STA642

Edited by
Mahar Afaq Safdar Muhammadi
MSc Math, Whatsapp 03494873115

دعاؤں میں یاد رکھیے گا۔

Scanned with CamScanner

00:25

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 $y^2/12$, $0 < y < 1$.

00:47

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 $0 < x < 1,\ x < y < 1$

02:02

First and foremost, we need to find each of the two **marginal** distributions:

02:20

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. $f_X(x)$)

Scanned with CamScanner

03:37

As such:

Marginal distribution of X:

$$f_X(x) = \int_x^1 2\,dy = 2.\,y\Big|_x^1 = 2(1-x),\quad 0<x<1$$

and

Marginal distribution of Y:

$$f_Y(y) = \int_0^y 2\,dx = 2.\,x\Big|_0^y = 2(y-0) = 2y,\quad 0<y<1$$

04:38

Then, the conditional distributions are as follows:

$$f(x|y) = \frac{f(x,y)}{f_Y(y)} = \frac{2}{2y} = \frac{1}{y},\quad 0<x<y$$

and

$$f(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2}{2(1-x)} = \frac{1}{1-x},\quad x<y<1$$

(Note the domains (i.e. supports) of the two distributions.)

06:21

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

$$E(X|Y=y) = \int_0^y x.f(x|y)\,dx = \int_0^y x.\frac{1}{y}dx = \frac{1}{y}\left(\frac{x^2}{2}\Big|_0^y\right) = \frac{1}{y}\left(\frac{y^2}{2}-0\right)$$

$$= \frac{1}{y}.\frac{y^2}{2} = \frac{y}{2},\quad 0<y<1$$

10:13

Similarly

$$E(X^2|y) = \int_0^y x^2.f(x|y)\,dx = \int_0^y x^2.\frac{1}{y}dx = \frac{1}{y}\left(\frac{x^3}{3}\Big|_0^y\right)$$

$$= \frac{1}{y}\left(\frac{y^3}{3}-0\right) = \frac{1}{y}.\frac{y^3}{3} = \frac{y^2}{3},\quad 0<y<1$$

Scanned with CamScanner

10:36

Therefore,

$$\text{Var}\ X \mid y = E\ X^2 \mid y - \left[E\ X \mid y\right]^2 = \frac{y^2}{3} - \left(\frac{y}{2}\right)^2$$

$$= \frac{y^2}{3} - \frac{y^2}{4} = \frac{4y^2 - 3y^2}{12} = \frac{y^2}{12}, \quad 0 < y < 1.$$

Scanned with CamScanner

03:09

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

06:38

Example 2:

Let X denote the number that we observe on the uppermost face if we roll a fair die once.

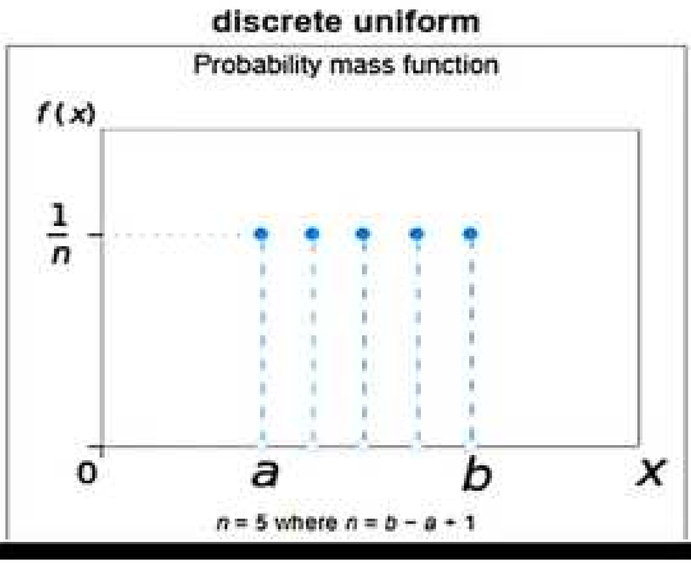
Then, we have

07:10

X	$p(x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

08:39

SHAPE OF THE DISTRIBUTION



Scanned with CamScanner

00:22

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.

00:53

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

01:41

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

02:54

Step 2: Write as follows:

$$p(x) = \begin{cases} 0 & -\infty < x < 1 \\ 1/6 & 1 \le x < 2 \\ 2/6 & 2 \le x < 3 \\ 3/6 & 3 \le x < 4 \\ 4/6 & 4 \le x < 5 \\ 5/6 & 5 \le x < 6 \\ 6/6 & 6 \le x < \infty \end{cases}$$

The entire X-axis (from $-\infty$ to ∞) must be covered.

Scanned with CamScanner

00:19

Derivation of the Mean of the Discrete Uniform Distribution

We derive the formula for the mean of the discrete random variable $X \sim U(1,n)$:

$$\mu = \sum_{x=1}^{n} x p_x \quad x \quad = \sum_{x=1}^{n} x . \frac{1}{n}$$

$$= \frac{1}{n} \; 1+2+\ldots+n \quad = \frac{1}{n} \times \frac{n(n+1)}{2}$$

or

$$\mu = \frac{n+1}{2}$$

Scanned with CamScanner

07:04

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

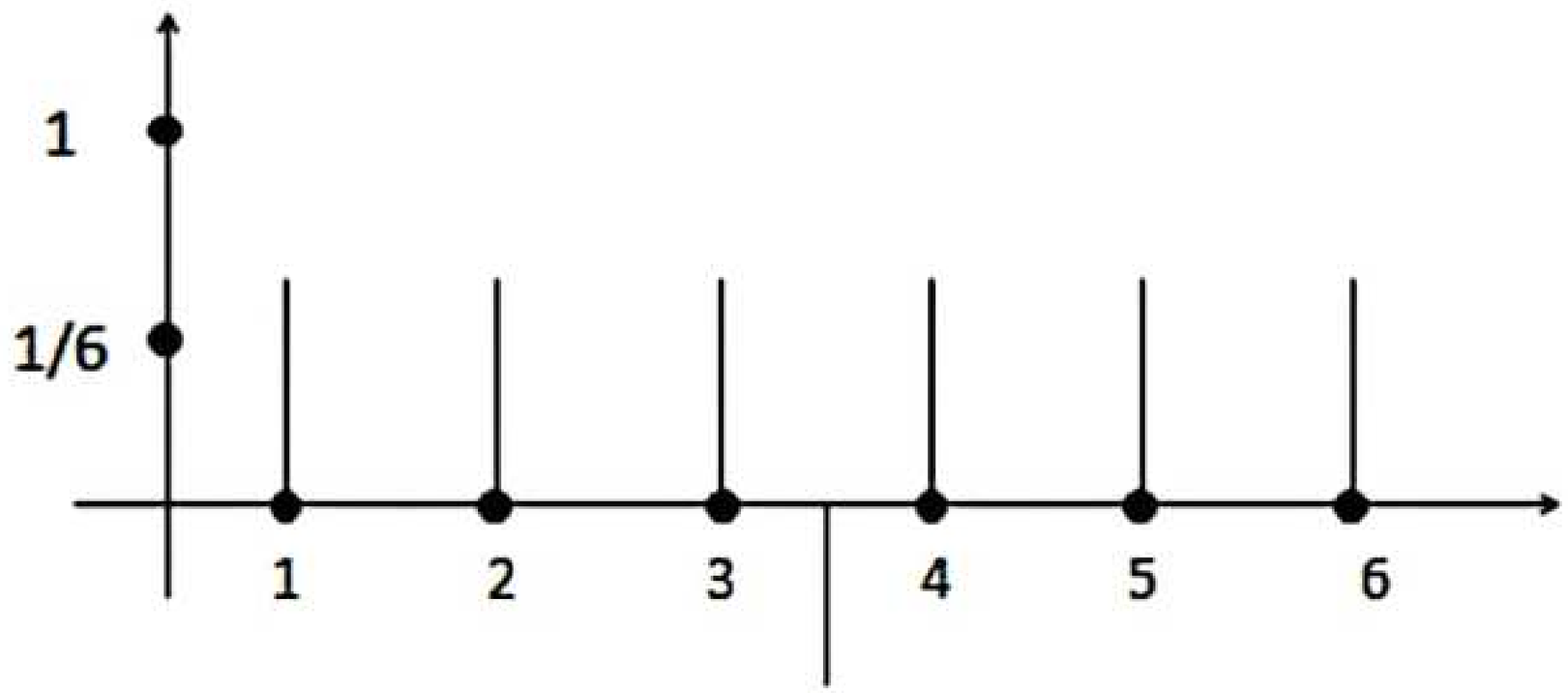
08:04

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

Scanned with CamScanner

00:36

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.

00:24

x	$P(X=x)=\binom{n}{x}p^x q^{n-x}$
0	$\binom{5}{x}\left(\frac{3}{8}\right)^x\left(\frac{5}{8}\right)^{5-x}$
1	$\binom{5}{x}\left(\frac{3}{8}\right)^x\left(\frac{5}{8}\right)^{5-x}$
2	$\binom{5}{x}\left(\frac{3}{8}\right)^x\left(\frac{5}{8}\right)^{5-x}$
3	$\binom{5}{x}\left(\frac{3}{8}\right)^x\left(\frac{5}{8}\right)^{5-x}$
4	$\binom{5}{x}\left(\frac{3}{8}\right)^x\left(\frac{5}{8}\right)^{5-x}$
5	$\binom{5}{x}\left(\frac{3}{8}\right)^x\left(\frac{5}{8}\right)^{5-x}$

06:03

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

The sum of all the probabilities equals to 1.

Scanned with CamScanner

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

$P(Y \geq 1) = 1 - P(Y<1) = 1 - P(Y=0) = 1 - (2/3)^n$.

04:32

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 $P(Y = 0) = (1 - 1/3)^n = (2/3)^n$

06:41

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 **$1-(2/3)^n > 0.80$** implying that $0.20 > (2/3)^n$.

07:20

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 $$n > 1.6094/0.4055 \quad \text{i.e. } n > 3.97.$$

Thus, we see that $n = 4$ is the solution.

Scanned with CamScanner

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?

(iii) What is the probability that, out of the nine seeds, at least 8 will germinate ?

02:24

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.

05:14

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.

Scanned with CamScanner

08:18

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.

09:30 What is the probability that, out of the nine seeds, at least 8 will germinate ?

We have

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{9}{x} 0.80^x (0.20)^{9-x}$$

$$So\ P(X = 8\ or\ X = 9) = \binom{9}{8} 0.80^8 (0.20)^{9-8} + \binom{9}{9} 0.80^9 (0.20)^{9-9}$$

$$= 9 \times 0.80^8 (0.20) + 1 \times 0.80^9 \times 1 = 9 \times 0.1678 \times 0.20 + 0.1342$$

$$= 0.3020 + 0.1342 = 0.4362 = 43.62\%$$

$$(a\ non-trivial\ amount\ of\ probability)$$

Scanned with CamScanner

01:45

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 $p = 1/3$;

06:07

that is, the **pmf** of X is

$$p\ x\ = \begin{cases} \binom{5}{x}\left(\frac{1}{3}\right)^{x}\left(\frac{2}{3}\right)^{5-x} & x = 0,1,2,\ldots,5 \\ 0 & elsewhere. \end{cases}$$

$$\mu = np = \frac{5}{3} \quad and \quad \sigma^{2} = np\ 1-p\ = \frac{10}{9}.$$

Scanned with CamScanner

02:41

Example:

Suppose that a game consists of tossing a fair coin until we obtain the 3rd head.

Then

head is 'success',

$p=P(H)=0.5$

&

$r=3$

03:53

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

$$P\ X=x\ =\binom{x-1}{2}\ 0.5^{\ x},\ x=3,4,5,\ldots$$

05:44

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.	

Scanned with CamScanner

00:22

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**?

01:33

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:

$$P(X=x)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r},\quad x=r,r+1,r+2,\ldots$$

03:17

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 $P(X=x)=\binom{x-1}{r-1}p^{r}(1-p)^{x-r},\quad x=r,r+1,r+2,\ldots$

putting the values

$$P(X=x)=\binom{x-1}{2-1}0.2^{2}\,0.8^{x-2},\quad x=2,3,4,\ldots$$

$$\text{or } P(X=x)=\binom{x-1}{1}0.2^{2}\,0.8^{x-2},\quad x=2,3,4,\ldots$$

$$\text{or } P(X=x)=(x-1)\,0.2^{2}\,0.8^{x-2},\quad x=2,3,4,\ldots$$

06:40

To find the required probability, we need to find $P(X=5)$.

$$P(X=x)=(x-1)\,0.2^{2}\,0.8^{x-2}$$

$$\Rightarrow P(X=5)=(5-1)\,0.2^{2}\,0.8^{5-2}$$

$$=4\,(0.2)^{2}\,(0.8)^{3}$$

$$=0.082$$

$$=8.2\%$$

Scanned with CamScanner

01:08

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.

02:04

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:

$$p_X\ x\ = 1-p\ ^{x-1}\ p,\ x=1,2,3,\ldots \quad (1)$$

04:33

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

$$p_X\ x\ = 0.75\ ^{x-1}\ 0.25\ ,\ x=1,2,3,\ldots$$

05:40

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

$$P\ X=6\ = 0.75\ ^{6-1}\ 0.25\ = 0.75\ ^{5}\ 0.25$$

$$=0.2373\times 0.25=0.059=5.9\%$$

Scanned with CamScanner

00:24

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?

05:52

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

$$\mu = E\ X\ = \frac{1}{0.20} = 5$$

09:08

For the geometric random variable X, the average number is given by:

$$\sigma^2 = Var\ X\ = \frac{1-p}{p^2}$$

so here

$$\sigma^2 = Var\ X\ = \frac{1-p}{p^2} = \frac{1-0.20}{0.20^2} = \frac{0.80}{0.04} = 20$$

Scanned with CamScanner

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.

01:22

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.

02:09

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?

02:40

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.

Scanned with CamScanner

03:26

Hence, the answer to the first question is:

$$P\ \ X_1 = 7, X_2 = 3, X_3 = 2\ \ = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}$$

$$= \frac{12!}{7!3!2!}\ 0.65^{\ 7}\ \ 0.20^{\ 3}\ \ 0.15^{\ 2} = 0.0699 = 6.99\% \text{ or } 7\%.$$

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%

09:00

The resulting distribution is none other than the **Binomial** distribution.

Hence, the answer to the second question is:

$$P\ \ X_1 = 4, X_2 = 8\ \ = \frac{n!}{x_1!x_2!} p_1^{x_1} p_2^{x_2}$$

$$= \frac{12!}{4!8!}\ 0.65^{\ 4}\ \ 0.35^{\ 8} = 0.0199 = 1.99\% \text{ or } 2\%.$$

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%

Scanned with CamScanner

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 $X = 1$.

11:12

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:

$$P(X = 0 \text{ or } X = 1) = P(X = 0) + P(X = 1)$$

$$= 0.069 + 0.206 = 0.275 = 27.5\%.$$

Scanned with CamScanner

00:22

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?

05:00

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)

06:24 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

$$h\ 2;10,4,5 = \frac{\binom{5}{2}\binom{5}{4-2}}{\binom{10}{4}} = \frac{\binom{5}{2}\binom{5}{2}}{\binom{10}{4}} = \frac{10}{21}.$$

Scanned with CamScanner

00:15

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 ?

00:45

Solution:

This is a Poisson process with Occurrence Rate = $\lambda = 2$ per day.

Also, one week = 7 days implies t=7

Hence $\lambda t = 2 \times 7 = 14$

So that

$$P(X = x) = \begin{cases} \dfrac{e^{-\lambda t}(\lambda t)^x}{x!} = \dfrac{e^{-14}(14)^x}{x!} & x = 0,1,2,\ldots \\ 0 & elsewhere, \end{cases} \qquad (1)$$

04:44

Now, we require X<2.

So

$$P(X \le 2) = P(X = 0 \text{ or } X = 1 \text{ or } X = 2)$$

$$= \frac{e^{-14}\,14^{0}}{0!} + \frac{e^{-14}\,14^{1}}{1!} + \frac{e^{-14}\,14^{2}}{2!}$$

$$= e^{-14} + 14e^{-14} + 98e^{-14} = 113e^{-14}$$

$$= 113 / 1202592.9591 = 0.00009396$$

Scanned with CamScanner

04:19

Example 1:

Suppose that X has a Poisson distribution with $\lambda = 4$.

Then the pmf of X is

$$p(x) = \begin{cases} \dfrac{0.5^x e^{-0.5}}{x!} & x = 0,1,2,\ldots \\ 0 & elsewhere. \end{cases}$$

« « « Mistake

0.5 ki jaga 4 ata hy.

05:00

The mean of this distribution is 4.

&

The variance of this distribution is 4.

Therefore,

The **coefficient of variation** of this distribution is

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{4}}{4} \times 100\% = \frac{1}{2} \times 100\% = 50\%$$

05:27

Example 2:

Suppose that Y has a Poisson distribution with $\lambda = 9$.

Then the pmf of Y is

$$p(y) = \begin{cases} \dfrac{2^y e^{-2}}{y!} & y = 0,1,2,\ldots \\ 0 & elsewhere. \end{cases}$$

« « « Mistake

2 ki jaga 9 ata hy.

06:00

The mean of this distribution is 3.

&

The variance of this distribution is 3.

Therefore,

The **coefficient of variation** of this distribution is

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{9}}{9} \times 100\% = \frac{1}{\sqrt{9}} \times 100\% = 33.33\%$$

Scanned with CamScanner

06:20

Example 3:

Suppose that Z has a Poisson distribution with $\lambda = 16$.

Then the pmf of Z is

$$p(z) = \begin{cases} \dfrac{5^{z} e^{-5}}{z!} & z = 0,1,2,\ldots \\ 0 & elsewhere. \end{cases}$$

« « « Mistake

5 ki jaga 16 ata hy.

06:57

The mean of this distribution is 16.

&

The variance of this distribution is 16.

Therefore,

The **coefficient of variation** of this distribution is

$$COV = \frac{\sigma}{\mu} = \frac{\sigma}{\mu} \times 100\% = \frac{\sqrt{\lambda}}{\lambda} \times 100\% = \frac{1}{\sqrt{16}} \times 100\%$$

$$= \frac{1}{4} \times 100\% = 25\%$$

07:34

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.

Scanned with CamScanner

03:35

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:

$$P\ \ X=1\ \equiv \frac{e^{-\ 20\ \ .08}\ \left[\ 20\ \ \ .08\ \right]^{1}}{1!}=\frac{e^{-1.6}\ \ 1.6\ ^{1}}{1!}=0.3230$$

06:03

Had the true distribution, the binomial, been used instead of the approximation,

$$P\ \ X=1\ \equiv\binom{20}{1}\ .08\ ^{1}\ \ .92\ ^{19}=0.3282$$

Scanned with CamScanner

00:19

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?

01:05

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,

$f(x) = 1/(25-0) = 1/25$

03:00

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 $a \leq x \leq b$, the Cumulative Distribution Function of the continuous uniform distribution is $F(x) = \frac{x-a}{b-a}$

So, here $F(6) = \frac{6-a}{b-a} = \frac{6-0}{25-0} = \frac{6}{25}$.

05:35

ii) There are 30 participants in the quiz

Hence the number of participants likely to answer it in 6 seconds = $(6/25) \times 30 \approx 7$.

Scanned with CamScanner

00:17

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.

01:11

Solution:

The pdf of the gamma distribution is given by

$$f(x) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, \quad 0 < x < \infty, \alpha > 0, \beta > 0$$

01:43

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 $\alpha = 2$.

02:58

- 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,

$$Mode = (2-1)\beta = 2 \Rightarrow \beta = \frac{2}{2-1} = 2.$$

Scanned with CamScanner

06:34

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 ?

07:07

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.**

07:20

The probability density function of the Beta distribution is given by

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, 0 \le x \le 1, \alpha > 0, \beta > 0$$

08:18

In this example, $\alpha = 2$ and $\beta = 5$.

Hence, our PDF is

$$f(x) = \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)}, 0 \le x \le 1$$

Scanned with CamScanner

09:04

We need to compute

$$P(20\% \le x \le 30\%) = P(0.20 \le x \le 0.30)$$

which is given by

$$P(0.2 \le x \le 0.3) = \int_{0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)} dx$$

09:04

Solving the beta function

$$B(a,b) = \frac{\Gamma\ a\ \Gamma\ b}{\Gamma\ a+b}$$

So,

$$B(2,5) = \frac{\Gamma\ 2\ \Gamma\ 5}{\Gamma\ 2+5} = \frac{\Gamma\ 2\ \Gamma\ 5}{\Gamma\ 7} = \frac{2-1\ !\ 5-1\ !}{7-1\ !}$$

$$= \frac{1\ !\ 4\ !}{6\ !} = 0.03333$$

12:24

We need to compute

$$P(20\% \le x \le 30\%) = P(0.20 \le x \le 0.30)$$

which is given by

$$P(0.2 \le x \le 0.3) = \int_{0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{B(2,5)} dx = \int_{0.2}^{0.3} \frac{x^{2-1}(1-x)^{5-1}}{0.033333} dx = 0.2352$$

13:00

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%.

Scanned with CamScanner

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0	0.5	0.504	0.508	0.512	0.516	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.591	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.648	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.67	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.695	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.719	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.758	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.791	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.834	0.8365	0.8389
1	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.877	0.879	0.881	0.883
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.898	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.937	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.975	0.9756	0.9761	0.9767
2	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.983	0.9834	0.9838	0.9842	0.9846	0.985	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.989
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.992	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.994	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.996	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.997	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.998	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.999	0.999
3.1	0.999	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Table 1: Standard Normal Cumulative Probabilities

Scanned with CamScanner

06:25

3. The Computation of Probabilities

Example:

Let X be $N(2,25)$ and suppose that we are interested in computing the probability

$$P\ 0<X<10\ = F(10)-F(0).$$

07:48

Applying the standardization formula $Z=\dfrac{X-\mu}{\sigma}=\dfrac{X-2}{5}$, we obtain

$$P\ 0<X<10\ = F(10)-F(0)$$

$$=\Phi\left(\frac{10-2}{5}\right)-\Phi\left(\frac{0-2}{5}\right)$$

$$=\Phi\left(\frac{8}{5}\right)-\Phi\left(\frac{-2}{5}\right)=\Phi\ 1.60\ -\Phi\ -0.40\ .$$

Area Table of the Standard Normal Distribution

$P\ 0<X<10\ =\Phi\ 1.60\ -\Phi\ -0.40\ .$

z	.00	.01	
0	0.5	0.504	0
0.1	0.5398	0.5438	0
0.2	0.5793	0.5832	0
0.3	0.6179	0.6217	0
0.4	0.6554	0.6591	0
0.5	0.6915	0.695	0
0.6	0.7257	0.7291	0
0.7	0.758	0.7611	0
0.8	0.7881	0.791	0
0.9	0.8159	0.8186	0
1	0.8413	0.8438	0
1.1	0.8643	0.8665	0
1.2	0.8849	0.8869	0
1.3	0.9032	0.9049	0
1.4	0.9192	0.9207	0
1.5	0.9332	0.9345	0
1.6	0.9452	0.9463	0
1.7	0.9554	0.9564	0
1.8	0.9641	0.9649	0
1.9	0.9713	0.9719	0

15:50

So

$$P\ 0<X<10\ =\Phi\left(\frac{10-2}{5}\right)-\Phi\left(\frac{0-2}{5}\right)$$

$$=\Phi\ 1.60\ -\Phi\ -0.40\ =0.945-(1-0.655)=0.600=60\%.$$

Scanned with CamScanner

03:32

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 ?

04:25

First of all, let us see what is <u>meant</u> 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 <u>fatness.</u>**
- 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.

07:42

- 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.

09:27

- 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.

Scanned with CamScanner

11:11

- 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.

Scanned with CamScanner

An example of the Normal Approx to the Binomial distribution

Example:
A fair coin is tossed 20 times. Find the probability that the number of heads occurring is between 10 and 14 inclusive by using
(a) the binomial distribution,
(b) the normal approximation to the binomial distribution.

VU

An example of the Normal Approx to the Binomial distribution

(a) Let X denote the number of heads occurring.

Then the pmf of X is given by

$$P(X=x)=\binom{20}{x}\left(\frac{1}{2}\right)^{x}\left(\frac{1}{2}\right)^{20-x}=\binom{20}{x}\left(\frac{1}{2}\right)^{20-x+x}=\binom{20}{x}\left(\frac{1}{2}\right)^{20}$$

and the desired probability is P(10≤ X≤14).

VU

An example of the Normal Approx to the Binomial distribution

Using the binomial distribution, we have:

$$P(10\le X\le 14)=\sum_{x=10}^{14}\binom{20}{x}\left(\frac{1}{2}\right)^{20}$$

$$\binom{20}{10}\left(\frac{1}{2}\right)^{20}+\binom{20}{11}\left(\frac{1}{2}\right)^{20}+\binom{20}{12}\left(\frac{1}{2}\right)^{20}+\binom{20}{13}\left(\frac{1}{2}\right)^{20}+\binom{20}{14}\left(\frac{1}{2}\right)^{20}$$

$$=0.1762+0.1602+0.1201+0.0739+0.0370$$

$$=0.5674.$$

VU

Scanned with CamScanner

An example of the Normal Approx to the Binomial distribution

Here
n=20, p=q=1/2

So
np=20(1/2)=10>5 and nq=20(1/2)=10>5.

So, the normal approximation should be adequate.

An example of the Normal Approx to the Binomial distribution

$$\mu = np = 20(0.5) = 10$$

and

$$\sigma = \sqrt{npq} = \sqrt{20(0.5)(0.5)} = 2.24.$$

An example of the Normal Approx to the Binomial distribution

Applying continuity correction, the set of discrete values X=10, 11, 12, 13,14 is replaced by the interval 9.5 ≤X≤ 14.5.
We compute the z- values as below:

$$\text{At, } x = 9.5, \text{ we find } z = \frac{9.5-10}{2.24} = -0.22, \quad \text{and}$$

$$\text{at, } x = 14.5, \text{ we find } z = \frac{14.5-10}{2.24} = 2.01$$

An example of the Normal Approx to the Binomial distribution

Hence, using the table, we find

$$P(9.5 \le X \le 14.5) = P(-0.22 \le Z \le 2.01)$$
$$= P(-0.22 \le Z \le 0) + P(0 \le Z \le 2.01)$$
$$= 0.0871 + 0.4778 = 0.5649$$

Hence, using the normal approximation to the binomial distribution, the probability of obtaining between 10 and 14 heads is 0.5649.

Scanned with CamScanner

04:32

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.

05:29

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 $N(25,25)$

05:50

(a) $P(23 \leq X \leq 26)$ becomes on continuous scale $P(22.5 \leq X \leq 26.5)$, X is $N(25,25)$.

The z-values are:

$$\text{At, } x=22.5,\ z=\frac{22.5-25}{5}=-0.5, \quad \text{and}$$

$$\text{at} \quad x=26.5,\ z=\frac{26.5-25}{5}=0.3,$$

$$\therefore\ P\ 22.5 \leq X \leq 26.5\ = P\ -0.5 \leq Z \leq 0.3$$

$$= P\ 0<Z<0.5\ + P\ 0<Z<0.3\ = 0.1915+0.1179$$

$$= 0.3094 = 30.94\%.$$

09:29

(b) $P(X>30)$ becomes on continuous scale $P(X>29.5)$

$$\therefore\ P\ X>29.5\ = P\left(\frac{X-25}{5}>\frac{29.5-25}{5}\right)$$

$$= P\left(Z>\frac{4.5}{5}\right) = P\ Z>0.9\ = 1-P\ Z<0.9$$

$$= 1-0.8159 = 0.1841 = 18.41\%.$$

Scanned with CamScanner

01:02

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).

01:59

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)?

02:50

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. $X \sim N(70, 100)$.

04:03

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

Scanned with CamScanner

07:44

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)?

09:32

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 $X=x=75$.

09:48

The pdf of the conditional distribution of Y given $X=x$ is

$$f_{Y|X}(y \mid X=x) = \frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2\sigma_Y^2(1-\rho^2)}\left(y-\left[\mu_Y+\rho\frac{\sigma_Y}{\sigma_x}(x-\mu_x)\right]\right)^2}$$

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 $\sigma_y^2\ 1-\rho^2$.

10:42

Here, $\mu_x = 70$, $\mu_y = 60$, $\sigma_x = 10$, $\sigma_y = 15$ and $\rho = 0.6$ Hence, we have

$$f_{Y|X}(y \mid X=x)$$

$$= \frac{1}{\sqrt{2\pi}\ 15\ \sqrt{1-\ 0.6^{\ 2}}} e^{-\frac{1}{2\ 15^{\ 2}(1-\ 0.6^{\ 2})}\left(y-\left[60+\ 0.6\ \frac{15}{10}(x-70)\right]\right)^2}$$

$$= \frac{1}{\sqrt{2\pi}\ 15\ \sqrt{1-0.36}} e^{-\frac{1}{2\ 15^{\ 2}(1-0.36)}\ y-\left[60+\ 0.6\ \ 1.5\ (x-70)\right]^{\ 2}}$$

Scanned with CamScanner

11:16

or

$$f_{Y|X}(y \mid X = x) = \frac{1}{\sqrt{2\pi}\ 15\ \sqrt{0.64}} e^{-\frac{1}{2\ 15^{\ 2}(0.64)}\ y-\left[60+\ 0.9\ (x-70)\right]^{\ 2}}$$

$$= \frac{1}{\sqrt{2\pi}\ 15\ \ 0.8} e^{-\frac{1}{2\ 225\ (0.64)}\ y-\left[60+\ 0.9\ (x-70)\right]^{\ 2}}$$

$$= \frac{1}{12\ \sqrt{2\pi}} e^{-\frac{1}{2\ 144}\ y-\left[60+\ 0.9\ (x-70)\right]^{\ 2}}$$

11:24

Now, since the **given value** of X is x**=75,** therefore we have

$$f_{Y|X}(y \mid X = 75) = \frac{1}{12\ \sqrt{2\pi}} e^{-\frac{1}{2\ 12^{\ 2}}\ y-\left[60+\ 0.9\ (75-70)\right]^{\ 2}}$$

$$= \frac{1}{12\ \sqrt{2\pi}} e^{-\frac{1}{2\ 12^{\ 2}}\ y-\left[60+\ 0.9\ (5)\right]^{\ 2}} = \frac{1}{12\ \sqrt{2\pi}} e^{-\frac{1}{2\ 12^{\ 2}}\ y-64.5^{\ 2}}$$

which is the pdf of the **univariate normal distribution** with mean $\tilde{\mu}_y = 64.5$ and standard deviation $\tilde{\sigma}_y = 12$.

13:34

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%.

Scanned with CamScanner

End

Edited by

Mahar Afaq Safdar Muhammadi

MSc Math, Whatsapp 03494873115

دعاؤں میں یاد رکھیے گا۔

Scanned with CamScanner