CA FOUNDATION - PAPER 3 - BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

This capsule on Foundation - Paper 3 - Business Mathematics, Logical Reasoning and Statistics , will enable the students to understand and apply the techniques of developing discrete and continuous probability distributions.

Chapter 17 : Theoretical Distributions

In this chapter we will discuss the probability theory by considering a concept and analogous to the idea of frequency distribution. In frequency distribution where to the total frequency to different class intervals, the total probability (i.e. one) is distributed to different mass points is known as theoretical probability distributions.

- Discrete Random variable.
- Continues Random variable.

Importance of theoretical probability distribution.☛:

(a) An observed frequency distribution, in many a case, may be regarded as a sample i.e. a representative part of a large, unknown, boundless universe or population and we may be interested to know the form of such a distribution.
For Example: By fitting a theoretical probability distribution.
  - Length of life of the lamps produced by manufacturer up to a reasonable degree of accuracy.
  - The effect of a particular type of missiles, it may be possible for our scientist to suggest the number of such missiles necessary to destroy an army position.
  - By knowing the distribution of smokers, a social activist may warn the people of a locality about the nuisance of active and passive smoking and so on.

(b) Theoretical probability distribution may be profitably employed to make short term projections for the future.

(c) Statistical analysis is possible only on the basis of theoretical probability distribution.

A probability distribution also possesses all the characteristics of an observed distribution. We define population mean, population median, population mode, population standard deviation etc. exactly same way we have done earlier. These characteristics are known as population parameters.

A probability distribution may be either a discrete probability distribution or a Continuous probability distribution depending on the random variable under study.

Two important discrete probability distributions

(a) Binomial Distribution
(b) Poisson distribution.

Important continuous probability distribution

Normal Distribution

Binomial Distribution

One of the most important and frequently used discrete probability distribution is Binomial Distribution. It is derived from a particular type of random experiment known as Bernoulli process named after the famous mathematician Bernoulli. Noting that a 'trial' is an attempt to produce a particular outcome which is neither certain nor impossible, the characteristics of Bernoulli trials are stated below:

(i) Each trial is associated with two mutually exclusive and exhaustive outcomes, the occurrence of one of which is known as a 'success' and as such its non occurrence as a 'failure'. As an example, when a coin is tossed, usually occurrence of a head is known as a success and its non–occurrence i.e. occurrence of a tail is known as a failure.

(ii) The trials are independent.

We may note the following important points in connection with binomial distribution:

(a) As $n > 0, p, q \geq 0$, it follows that $f(x) \geq 0$ for every x
Also $\sum f(x) = f(0) + f(1) + f(2) + ... + f(n) = 1$

(b) Binomial distribution is known as biparametric distribution as it is characterised by two parameters n and p. This means that if the values of n and p are known, then the distribution is known completely.

(c) The mean of the binomial distribution is given by $\mu = np$

(d) Depending on the values of the two parameters, binomial distribution may be unimodal or bi- modal., the mode of binomial distribution, is given by μ_0 = the largest integer contained in $(n+1)p$
if $(n+1)p$ is a non-integer $(n+1)p$ and $(n+1)p - 1$ if $(n+1)p$ is an integer

(e) The variance of the binomial distribution is given by
$\sigma^2 = npq$
Since p and q are numerically less than or equal to 1, $npq < np$ variance of a binomial variable is always less than its mean.
Also variance of X attains its maximum value at $p = q = 0.5$ and this maximum value is $n/4$.

(f) Additive property of binomial distribution.
If X and Y are two independent variables such that
$X \sim \beta(n_1, P)$
and $Y \sim \beta(n_2, P)$
Then $(X+Y) \sim (n_1 + n_2, P)$

Applications of Binomial Distribution

Binomial distribution is applicable when the trials are independent and each trial has just two outcomes success and failure. It is applied in coin tossing experiments, sampling inspection plan, genetic experiments and so on.

Poisson Distribution

Poisson distribution is a theoretical discrete probability distribution which can describe many processes. Simon Denis Poisson of France introduced this distribution way back in the year 1837.

Poisson Model

Let us think of a random experiment under the following conditions:

I. The probability of finding success in a very small time interval (t, t + dt) is kt, where k (>0) is a constant.

II. The probability of having more than one success in this time interval is very low.

III. The probability of having success in this time interval is independent of t as well as earlier successes.

The above model is known as Poisson Model. The probability of getting x successes in a relatively long time interval T containing m small time intervals t i.e. T = mt. is given by

$$\frac{e^{-kt}.(kt)^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots\ldots\ldots\ldots$$

Taking kT = m, the above form is reduced to

$$\frac{e^{-m}.m^x}{x!} \quad \text{for } x = 0, 1, 2, \ldots\ldots\ldots\ldots$$

Definition of Poisson Distribution

A random variable X is defined to follow Poisson distribution with parameter λ, to be denoted by X ~ P (m) if the probability mass function of x is given by

$$f(x) = P(X = x) = \frac{e^{-m}.m^x}{x!} \text{ for } x = 0, 1, 2, \ldots$$

$$= 0 \quad \text{otherwise}$$

Here e is a transcendental quantity with an approximate value as 2.71828.

Important points in connection with Poisson distribution:

(i) Since $e^{-m} = 1/e^m > 0$, whatever may be the value of m, m > 0, it follows that f (x)≥ 0 for every x.
Also it can be established that $\sum_x f(x) = 1$ i.e. f(0) + f(1) + f(2) +....... = 1

(ii) Poisson distribution is known as a uniparametric distribution as it is characterised by only one parameter m.

(iii) The mean of Poisson distribution is given by m i,e μ = m

(iv) The variance of Poisson distribution is given by $\sigma^2 = m$

(v) Like binomial distribution, Poisson distribution could be also unimodal or bimodal depending upon the value of the parameter m.
We have μ_0 = The largest integer contained in m if m is a non-integer
= m and m−1 if m is an integer

(vi) **Poisson approximation to Binomial distribution**
If n, the number of independent trials of a binomial distribution, tends to infinity and p, the probability of a success, tends to zero, so that m = np remains finite, then a binomial distribution with parameters n and p can be approximated by a Poisson distribution with parameter m (= np).
In other words when n is rather large and p is rather small so that m = np is moderate then $\beta(n, p) \cong P(m)$

(vii) **Additive property of Poisson distribution**
If X and y are two independent variables following Poisson distribution with parameters m_1 and m_2 respectively, then Z = X + Y also follows Poisson distribution with parameter (m_1 + m_2).
i.e. if X ~ P (m_1)
and Y ~ P (m_2)
and X and Y are independent, then
Z = X + Y ~ P (m_1 + m_2)

Application of Poisson distribution

Poisson distribution is applied when the total number of events is pretty large but the probability of occurrence is very small. Thus we can apply Poisson distribution, rather profitably, for the following cases:

a) The distribution of the no. of printing mistakes per page of a large book.
b) The distribution of the no. of road accidents on a busy road per minute.
c) The distribution of the no. of radio-active elements per minute in a fusion process.
d)

Normal or Gaussian distribution

The two distributions discussed so far, namely binomial and Poisson, are applicable when the random variable is discrete. In case of a continuous random variable like height or weight, it is impossible to distribute the total probability among different mass points because between any two unequal values, there remains an infinite number of values. Thus a continuous random variable is defined in term of its probability density function f (x), provided, of course, such a function really exists, f (x) satisfies the following condition:

$$f(x) \geq 0 \text{ for } x(-\infty,\infty) \text{ and } = \int_{-\infty}^{+\infty} f(x) = 1$$

The most important and universally accepted continuous probability distribution is known as normal distribution. Though many mathematicians like De-Moivre, Laplace etc. contributed towards the development of normal distribution, Karl Gauss was instrumental for deriving normal distribution and as such normal distribution is also referred to as Gaussian Distribution.

A continuous random variable x is defined to follow normal distribution with parameters μ and σ^2, to be denoted by

$$X \sim N(\mu, \sigma^2)$$

If the probability density function of the random variable x is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}.e^{-(\bar{x}-\mu)^2/2\sigma^2} \quad \text{for } -\infty<x<\infty$$

where μ and σ are constants, and > 0

CA FOUNDATION - PAPER 3: BUSINESS MATHEMATICS, LOGICAL REASONING AND STATISTICS

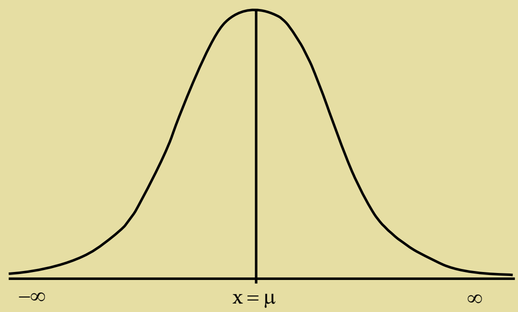
This capsule is in continuation to the previous edition featured in November 2021. Further here presented properties of Normal Distribution and their applications. Here an attempt is made to the enable the students to understand the concepts Binomial, Poisson and Normal distribution with the help of examples.

Chapter 17 : Theoretical Distributions

Some important points relating to normal distribution are listed below:

(a) The name Normal Distribution has its origin some two hundred years back as the then mathematician were in search for a normal model that can describe the probability distribution of most of the continuous random variables.

(b) If we plot the probability function y = f (x), then the curve, known as probability curve, takes the following shape:



Mean = Median = Mode

Showing Normal Probability Curve. A quick look at figure reveals that the normal curve is bell shaped and has one peak, which implies that the normal distribution has one unique mode. The line drawn through x = μ has divided the normal curve into two parts which are equal in all respect. Such a curve is known as symmetrical curve and the corresponding distribution is known as symmetrical distribution. Thus, we find that the normal distribution is symmetrical about x = μ. It may also be noted that the binomial distribution is also symmetrical about p = 0.5. We next note that the two tails of the normal curve extend indefinitely on both sides of the curve and both the left and right tails never touch the horizontal axis. The total area of the normal curve or for that any probability curve is taken to be unity i.e. one. Since the vertical line drawn through x = μ divides the curve into two equal halves, it automatically follows that,
The area between – ∞ to μ= the area between μ to ∞ = 0.5
When the mean is zero, we have
the area between – ∞ to 0 = the area between 0 to ∞ = 0.5

(c) If we take μ = 0 and σ = 1 , we have

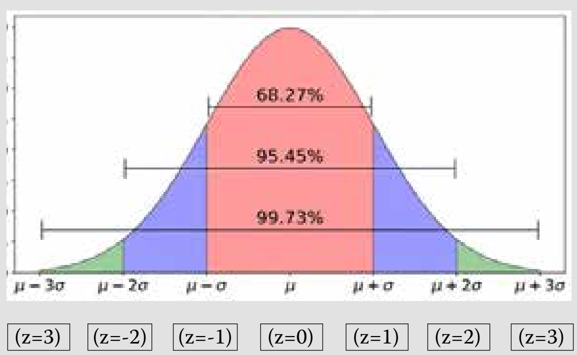
$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-2^2/2}$ for $-\infty < z < \infty$

The random variable z is known as standard normal variate (or variable) or standard normal deviate. The probability that a standard normal variate X would take a value less than or equal to a particular value say X = x is given by
$\phi(x) = p (X \le x)$
(x) is known as the cumulative distribution function.
We also have (0) = P (X ≤ 0) = Area of the standard normal curve between –∞ and 0 = 0.5

(d) **The normal distribution is known as biparametric distribution as it is characterised by two parameters μ and σ^2. Once the two parameters are known, the normal distribution is completely specified.**

Properties of Normal Distribution

1. Since π = 22/7 , $e^{-\theta} = 1 / e^{\theta} > 0$, whatever θ may be, it follows that f (x) 0 for every x.
   It can be shown that $\int_{-\infty}^{\infty} f(x)dx = 1$
2. The mean of the normal distribution is given by μ. Further, since the distribution is symmetrical about x = μ, it follows that the mean, median and mode of a normal distribution coincide, all being equal to μ.
3. The standard deviation of the normal distribution is given by
   Mean deviation of normal distribution is $\sigma \frac{\sqrt{2}}{\pi}$
   The first and third quartiles are Q_1 = μ– 0.675 σ and Q_3 = μ+ 0.675 σ
   so that, quartile deviation = 0.675 σ
4. The normal distribution is symmetrical about x =μ. As such, its skewness is zero i.e. the normal curve is neither inclined move towards the right (negatively skewed) nor towards the left (positively skewed).
5. The normal curve y = f (x) has two points of inflexion to be given by x = μ – σ and
   x = μ + σ i.e. at these two points, the normal curve changes its curvature from concave to convex and from convex to concave.
6. If x ~ N (μ, σ^2) then z = x – μ/σ ~ N (0, 1), z is known as standardised normal variate or normal deviate.
   We also have P (z ≤ k) = ϕ (k) The values of ϕ(k) for different k are given in a table known as "Biometrika."
7. Area under the normal curve is shown in the following figure:



Area Under Normal Curve

From the figure , we have

$P(\mu - \sigma < x < \mu + \sigma) = 0.6828$

=> $P(-1 < z < 1) = 0.6828$

$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0.9546$

=> $P(-2 < z < 2) = 0.9546$

and $P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973$

=> $P(-3 < z < 3) = 0.9973$.

We note that 99.73 per cent of the values of a normal variable lies between $(\mu - 3\sigma)$ and $(\mu + 3\sigma)$. Thus the probability that a value of x lies outside that limit is as low as 0.0027.

8. If x and y are independent normal variables with means and standard deviations as μ_1 and μ_2 and σ_1, and σ_2 respectively, then $z = x + y$ also follows normal distribution with mean $(\mu_1 + \mu_2)$ and SD = $\sqrt{\sigma_1^2 + \sigma_2^2}$ respectively.

   i.e. If $x \sim N(\mu_1, \sigma_1^2)$ and $y \sim N(\mu_2, \sigma_2^2)$ and x and y are independent,

   then $z = x + y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

Applications of Normal Distribution

Most of the continuous variables like height, weight, wage, profit etc. follow normal distribution. If the variable under study does not follow normal distribution, a simple transformation of the variable, in many a case, would lead to the normal distribution of the changed variable.

When n, the number of trials of a binomial distribution, is large and p, the probability of a success, is moderate i.e. neither too large nor too small then the binomial distribution, also, tends to normal distribution. Poisson distribution, also for large value of m approaches normal distribution. Such transformations become necessary as it is easier to compute probabilities under the assumption of a normal distribution.

Some Important Problems

1. What is the probability that out of 10 missiles fired, atleast 2 will hit the target

Solution: Probability of atleast 2 will hit the target is given as,

$P(X \geq 2) = 1 - P(X < 2)$

The probability of a missile hitting a target is 1 /8

$P(X \geq 2) = 1 - [P(X = 0) + P(X = 1)]$

$= 1 - [{}^{10}C_0 \times (1/8)\text{^}0 \times (7/8)\text{^}10 + {}^{10}C_1 \times (1/8)\text{^}1 \times (7/8)\text{^}9]$

= 1 – [7^10/8^10 + (10 × 7^9)/8^10]

= 1 – [7^10 + 10 × 7^9] / 8^10

= 1 – (17 × 7^9) / 8^10

$P(X \geq 2) \approx 0.3611 = 36.11\%$

Therefore, the probability that out of 10 missiles fired, atleast 2 will hit the target is 0.3611.

2. Given X is a binomial variable such that 2 P(X = 2) = P(X = 3) and mean of X is known to be 10/3. What would be the probability that X assumes at most the value 2

Solution:

mean = 10/3 , Mean = np = 10/3

$P(X) = {}^nC_x p^x (1-p)^{n-x}$

=> $P(2) = {}^nC_2 p^2 (1-p)^{n-2}$

$P(3) = {}^nC_3 p^3 (1-p)^{n-3}$

P(3) = 2 P(2)

=> ${}^nC_3 p^3 (1-p)^{n-3} = 2\,{}^nC_2 p^2 (1-p)^{n-2}$

=> p/3!(n-3)! = (2 /2!(n-2)!)(1 - p)

=> p/6 = (1-p)/(n - 2)

=> np - 2p = 6 - 6p

=> 4p = 6 - np

=> 4p = 6 - 10/3

=> 4p = 8/3

=> p = 2/3

np = 10/3 => n = 5

probability that X assumes at most the value 2 = P(0) + P(1) + P(2)

$= {}^5C_0(2/3)^0(1/3)^5 + {}^5C_1(2/3)^1(1/3)^4 + {}^5C_2(2/3)^2(1/3)^3$

= 1/243 + 10/243 + 40/243 = 51/243= 17/81

3. Assuming that one-third of the population is tea drinkers and each of 1000 enumerators takes a sample of 8 individuals to find out whether they are tea drinkers or not, how many enumerators are expected to report that five or more people are tea drinkers?

Solution: Assume that being a tea drinker is like taking a flip of coin.(i.e. either the person drinks tea or not) with the probability of heads being 1/3 and tails being 2/3

now $P(X\geq 5)=P(X=5)+P(X=6)+P(X=7)+P(X=8)=(56*8+28*4+8*2+1)/6561$ = 577/6561 =0.087943 so for 1000 trials the number of trials which report greater than or equal to 5 is 1000*P(X>=5) = 87.94 =88

4. If a random variable x follows binomial distribution with mean as 5 and satisfying the condition 10. P (x= 0) = P(x= 1) , what is the value of $P(X\geq 1/X>1)$?

Solution:

$$\text{Here Mean } np = 5$$

$$10.P(x=0) = P(x=1)$$

$$10.{}^nC_0.p^0.(1-p)^n = {}^nC_1.p.(1-p)^{n-1}$$

$$10.(1-p)^n = np(1-p)^{n-1}$$

$$10(1-p) = 5$$

$$1-p = 1/2$$

$$p = 1/2, n = 10$$

$$p\left(x \geq 1 / x \geq 0\right) = \frac{p(x \geq 1)}{p(x \geq 0)}$$

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

$$= 1 - \frac{1}{2^{10}}$$

$$= 0.99$$

5. Out of 128 families with four children each, how many are expected to have atleast have one boy and one girl?

Solution: 4 children in a family can be in

2 * 2 * 2 * 2 = 16 ways

at least one boy and one girl = Total cases - all boys - all girs

All boys = 1 case, All girls = 1 case

=> at least one boy and one girl = 16 - 1 - 1 = 14

Probability of at least one boy and one girl = 14/16

out of 128 families expected to have = 128 * 14/16 = 8 * 14 = 112

112 Families expected to have at least one boy and one girl

6. In 10 independent rollings of a biased die, the probability that an even number will appear 5 times is twice the probability that an even number will appear 4 times. What is the probability that an even number will appear twice when the die is rolled 8 times?

Solution: Probability of even number p = p

Then probability of odd number (or not even number) = q = 1 - p

Probability Appearing 5 times

${}^{10}C_5 * p^5 * q^{10-5}$

$= {}^{10}C_5 * p^5 * q^5$

Probability Appearing 5 times

$^{10}C_4 * p^4 * q^{10-4}$

$= {}^{10}C_4 * p^4 * q^6$

$^{10}C_5 * p^5 * q^5 = 2 * {}^{10}C_4 * p^4 * q^6$

=>p * 10!/5!5! = 2q * 10!/6!4!

=> p * 6 = 2q * 5

=> 3p = 5q => 3p = 5(1 - p)

=> 8p = 5

=> p = 5/8 & q = 1 - 5/8 = 3/8

probability that an even number will appear twice when the die is rolled 8 times

$= {}^8C_2 * p^2q^{8-2}$

$= 28 * (5/8)^2 (3/8)^6$

$= 28 * 25 * 3^6 / 8^8$

$= 700 * 3^6 / 8^8$

= 5,10,300/1,67,77,216 = 0.0304

7. What is the probability of making 3 correct guesses in 5 true-False answer type questions?

Solution: Here P = 0.5 and q = 0.5 (Since answer can be either True or False) n= 5, r = 3

$P(x=r) = {}^nC_r\, p^x.q^{n-x}$

$P(x=3) = {}^5C_3\, (0.5)^5 .(0.5)^{5-3} = 0.3125$

8. Suppose that weather records show that on an average 5 out of 31 days in October are rainy days. Assuming a binomial distribution which each day of October as an independent trail, then the probability that the next October will have at most three rainy days is:

Solution: p = Probability of a rainy day in October :p = 5/31, q = probability of a non-rainy day in October q = 1- p = 1-5/31 = 26/31

n = 31 (number of days in October)

$$P(x) = 31Cr.\left(\frac{5}{31}\right)^r.\left(\frac{26}{31}\right)^{31-r}$$

Required Probability = P (0) + P (1) +P(2) +P(3)

$$= 31C_0.\left(\frac{5}{31}\right)^0.\left(\frac{26}{31}\right)^{31} + 31C_1.\left(\frac{5}{31}\right)^1.\left(\frac{26}{31}\right)^{31-1} +$$

$$31C_2.\left(\frac{5}{31}\right)^2.\left(\frac{26}{31}\right)^{31-2} + 31C_3.\left(\frac{5}{31}\right)^3.\left(\frac{26}{31}\right)^{31-3}$$

= 0.2403

9. If 5 days are selected at random, then the probability of getting two Sundays is:

Solution: Let P = Probability of getting a Sunday in a week (P) = 1/7

Therefore P = 1/7 and q= 1-p = 1-1/7 = 6/7

$$\text{Required probability} = 15C2.\left(\frac{1}{7}\right)^2.\left(\frac{6}{7}\right)^{15-2} = 0.288 = 0.29$$

10. An experiment of succeeds twice as often as it falls. What is the probability that in next five trials there will be at least three successes?

Solution: According to the given statement p = 2q

We know that p = 2/3 q= 1/3

Required probability P(X≥3) = P(3) +P(4) +P(5)

$$= 5C3.\left(\frac{2}{3}\right)^2.\left(\frac{6}{7}\right)^{15-2} + 5C4.\left(\frac{2}{3}\right)^2.\left(\frac{6}{7}\right)^{15-2} + 5C5.\left(\frac{2}{3}\right)^3.\left(\frac{1}{3}\right)^2$$

11. What is the probability of getting 3 head if 6 unbiased coins are tossed simultaneously?

Solution: if x denotes the number of heads, then x follows binomial distribution with parameters n = 6 and p = probability of success = ½

q= probability of failure = 1-1/2 = ½ , being given the coins are unbiased

The probability mass function of x is given by

$$f(x) = {}^6C_3.\left(\frac{1}{2}\right)^3.\left(\frac{1}{2}\right)^3 = 20\times\left(\frac{1}{2}\right)^3 = 0.3125$$

12. In Binomial Distribution n = 9 and p = 1/3 , What is the value of variance .

Solution: In Binomial Distribution variance = npq , here n= 9 , p = 1/3 and q= 2/3

Therefore variance = 9*1/3*2/3 = 2

13. For a Binomial Distribution E(x) = 2 , V(X)= 4/3 . Find the value of n

Solution: Here E(X) = np = 2

V(x)= npq = 4/3 then substituting the value of np

2×q = 4/3

2q = 4/3

q= 2/3 then p = 1-2/3 = 1/3

np = 2, n×1/3 = 2 , n= 6

14. The mode of Binomial Distribution for which the mean is 4 and variance 3 is equal to?

Solution: In Binomial Distribution Mean = np = 4 and Variance = npq = 3

Then 4q = 3 q = ¾ , p = 1-q = 1-3/4 = ¼

n×1/4 = 4 therefore n = 16

(n+1) p = (16+1) ×1/4 = 4.25 which is no longer integer.

So mode = 4

15. In a Binomial Distribution with 5 independent trials, probability of 2 and 3 successes are 0.4362 and 0.2181 respectively. Parameter 'p 'of the Binomial Distribution is

Solution: Given n = 5 , P(x=2) = 0.4362

P(x=3) =0.2181

$P(x=3) = 5C3.(p)^3.(q)^{5-3} = 10(p)^3.(q)^2$

$0.2181 = 10.P^{3}.q^{2}$

And $P(X=2) = 5C_2.(p)^3.(q)^{5-2} = 10(p)^2.(q)^3$

$0.4362 = 10(p)^2.(q)^3$

By dividing

$$0.2181/0.4362 = \frac{10.p^3.q^2}{10.p^2.q^3}$$

½= q/p

q= 2p; 2p+p = 1

3p = 1, then p = 1/3

16. What is the first quartile of X having the following probability density function?

$$f(x) = \frac{1}{\sqrt{72\pi}}\, e^{\frac{-(x-10)^2}{72}} \quad \text{for} -\infty < x < \infty$$

Solution: First Quartile Deviation (Q_1) = μ- 0.675σ here

By comparing probability density function of Normal Distribution, Mean = 10 and σ = 4

Q_1 = 10-0.675(6) =10-4.05 = 5.95