TUTORIAL-II
Statistical Learning
(Part-A)

OBJECTIVES

1. Which of the following mentioned standard Probability density functions is applicable to discrete Random Variables?

a) Gaussian Distribution
b) Poisson Distribution ✓
c) Gamma Distribution
d) Exponential Distribution

OBJECTIVES

2. If the values taken by a random variable are negative, the negative values will have ________

a) Positive probability ✓
b) Negative probability
c) May have negative or positive probabilities
d) Insufficient data

OBJECTIVES

3. The expected value of a random variable is its ____________

a) Mean ✓
b) Standard Deviation
c) Mean Deviation
d) Variance

OBJECTIVES

4. In a Binomial Distribution, if 'n' is the number of trials and 'p' is the probability of success, then the mean and variance is given by ____________

a) np, np(1-p)
b) np(1-p), np
c) n, np^2
d) p, np

OBJECTIVES

5. For larger values of ‘n’, Binomial Distribution ______________

a) loses its discreteness
b) tends to Poisson Distribution ✓
c) stays as it is
d) gives oscillatory values

OBJECTIVES

6. The recurrence relation between P(x) and P(x +1) in a Poisson distribution is given by ____________

(a) $P(x+1) - m\,P(x) = 0$
(b) $m\,P(x+1) - P(x) = 0$
(c) $(x+1)\,P(x+1) - m\,P(x) = 0$
(d) $(x+1)\,P(x) - x\,P(x+1) = 0$

OBJECTIVES

7. In the following Table, Column A lists some sampling distributions, whereas Column B lists the name of sampling distributions. All symbols bear their usual meanings.

Column A		Column B	
(A)		(W)	Normal distribution
(B)		(X)	Chi-squared distribution
(C)		(Y)	t-distribution
(D)		(Z)	F distribution

Some matchings from Column A and Column B are given below. Select the correct matching?

a) (A)-(Z), (B)-(X), (C)-(Y), (D)-(Z)
b) (A)-(X), (B)-(Z), (C)-(W), (D)-(W)
c) (A)-(Y), (B)-(W), (C)-(X), (D)-(Y)
d) (A)-(W), (B)-(Y), (C)-(Z), (D)-(X) ✓

OBJECTIVES

8. Which of the following statement sounds reasonable?

a) *If the sample size increases sampling distribution must approach normal distribution.* ✓

a) *If the sample size decreases then the sample distribution must approach normal distribution.*

a) *If the sample size increases then the sampling distribution much approach an exponential distribution.*

a) *If the sample size decreases then the sampling distribution much approach an exponential distribution.*

OBJECTIVES

9. If μ and σ denote the mean and standard deviation of a population, then the standard normal distribution is better described as (select the correct option from the list of options given below):

a) $f(x: A, B) = \begin{cases} \frac{1}{B-A} & A \leq x \leq B \\ 0 & Otherwise \end{cases}$

b) $f(x: \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$

c) $f(z: 0,1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \qquad -\infty < z < \infty$

d) $f(x: \mu, \sigma) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[\ln(x)-\mu]^2} & x \geq 0 \\ 0 & x < 0 \end{cases}$

OBJECTIVES

10. If f(x) is a probability density function of any continuous random variable, then which of the follow statement(s) is(are) NOT correct?

$a)\ 0 \le f(x) \le 1$ ✓

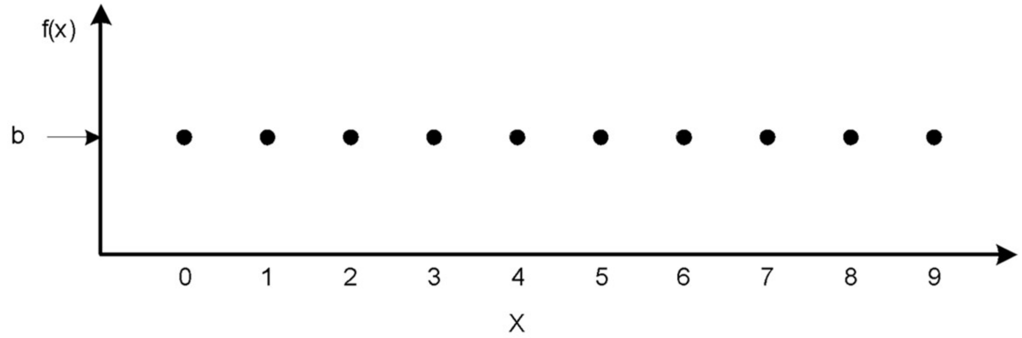
$b)\ P(a \le X \le b) = \int_b^a f(x)dx\ <\ 1$ ✓

$c)\ y = \int_{-\propto}^{\propto} xf(x)dx\ \ there\ exist\ y \in R$

$d)\ z = \int_{-\propto}^{\propto} (x-\mu)^2 f(x)dx\ \ there\ exist\ \mu \in R\ and\ z \in R$

NUMERICALS

1. A bitcoin if you toss it gives any value in the range [0…9] both inclusive. Assume that the random variable X represents the toss value of the bitcoin which has a discrete uniform distribution and is shown in the following figure.



a) What is $P(X = x_i) = f(x_i)$ for any value of $x_i \in [0 \dots 9]$?
b) What is the value of b (as marked in the figure)?
c) Calculate the mean μ of this distribution.
d) Calculate the variance σ^2 of this distribution.

ANS:

a) $P(X = x_i) = f(x_i) = \frac{1}{10} = 0.1$

b) $b = f(x_i) = \frac{1}{10} = 0.1$

c) $\mu = \sum_1^{10} x_i \cdot f(x_i) = \frac{b+a}{2} = \frac{9+0}{2} = 4.5$

d) $\sigma^2 = \sum_1^{10} (x_i - \mu)^2 \cdot f(x_i) = \frac{(b-a+1)^2-1}{12} = \frac{10^2-1}{12} = 8.25$

NUMERICALS

2. A quiz test for a course Data Analytics was conducted for a total score of 100 where 600 students took the test. From the result of the test it was found that mean score μ = 90 and standard deviation σ = 20. Students are randomly distributed among six sections and each section includes 100 students. In one of the section of 100 students, the mean score is found as 86.

a) What is the standard error rate?
b) If you select any section at random, what is the probability of getting a mean score is 86 or lower?

ANS:-

As per the Central Limit Theorem, the standard error is $\varepsilon = \frac{\sigma}{\sqrt{n}} = \frac{20}{10} = 2.0$

The sample distribution statistics can be obtained with the z-distribution. For the sample,

The probability of getting 86 or lower is P(Z<-2.0). From the standard normal distribution table it is found that P(Z<-2.0) = 0.0228.

NUMERICALS

3. On average, 3 traffic accidents per month occur at a certain intersection. What is the probability that in any given month at this intersection

a) exactly 5 accidents will occur?
b) fewer than 3 accidents will occur?
c) at least 2 accidents will occur?

Ans:-

$$a)\ P(X=5)=\frac{e^{-3}.3^5}{5!}=0.1008$$

$$b)\ P(X<3)=P(X\le 2)=\ e^{-3}\cdot\sum_{x=0}^{2}\frac{3^x}{x!}=0.4232$$

$$c)\ P(X\ge 2)=1-P(X\le 1)=1-e^{-3}\cdot\sum_{x=0}^{1}\frac{3^x}{x!}=0.8009$$

NUMERICALS

4. A large chain retailer purchases a certain kind of electronic device from a manufacturer. The manufacturer indicates that the defective rate of the device is 3%.

a) The inspector randomly picks 20 items from a shipment. What is the probability that there will be at least one defective item among these 20?

b) Suppose that the retailer receives 10 shipments in a month and the inspector randomly tests 20 devices per shipment. What is the probability that there will be exactly 3 shipments each containing at least one defective device among the 20 that are selected and tested from the shipment?

ANS:-

a) Denote by X the number of defective devices among the 20. Then X follows a $b(x;\ 20, 0.03)$ distribution. Hence,

$$P(X \geq 1) = 1 - P(X = 0) = 1 - b(0;\ 20, 0.03) = 1 - (0.03)^0\,(1 - 0.03)^{20-0} = 0.4562.$$

b) In this case, each shipment can either contain at least one defective item or not. Hence, testing of each shipment can be viewed as a Bernoulli trial with $p = 0.4562$ from part (a). Assuming independence from shipment to shipment and denoting by Y the number of shipments containing at least one defective item, Y follows another binomial distribution $b(y;\ 10, 0.4562)$. Therefore,

$$P(Y = 3) = \binom{10}{3} 0.4562^3 (1 - 0.4562)^7 = 0.1602$$

NUMERICALS

5. A soft-drink machine is regulated so that it discharges an average of 200 milliliters per cup. If the amount of drink is normally distributed with a standard deviation equal to 15 milliliters,

a) what fraction of the cups will contain more than 224 milliliters?
b) what is the probability that a cup contains between 191 and 209 milliliters?
c) how many cups will probably overflow if 230- milliliter cups are used for the next 1000 drinks?
d) below what value do we get the smallest 25% of the drinks?

ANS:-

a) $z = \frac{224-200}{15} = 1.6$ Fraction of the cups containing more than 224 millimeters is $P(Z > 1.6) = 0.0548$

b) $z_1 = \frac{191-200}{15} = -0.6$, $z_2 = \frac{209-200}{15} = 0.6$;

$$P(191 < X < 209) = P(-0.6 < Z < 0.6) = 0.7257 - 0.2743 = 0.4514$$

c) $z = \frac{230-200}{15} = 2.0$; $P(X > 230) = P(Z > 2.0) = 0.0228$ Therefore, $(1000)(0.0228) = 22.8$ or approximately 23 cups will overflow.

d) $z = -0.67, x = (15)(-0.67) + 200 = 189.95$ millimeters