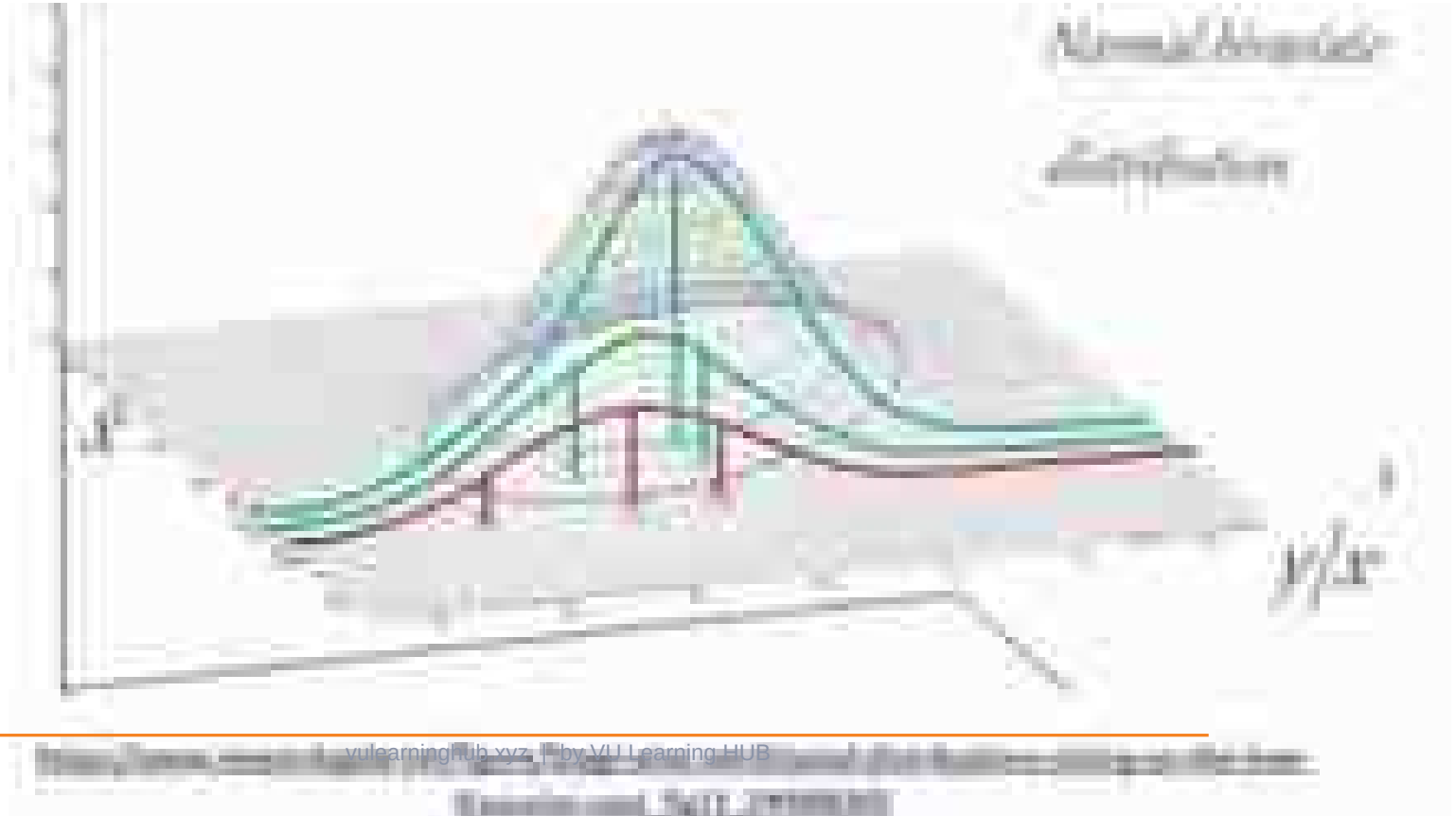
As far as the

Graphical interpretation

of the conditional distributions

associated with the bivariate normal distribution are concerned,

please consider the following diagram:



vulearninghub.xyz, Powered by VU Learning HUB

Important Note:

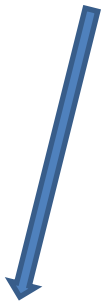
In equation (1) i.e.

$$f_{Y|X}(y|X=x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\left[\mu_y+(x-\mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho\right]\right)^2}$$

we are **talking about**

considering the distribution of Y against a **given** value of X which we are denoting by x.

In other words, the '*x*' occurring in the **exponent**



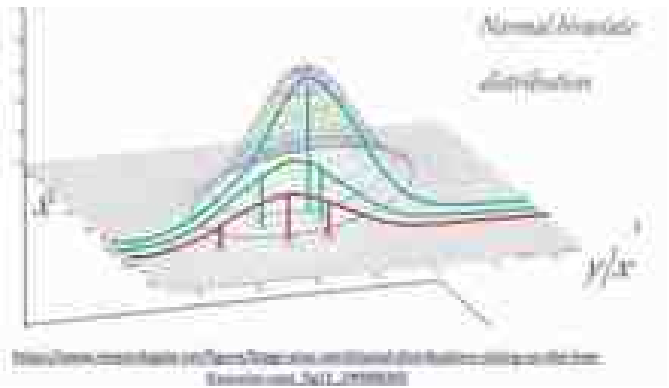
$$f_{Y \mid X}(y \mid X = x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\left[\mu_y+(x-\mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho\right]\right)^2}$$

is a **particular value** of *X*, i.e. it is a **constant.**

Because of the above, we can see that:

As far as the **mean** of the random variable Y corresponding to any particular value of X is concerned,

$$\tilde{\mu}_y = \mu_y + (x - \mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho$$



represents the mean value of y against some **particular value** of x.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 185

In the case of a Bivariate Normal Distribution,
the random variables X and Y are
independent
IF AND ONLY IF
the Correlation Coefficient
$\boldsymbol{\rho = 0}$

Now we will attempt to prove that

In the case of a Bivariate Normal Distribution,

the random variables X and Y are

independent

IF AND ONLY IF

the Correlation Coefficient

$\boldsymbol{\rho = 0}$

Theorem:

If X and Y have a bivariate normal distribution with correlation coefficient ρ. Then X and Y are independent **if and only if** $\rho = 0$.

The statement "if and only if" means:

1) If X and Y are independent, then $\rho = 0$,

&

2) If $\rho = 0$, then X and Y are independent.

Note that, **in general**, if X and Y are **independent** random variables, **then** their correlation coefficient is 0.

(In other words, it is **ALWAYS** true that in the case of two **independent** random variables X and Y, the correlation coefficient ρ is equal to **zero**.)

As such, there is **no real need** to prove the **first** part of the statement **“if and only if”.**

The **actual goal** is to prove the **second** part of the statement **"if and only if"**

i.e.

In the case of a **bivariate normal distribution** $N\left(\mu_x,\ \mu_y,\ \sigma_x^2, \sigma_y^2, \rho\right)$,

if $\rho = 0$, **then** X and Y are **independent**.

the basic condition by which two jointly distributed random variables X and Y can be regarded as being statistically independent.

Definition of Independence:

Two **random** variables X and Y are **independent, statistically independent** or **stochastically independent** if

$$f_{XY}(x, y) = f_X(x) f_Y(y) \qquad \ldots(1)$$

i.e.

if the **joint** density $f_{XY}(x, y)$ is equal to the product of the **marginal** density $f_X(x)$ and the **marginal** density $f_Y(y)$.

Note that, due to equation (1), if X and Y are **independent**, then the conditional distribution of Y given $X = x$ i.e. $f_{Y|X}(y \mid X = x)$ can be written as

$$f_{Y|X}(y \mid X = x) = \frac{f_{XY}(x, y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y).$$

So, in the case of independent random variables (X and Y), **conditioning on $X = x$ does not change the distribution of Y.**

(i.e. the conditional distribution of Y is the SAME as the unconditional distribution of Y)

The **proof** of the **second** part of the statement **“if and only if”**

i.e.

In the case of a **bivariate normal distribution**

if $\rho = 0$, **then** X and Y are **independent**.

$$N\left(\mu_x,\ \mu_y,\ \sigma_x^2,\ \sigma_y^2,\ \rho\right),$$

Proof: In order to prove that if X and Y have the bivariate normal distribution with **zero** correlation, **then** X and Y are **independent,** we need to show that the bivariate normal density function

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}$$

factorizes into the **product** of the marginal p.d.f of X and the marginal p.d.f. of Y.

Substituting $\rho = 0$ in the pdf of the bivariate normal distribution, we have

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-0^2}} e^{-\frac{1}{2(1-0^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2(0)\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}$$

which reduces to

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

Further simplifying, we have:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$= \frac{1}{\sqrt{2\pi}\sqrt{2\pi}\sigma_x\sigma_y} e^{-\frac{1}{2}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right]} \times e^{-\frac{1}{2}\left[\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \times \frac{1}{\sqrt{2\pi}\sigma_y} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2} = f_X(x)\, f_Y(y)$$

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 181

Bivariate Normal Distribution

Detailed discussion of the Shape of the distribution when $\rho = 0$)

RECALL Definition

Joint probability density is given by

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_2 < \infty$,

where $\mu_1 \in \square$, $\mu_2 \in \square$, $\sigma_1 > 0, \sigma_2 > 0$, and $-1 \leq \rho \leq 1$.

PDF of the bivariate normal distribution for the case $\rho = 0$.

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)\right\}}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-0^2}} e^{-\frac{1}{2(1-0^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 2(0)\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]}$$

$$= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1}} e^{-\frac{1}{2(1)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} - 0 + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}$$

Now, since $f(x_1,x_2)$ represents the **ordinate** of the distribution against the point (x_1,x_2) on the X_1X_2-floor, therefore, for any particular **contour**, we can write:

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2}e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)}=\underline{\text{constant}}=c$$

$$\text{Now } \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \underline{\text{constant}} = c$$

$$\Rightarrow e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = (2\pi\sigma_1\sigma_2).c = \text{a new constant}$$

$$\Rightarrow \ln e^{-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)} = \ln[(2\pi\sigma_1\sigma_2).c] = \text{another new constant} = e$$

$$\text{or } -\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right) = e$$

vulearninghub.xyz | by VU Learning HUB

Further,

$$-\frac{1}{2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=e \Rightarrow -\frac{1}{2e}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=1$$

$$\Rightarrow \frac{1}{k}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right)=1 \text{ where } k=-2e$$

Hence, we have

$$\frac{(x_1-\mu_1)^2}{k\sigma_1^2}+\frac{(x_2-\mu_2)^2}{k\sigma_2^2}=1 \qquad \ldots(2)$$

RECALL

the Mathematical Equation of An Ellipse

The equation of an ellipse with major axis parallel to the X_1-axis and minor axis parallel to the X_2-axis and having its center at (c_1,c_2) is as follows:

$$\frac{(x_1-c_1)^2}{a^2}+\frac{(x_2-c_2)^2}{b^2}=1 \qquad \ldots(1)$$

Next:

The Shape of the distribution when

$$\rho = 0 \quad \textbf{AND} \quad \sigma_1 = \sigma_2$$

In the special case $a = b$, equation (1) reduces to

$$\frac{(x_1 - c_1)^2}{a^2} + \frac{(x_2 - c_2)^2}{a^2} = 1 \Rightarrow (x_1 - c_1)^2 + (x_2 - c_2)^2 = a^2$$

and we **already** know that

$$(x_1 - c_1)^2 + (x_2 - c_2)^2 = a^2$$

is the equation of **that particular circle** with center at the point (c_1, c_2) and with radius a.

vulearninghub.xyz | by VU Learning HUB

Hence, if we put $\sigma_1 = \sigma_2 = \sigma$ in equation (2), we obtain

$$\frac{(x_1 - \mu_1)^2}{k\sigma^2} + \frac{(x_2 - \mu_2)^2}{k\sigma^2} = 1$$

$$\Rightarrow (x_1 - \mu_1)^2 + (x_2 - \mu_2)^2 = k\sigma^2 \qquad \ldots(3)$$

which is the equation of a circle

with center at (μ_1, μ_2) and with radius = $k\sigma$.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 182

Computation of Probabilities

in the case of

Bivariate Normal Distribution

explained through an

EXAMPLE

Example:

A statistics class takes two exams, Exam 1 (Midterm Exam) and Exam 2 (Final Exam), and let us suppose that the marks of the two exams (to be called X and Y respectively) follow a **bivariate normal distribution** with parameters:

$\mu_x = 70$ and $\mu_y = 60$ (which can be called the marginal means)

$\sigma_x = 10$ and $\sigma_y = 15$ (which can be called the marginal standard deviations)

and $\rho = 0.6$ (which is the correlation coefficient).

Suppose that we select a student **at random.**

1) What is the probability that the student scores **over 75** on **Exam 1 (Midterm Exam)**?

2) What is the probability that the student scores over 85 on Exam 2 (Final Exam) **given that** he/she scored 75 in Exam 1 (Midterm Exam)?

Solution of Part (1):

Note that marks obtained in Exam 1 (denoted by X) follow a **univariate normal distribution** with

mean = 70 and variance = 100 i.e. $X \sim N(70, 100)$.

Hence, the probability that a randomly selected student scores over 75 in Exam 1 is given by

$$P(X > 75) = P\left(Z > \frac{x-\mu_x}{\sigma_x}\right) = P\left(Z > \frac{75-70}{10}\right) = P(Z > 0.5)$$

$$= 1 - P(Z < 0.5) = 1 - \Phi(0.5) = 1 - 0.6915 = 0.3085 = 30.85\% \text{ or } 31\%$$

--- a little less than one-third.

Next,

Part (2) of the question is as follows:

Suppose that we select a student **at random.**

What is the probability that the student scores over 85 on Exam 2 (Final Exam) **given that** he/she scored 75 in Exam 1 (Midterm Exam)?

Solution of Part (2):

Note that the distribution of marks obtained in Exam 2 (denoted by Y) given that the marks in Exam 1 were 75 is the **conditional distribution** of Y given $X=x=75$.

The pdf of the conditional distribution of Y given $X=x$ is

$$f_{Y|X}(y \mid X = x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\left[\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right]\right)^2}$$

It is easy to see that the conditional distribution of Y given x i.e. $f_{Y|X}(y|x)$ is **normal** with mean $\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)$ and variance $\sigma_y^2(1-\rho^2)$.

Here, $\mu_x = 70$, $\mu_y = 60$, $\sigma_x = 10$, $\sigma_y = 15$ and $\rho = 0.6$ Hence, we have

$$f_{Y|X}(y \mid X = x)$$

$$= \frac{1}{\sqrt{2\pi}(15)\sqrt{1-(0.6)^2}} e^{-\frac{1}{2(15)^2(1-(0.6)^2)}\left(y-\left[60+(0.6)\frac{15}{10}(x-70)\right]\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}(15)\sqrt{1-0.36}} e^{-\frac{1}{2(15)^2(1-0.36)}\left(y-\left[60+(0.6)(1.5)(x-70)\right]\right)^2}$$

or

$$f_{Y|X}(y \mid X = x) = \frac{1}{\sqrt{2\pi}(15)\sqrt{0.64}} e^{-\frac{1}{2(15)^2(0.64)}\left(y-\left[60+(0.9)(x-70)\right]\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}(15)(0.8)} e^{-\frac{1}{2(225)(0.64)}\left(y-\left[60+(0.9)(x-70)\right]\right)^2}$$

$$= \frac{1}{(12)\sqrt{2\pi}} e^{-\frac{1}{2(144)}\left(y-\left[60+(0.9)(x-70)\right]\right)^2}$$

Now, since the **given value** of X is x**=75**, therefore we have

$$f_{Y|X}(y \mid X = 75) = \frac{1}{(12)\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2(12)^2}\left(y-\left[60+(0.9)(75-70)\right]\right)^2}$$

$$= \frac{1}{(12)\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2(12)^2}\left(y-\left[60+(0.9)(5)\right]\right)^2} = \frac{1}{(12)\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2(12)^2}(y-64.5)^2}$$

which is the pdf of the **univariate normal distribution** with mean and standard deviation

$\tilde{\mu}_y = 64.5$ $\tilde{\sigma}_y = 12.$

Hence, the probability that a randomly selected student scores **over 85** on Exam 2 (Final Exam) **given that** he/she scored 75 in Exam 1 (Midterm Exam) is given by

$$P(Y > 85) = P\left(Z > \frac{y - \tilde{\mu}_y}{\tilde{\sigma}_y}\right) = P\left(Z > \frac{85 - 64.5}{12}\right) = P\left(Z > \frac{20.5}{12}\right)$$

$$= P(Z > 1.71) = 1 - P(Z \leq 1.71) = 1 - \Phi(1.71)$$

$$= 1 - 0.9564 = 0.0436 = 4.36\%$$

--- a little less than 5%.

vulearninghub.xyz | by VU Learning HUB

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Topic No. 183 is at the end.

Topic No. 184

Marginals of the Bivariate Normal Distribution are themselves Normal

We know that the **joint p.d.f.** of Bivariate Normal distribution (X, Y) is

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]}$$

where $-\infty < x, y < \infty$ and the parameters are such that

$-\infty < \mu_x, \mu_y < \infty;$

$\sigma_x, \sigma_y > 0;$

$-1 < \rho < 1.$

Let us now **derive** the **Marginal distribution** of X:

For any bivariate distribution $f(x,y)$, the **marginal distribution of** $\boldsymbol{X}$ is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,dy, \quad -\infty < x < \infty.$$

Hence, in the case of the **bivariate normal**, the **marginal** distribution of X will be given by

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]} dy$$

Let us now do some algebraic manipulation in the expression of the exponent:

The exponent is given by:

$$-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]$$

Dividing and multiplying the above expression by $\sigma_x^2\sigma_y^2$, we have

$$=-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\frac{\sigma_x^2\sigma_y^2(x-\mu_x)^2}{\sigma_x^2}+\frac{\sigma_x^2\sigma_y^2(y-\mu_y)^2}{\sigma_y^2}-\frac{2\sigma_x^2\sigma_y^2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]$$

$$=-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\sigma_y^2(x-\mu_x)^2+\sigma_x^2(y-\mu_y)^2-2\rho\sigma_x\sigma_y(x-\mu_x)(y-\mu_y)\right]$$

Completing the square in the exponent by Adding and Subtracting $\left(x-\mu_x\right)^2\sigma_y^2\rho^2$ in the square brackets:

$$-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\times$$

$$\left[\sigma_y^2(x-\mu_x)^2+\sigma_x^2(y-\mu_y)^2-2\rho\sigma_x\sigma_y(x-\mu_x)\left(y-\mu_y\right)+\left(x-\mu_x\right)^2\sigma_y^2\rho^2-\left(x-\mu_x\right)^2\sigma_y^2\rho^2\right]$$

Combining (i) the second, third and fourth terms, and (ii) the first and fifth terms, we have

$$-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\times$$

$$\left[\left[\left(x-\mu_x\right)^2\sigma_y^2\rho^2-2\rho\sigma_x\sigma_y(x-\mu_x)\left(y-\mu_y\right)+\sigma_x^2(y-\mu_y)^2\right]+\left[\sigma_y^2(x-\mu_x)^2-\left(x-\mu_x\right)^2\sigma_y^2\rho^2\right]\right]$$

$$=-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\left[\sigma_x(y-\mu_y)-\left(x-\mu_x\right)\sigma_y\rho\right]^2+\sigma_y^2(x-\mu_x)^2\left[1-\rho^2\right]\right]$$

which can be re-written as

$$-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\sigma_x(y-\mu_y)-(x-\mu_x)\sigma_y\rho\right]^2-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\sigma_y^2(x-\mu_x)^2\left[1-\rho^2\right]$$

$$=-\frac{1}{2(1-\rho^2)}\left(\frac{\sigma_x(y-\mu_y)}{\sigma_x\sigma_y}-\frac{(x-\mu_x)\sigma_y\rho}{\sigma_x\sigma_y}\right)^2-\frac{(x-\mu_x)^2}{2\sigma_x^2}$$

$$=-\frac{1}{2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y}-\frac{(x-\mu_x)\rho}{\sigma_x}\right)^2-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2$$

Coming back to the overall expression of the marginal pdf:

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \mathrm{e}^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]} dy$$

Substituting the newly obtained expression of the exponent in the above pdf, we have

$$f_X(x) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y}-\frac{(x-\mu_x)\rho}{\sigma_x}\right)^2-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} dy$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y}-\frac{(x-\mu_x)\rho}{\sigma_x}\right)^2} \times e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} dy$$

Now, let us shift our attention to the first part of the expression inside the integral sign (the one that does NOT involve the exponent):

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} = \frac{1}{\sqrt{2\pi}\sigma_x\sqrt{2\pi}\left(\sigma_y\sqrt{1-\rho^2}\right)}$$

$$\Rightarrow f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\left(\sigma_y\sqrt{1-\rho^2}\right)} e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y} - \frac{(x-\mu_x)\rho}{\sigma_x}\right)^2} dy \quad ...(1)$$

vulearninghub.xyz | by VU Learning HUB

Multiplying and dividing the exponent inside the integral sign by σ_y^2, we have

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{\sigma_y^2}{2\sigma_y^2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y} - \frac{(x-\mu_x)\rho}{\sigma_x}\right)^2} dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(\frac{(y-\mu_y)\sigma_y}{\sigma_y} - \frac{(x-\mu_x)\rho\sigma_y}{\sigma_x}\right)^2} dy$$

Now, let $\tilde{\sigma}_y^2 = \sigma_y^2\left(1-\rho^2\right) \Rightarrow \tilde{\sigma}_y = \sigma_y\sqrt{1-\rho^2}$,

therefore

$$f_X\left(x\right) = \frac{1}{\sqrt{2\pi}\sigma_x}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\tilde{\sigma}_y}\mathrm{e}^{-\frac{1}{2\tilde{\sigma}_y^2}\left(\frac{(y-\mu_y)\sigma_y}{\sigma_y}-\frac{(x-\mu_x)\rho\sigma_y}{\sigma_x}\right)^2}dy$$

$$= \frac{1}{\sqrt{2\pi}\sigma_x}\mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\tilde{\sigma}_y}\mathrm{e}^{-\frac{1}{2\tilde{\sigma}_y^2}\left(\left(y-\mu_y\right)-\left(x-\mu_x\right)\frac{\sigma_y}{\sigma_x}\rho\right)^2}dy$$

vulearninghub.xyz | by VU Learning HUB

Now

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_y} e^{-\frac{1}{2\tilde{\sigma}_y^2}\left(y-\mu_y-(x-\mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho\right)^2} dy$$

can be re-written as

$$= \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_y} e^{-\frac{1}{2\tilde{\sigma}_y^2}\left(y-\left[\mu_y+(x-\mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho\right]\right)^2} dy$$

vulearninghub.xyz | by VU Learning HUB

Letting $\tilde{\mu}_y = \mu_y + (x - \mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho,$

we have $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_y} e^{-\frac{1}{2\tilde{\sigma}_y^2}\left(y-\tilde{\mu}_y\right)^2} dy$

Hence, the integrand is the pdf of a normal density of the variable Y with:

$$E(Y) = \tilde{\mu}_y = \mu_y + (x - \mu_x)\left(\frac{\sigma_y}{\sigma_x}\right)\rho, \text{Var}(Y) = \tilde{\sigma}_y^2 = (1 - \rho^2)\sigma_y^2$$

Now, it is obvious that, for any density function, the total area under the curve is unity.

Hence, $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\tilde{\sigma}_y} e^{-\frac{1}{2\tilde{\sigma}_y^2}\left(y-\tilde{\mu}_y\right)^2} dy = 1$

$$\Rightarrow f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}} (1) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{(x-\mu_x)^2}{2\sigma_x^2}}$$

Therefore, the marginal distribution of X is

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_1}\right)^2}, \quad -\infty < x < \infty$$

i.e. the normal distribution with mean μ_x and variance σ_x^2.

i.e.

$$X \sim N\left(\mu_x, \sigma_x^2\right) \text{ with } \mathrm{E}\left(X\right) = \mu_x,\ \mathrm{Var}\left(X\right) = \sigma_x^2,$$

Similarly,

$$Y \sim N\left(\mu_y, \sigma_y^2\right) \text{ with } \mathrm{E}\left(Y\right) = \mu_y,\ \mathrm{Var}\left(Y\right) = \sigma_y^2.$$

Important Note:

It is possible to have a joint p.d.f. which has marginal p.d.f.s which are Normal, yet which is NOT bivariate normal.

Virtual University of Pakistan
Probability Distributions

by

Dr. Saleha Naghmi Habibullah

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

Topic No.186

Conditional distributions in the case of Bivariate Normal Distribution

the Conditional Distributions
related to
the Bivariate Normal Distribution
and,
through
DERIVATION,
I will show that the Conditional Distributions
related to the bivariate normal distribution
are themselves Normal.

So, let us start with the pdf of the bivariate normal distribution:

We know that the **joint p.d.f.** of Bivariate Normal distribution (X, Y) is

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}$$

where $-\infty < x, y < \infty$ and the parameters are such that

$-\infty < \mu_x, \mu_y < \infty;$

$\sigma_x, \sigma_y > 0;$

$-1 < \rho < 1.$

Let us now attempt to derive the pdf of the conditional distributions associated with the bivariate normal distribution:

In general, we know that, in the case of **any** bivariate distribution $f(x,y)$, the **conditional distribution** of Y given $X = x$ is expressed as follows:

$$f_{Y|X}(y \mid X = x) = \frac{f_{XY}(x, y)}{f_X(x)}$$

where $f_{XY}(x, y)$ is the joint pdf of X and Y and $f_X(x)$ is the marginal pdf of X.

vulearninghub.xyz | by VU Learning HUB

Now, we know that, in the case of bivariate normal distribution, the **marginal** distribution of X is given by:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma_x} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}, \quad -\infty < x < \infty$$

i.e. the normal distribution with mean μ_x and variance σ_x^2.

Therefore, in the case of the **bivariate normal** distribution, we will have:

$$f_{Y|X}(y \mid X = x) = \frac{f_{XY}(x,\ y)}{f_X(x)} = \frac{\left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\right) e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}}{\left(\frac{1}{\sqrt{2\pi}\sigma_x}\right) e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}$$

Let us now do some algebraic manipulation in the expression of the exponent in the numerator:

The exponent is given by:

$$-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_x)^2}{\sigma_x^2}+\frac{(y-\mu_y)^2}{\sigma_y^2}-\frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]$$

Dividing and multiplying the above expression by $\sigma_x^2\sigma_y^2$, we have

$$=-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\frac{\sigma_x^2\sigma_y^2(x-\mu_x)^2}{\sigma_x^2}+\frac{\sigma_x^2\sigma_y^2(y-\mu_y)^2}{\sigma_y^2}-\frac{2\sigma_x^2\sigma_y^2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}\right]$$

$$=\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\sigma^2(x-\mu_x)^2+\sigma^2(y-\mu_y)^2-2\rho\sigma_x\sigma_y(x-\mu_x)(y-\mu_y)\right]$$

yulearninghub.xyz | by VU learning HUB

Completing the square in the exponent by Adding and Subtracting $\left(x-\mu_x\right)^2\sigma_y^2\rho^2$ in the square brackets:

$$-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\times$$

$$\left[\sigma_y^2(x-\mu_x)^2+\sigma_x^2(y-\mu_y)^2-2\rho\sigma_x\sigma_y(x-\mu_x)\left(y-\mu_y\right)+\left(x-\mu_x\right)^2\sigma_y^2\rho^2-\left(x-\mu_x\right)^2\sigma_y^2\rho^2\right]$$

Combining (i) the second, third and fourth terms, and (ii) the first and fifth terms, we have

$$-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\times$$

$$\left[\left[\left(x-\mu_x\right)^2\sigma_y^2\rho^2-2\rho\sigma_x\sigma_y(x-\mu_x)\left(y-\mu_y\right)+\sigma_x^2(y-\mu_y)^2\right]+\left[\sigma_y^2(x-\mu_x)^2-\left(x-\mu_x\right)^2\sigma_y^2\rho^2\right]\right]$$

$$=-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\left[\sigma_x(y-\mu_y)-(x-\mu_x)\sigma_y\rho\right]^2+\sigma_y^2(x-\mu_x)^2\left[1-\rho^2\right]\right]$$

vulexamhub.xyz | by VU Learning HUB

which can be re-written as

$$-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\left[\sigma_x(y-\mu_y)-(x-\mu_x)\sigma_y\rho\right]^2-\frac{1}{2(1-\rho^2)\sigma_x^2\sigma_y^2}\sigma_y^2(x-\mu_x)^2\left[1-\rho^2\right]$$

$$=-\frac{1}{2(1-\rho^2)}\left(\frac{\sigma_x(y-\mu_y)}{\sigma_x\sigma_y}-\frac{(x-\mu_x)\sigma_y\rho}{\sigma_x\sigma_y}\right)^2-\frac{(x-\mu_x)^2}{2\sigma_x^2}$$

$$=-\frac{1}{2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y}-\frac{(x-\mu_x)\rho}{\sigma_x}\right)^2-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2$$

Multiplying and dividing the first part of the exponent by σ_y^2, we have

$$-\frac{\sigma_y^2}{2\sigma_y^2(1-\rho^2)}\left(\frac{(y-\mu_y)}{\sigma_y}-\frac{(x-\mu_x)\rho}{\sigma_x}\right)^2-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2$$

$$=-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(\frac{(y-\mu_y)\sigma_y}{\sigma_y}-\frac{(x-\mu_x)\rho\sigma_y}{\sigma_x}\right)^2-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2$$

Now, let us revert back to the expression of the **conditional** distribution:

$$f_{Y|X}(y \mid X = x)$$

$$= \frac{f_{XY}(x,\, y)}{f_X(x)} = \frac{\left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\right) e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]}}{\left(\frac{1}{\sqrt{2\pi}\sigma_x}\right) e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}$$

This can be re-written as:

$$f_{Y|X}(y \mid X = x) = \frac{\left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\right) e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left((y-\mu_y)-(x-\mu_x)\frac{\sigma_y}{\sigma_x}\rho\right)^2 - \frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}{\left(\frac{1}{\sqrt{2\pi}\sigma_x}\right) e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}$$

$$= \frac{\left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\right) e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left((y-\mu_y)-(x-\mu_x)\frac{\sigma_y}{\sigma_x}\rho\right)^2} \times e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}{\left(\frac{1}{\sqrt{2\pi}\sigma_x}\right) e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}}$$

Cancelling out the term $e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}$ in the numerator with the same term in the denominator, we have

$$f_{Y|X}(y \mid X = x) = \frac{\left(\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\right)}{\left(\frac{1}{\sqrt{2\pi}\sigma_x}\right)} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left((y-\mu_y)-(x-\mu_x)\frac{\sigma_y}{\sigma_x}\rho\right)^2}$$

$$= \frac{\sqrt{2\pi}\sigma_x}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\left[\mu_y+(x-\mu_x)\frac{\sigma_y}{\sigma_x}\rho\right]\right)^2}$$

Now,

$$f_{Y|X}(y \mid X = x)$$

$$= \frac{\sqrt{2\pi\sigma_x}}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\left[\mu_y+(x-\mu_x)\frac{\sigma_y}{\sigma_x}\rho\right]\right)^2}$$

$$= \frac{1}{\sqrt{2\pi\sigma_y}\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_y^2(1-\rho^2)}\left(y-\left[\mu_y+(x-\mu_x)\frac{\sigma_y}{\sigma_x}\rho\right]\right)^2} \quad \ldots(1)$$

Now, let $\tilde{\sigma}_y^2 = \sigma_y^2\left(1-\rho^2\right) \Rightarrow \tilde{\sigma}_y = \sigma_y\sqrt{1-\rho^2}$

and let $\tilde{\mu}_y = \mu_y + \left(x-\mu_x\right)\left(\frac{\sigma_y}{\sigma_x}\right)\rho,$

Then, equation (1) can be re-written as

$$f_{Y|X}(y \mid X=x) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_y} e^{-\frac{1}{2\tilde{\sigma}_y^2}\left(y-\tilde{\mu}_y\right)^2}$$

vulearninghub.xyz | by VU Learning HUB

Now, $f_{Y|X}(y|X=x) = \frac{1}{\sqrt{2\pi}\tilde{\sigma}_y} e^{-\frac{1}{2\tilde{\sigma}_y^2}\left(y-\tilde{\mu}_y\right)^2}$

is the pdf of a normal distribution of the variable Y with:

$$E(Y) = \tilde{\mu}_y = \mu_y + \left(x-\mu_x\right)\left(\frac{\sigma_y}{\sigma_x}\right)\rho,$$

$$\mathrm{Var}(Y) = \tilde{\sigma}_y^2 = \left(1-\rho^2\right)\sigma_y^2.$$

The conditional distribution of Y given x can be expressed as follows:

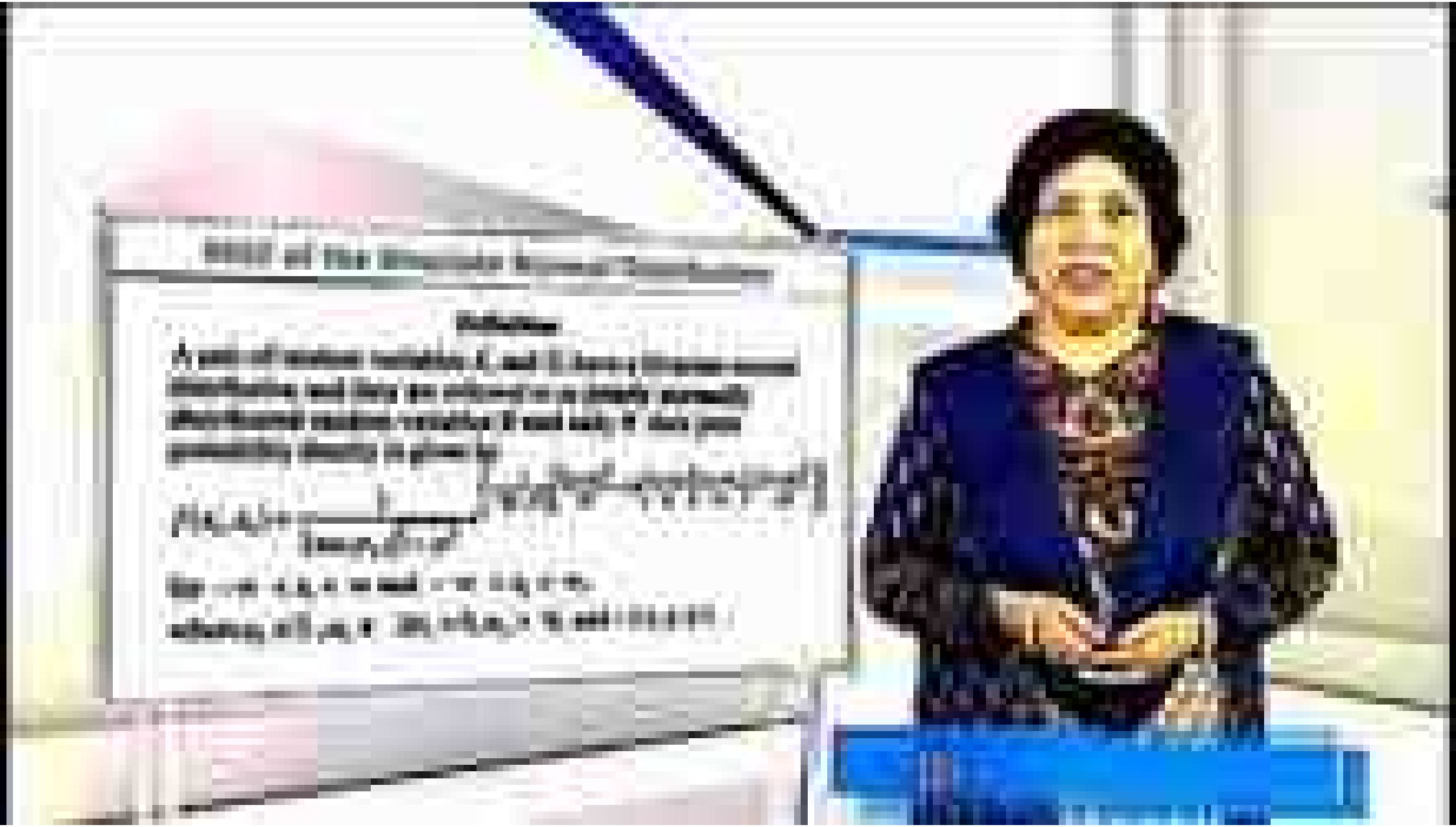
$$N\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}\left(x-\mu_x\right),\ \sigma_y^2\left(1-\rho^2\right)\right) = N\left(\tilde{\mu}_y,\ \tilde{\sigma}_y^2\right)$$

تمام دوست و احباب دعاؤں میں یاد رکھیے گا۔

طالب دعا مہر آفاق صفدر محمدی

Wa.me/+923494873115

Scanned with CamScanner



Scanned with CamScanner

MGF of the Bivariate Normal Distribution

NOW

Let us consider the Moment Generating Function of the Bivariate Normal distribution.

Scanned with CamScanner

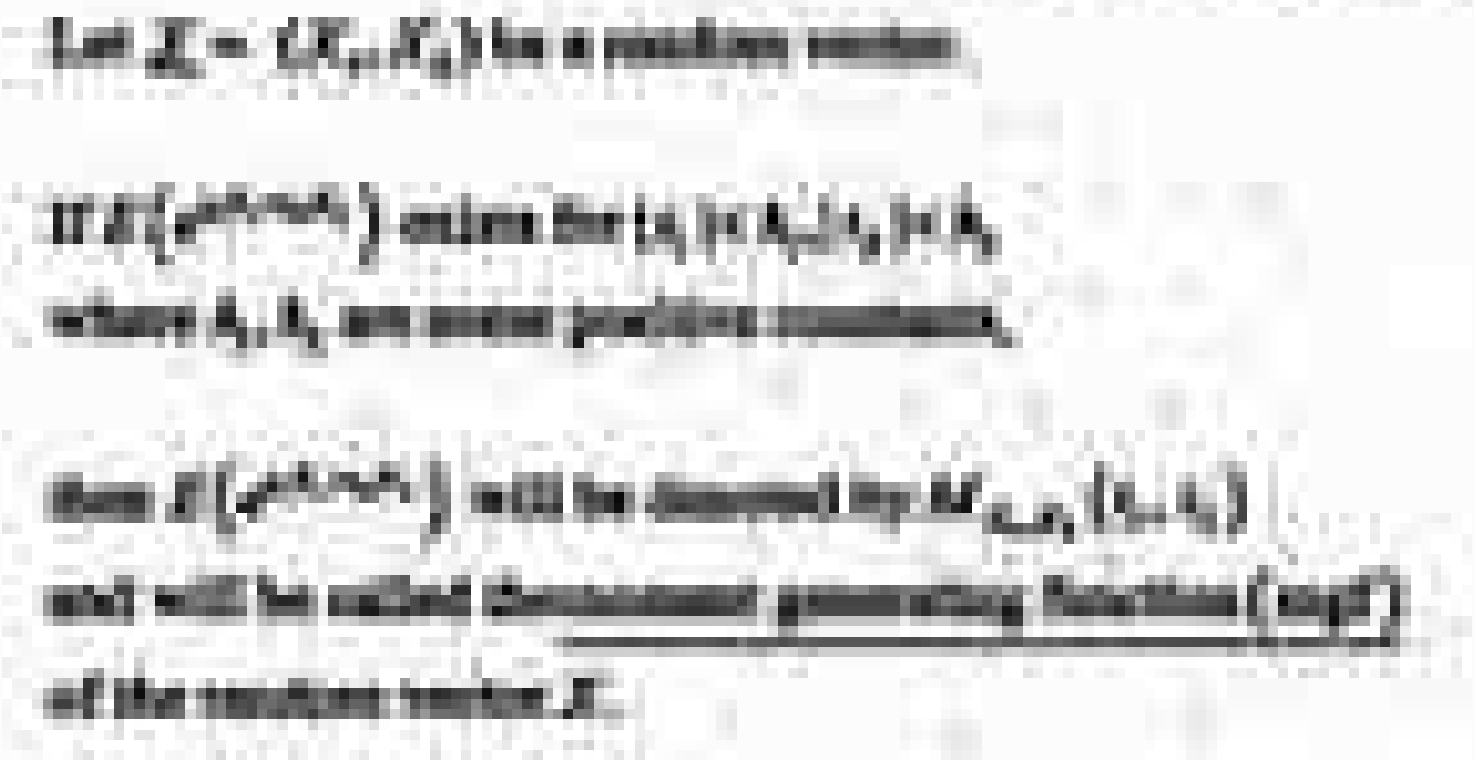
MGF of the Bivariate Normal Distribution

For this purpose, we begin with the definition of the mgf of a bivariate distribution (in general):

Scanned with CamScanner

MGF of the Bivariate Normal Distribution

Definition of Moment Generating Function of a Random Vector having 2 components:



Scanned with CamScanner

MGF of the Bivariate Normal Distribution

As far as the moment generating function of the bivariate normal distribution is concerned, it can be shown that:

$$M_{\underline{X}}(t) = M_{\underline{X}}(t_1, t_2) = E\left(e^{\sum_{i=1}^{2} t_i X_i}\right)$$

$$= e^{\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)}, \quad -\infty < t_1 < \infty, \; -\infty < t_2 < \infty$$

Scanned with CamScanner

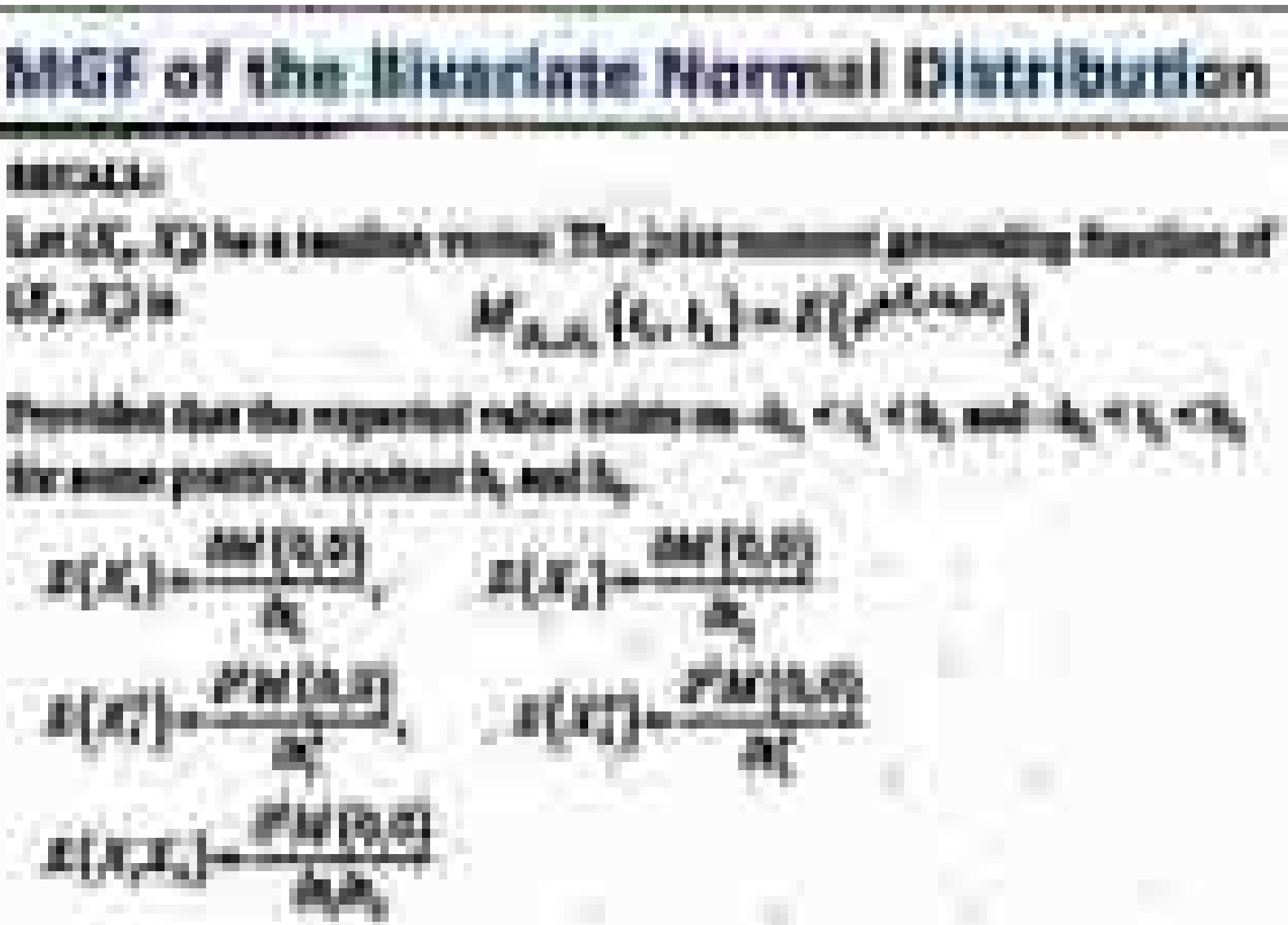
MGF of the Bivariate Normal Distribution

For this, RECALL:

The general method of obtaining Product Moments and simple Moments from the MGF of a random vector

Scanned with CamScanner

MGF of the Bivariate Normal Distribution



Scanned with CamScanner

MGF of the Bivariate Normal Distribution

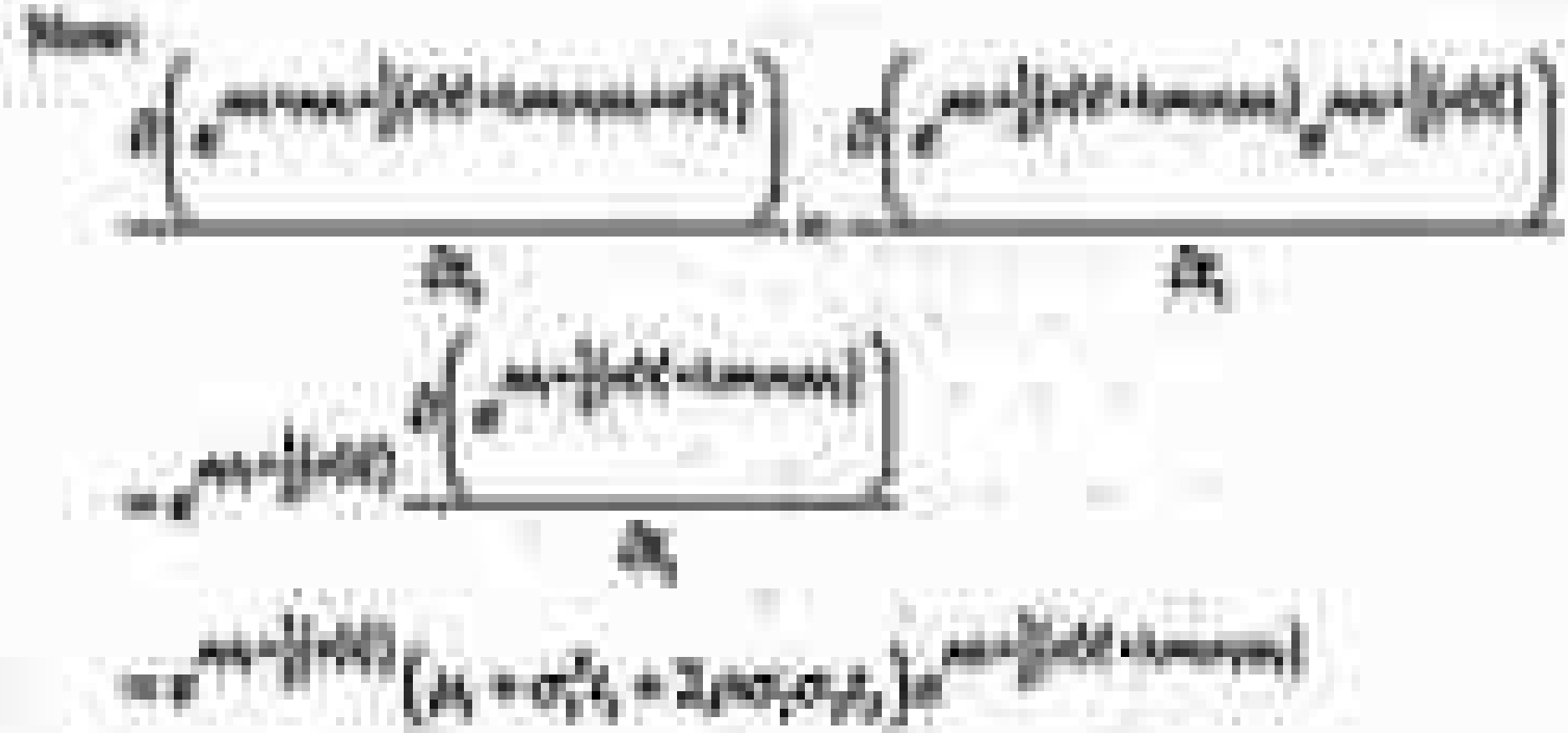
Hence: $M(t_1,t_2) = e^{\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)}$

Therefore

$$E(X_1) = \frac{\partial M(0,0)}{\partial t_1} = e^{\left\{\mu_1 t_1 + \mu_2 t_2 + \frac{1}{2}(\sigma_1^2 t_1^2 + 2\rho\sigma_1\sigma_2 t_1 t_2 + \sigma_2^2 t_2^2)\right\}} \left\{ \qquad \right\}$$

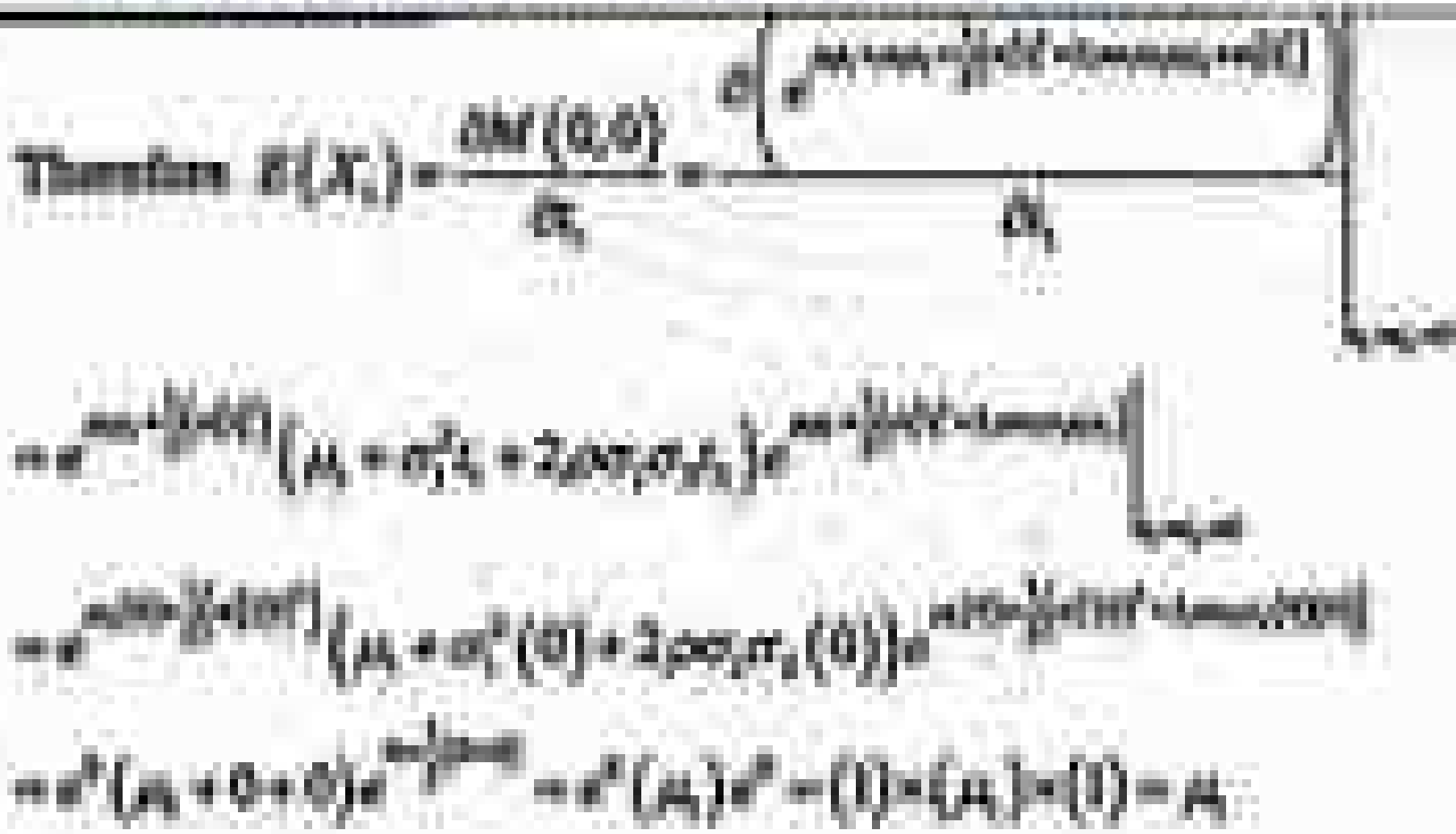
Scanned with CamScanner

MGF of the Bivariate Normal Distribution



Scanned with CamScanner

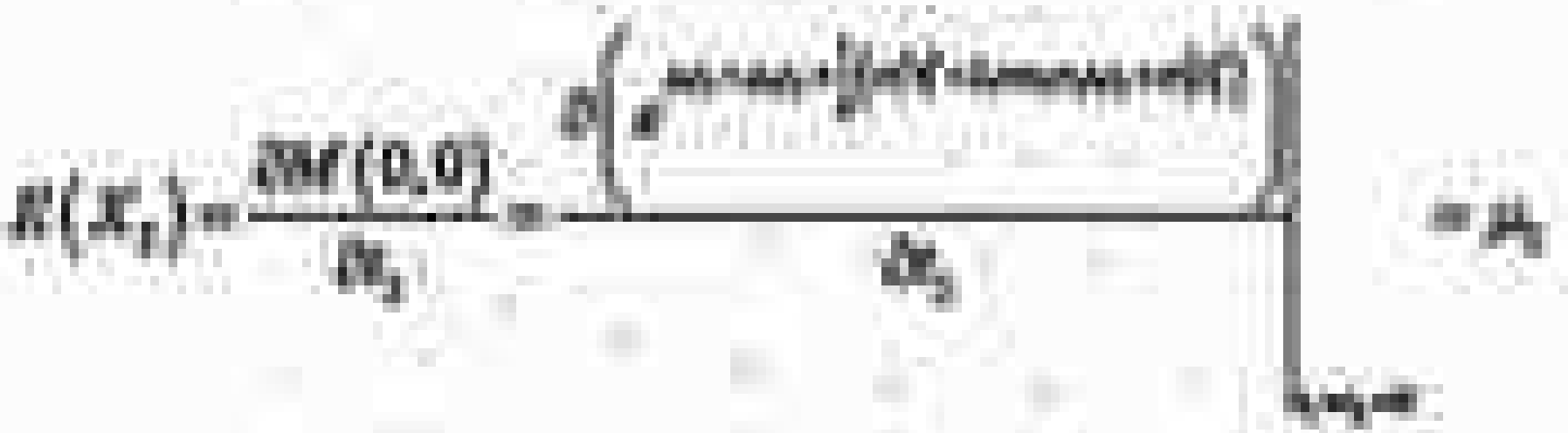
MGF of the Bivariate Normal Distribution



Scanned with CamScanner

MGF of the Bivariate Normal Distribution

Similarly, it can be proved that



Scanned with CamScanner

MGF of the Bivariate Normal Distribution

Second moments about zero i.e. $E(X_1)^2$ & $E(X_2)^2$ can be found through the formulas

$$E(X_1^2) = \frac{\partial^2 M(0,0)}{\partial t_1^2}, \qquad E(X_2^2) = \frac{\partial^2 M(0,0)}{\partial t_2^2}$$

And, once they are obtained, the variances can be found very quickly.

Scanned with CamScanner

MGF of the Bivariate Normal Distribution

The product moment $E(X_1X_2)$ can be found through the formula

$$E(X_1X_2) = \frac{\partial^2 M(0,0)}{\partial t_1 \partial t_2}$$

And, once this is obtained, the covariance and the correlation coefficient can be found easily.

Scanned with CamScanner