- example - 7 : If $E = \{H\}$ & $F = \{T\}$ then EF would be 'null' or 'ϕ'. Then E and F are mutually exclusive.

* Union and Intersection of more than two events can be denoted $\bigcup_{n=1}^{\infty} E_n$ & $\bigcap_{n=1}^{\infty} E_n$

19/2/24

~~Discrete~~ Random Variable

A discrete random variable takes a finite set of values where the continuous random variable takes an infinite set of values

In probability, we define discrete variable with Probability Mass Function (PMF). The PMF assign probability to every possible variable specific to the given data attribute

$$DRV = \sum_{x=X} P(X) = 1 \quad \text{—— (1)}$$

The probability of continuous variable can be defined using Probability Density Function (PDF). ~~The~~ As the continuous variables are not finite, we use an integral in defining PDF

$$CRV = \int p(x) \cdot dx = 1$$

- Probability Distribution :

It defines the likelihood of possible values that a random variable can take. ~~and~~

- Joint probability Distribution

This determines the probability distribution for two or more random variables. $P(XY)$

- Conditional probability

There are cases where we want to compute the probability of any value having the different event already occured.

$$\therefore P(X|Y) = \frac{P(Y|X)\,P(X)}{P(Y)}$$

eg: Suppose a number of item manufactured ~~du~~ in a factory during a week is a random variable with a mean of 50. i) What is the probability that the current week's production will exceed 75

26/2/24

Bayes' Rule :- It defines the probability of an event computed using the prior knowledge of the related events of the occurring event.

- Bayes' Rule, plays a significant role in statistics and machine learning where probability is believed to be a degree of belief in an event.

Prior

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)}$$

posterior

* Expected value : The expectation or expected value is the mean of multiple repetitions of an event.

> A variable can take several possible values or probabilities where each value has the attached likelihood. Summing up this details into a single variable return us the expectation. $E[X]$

* Variance : It defines, how the output of an event varies as the values influencing the event are picked from the probability distribution. It basically defines how the one value differs from the other values. $Var(X)$

* Standard Deviation : It gives the spread of the data set, as, how far the values are from the mean or expected value. It is given by the square root of variance.

$$Std(X) = \sqrt{Var(X)}$$

* Co-Variance : It defines, the ~~it~~ linear relationship between two variables.

> If it returns positive then both the variables tend to take higher value. And if its negative then one variable takes higher value and another one takes lower value.

Types of probability distribution:

1. Bernoulli's Distribution

It is the probability distribution of a RV where the RV is 1 with a probability of 'P' and it is 0 with a probability of '1-P'. This typically returns the true/false type of classification scenarios.

2. Binomial Distribution

Multiple Bernoulli's trial constitute a binomial distribution.

3. Multinoulli Distribution

This is the case where a single variable can have multiple outcomes so it is the transition from the binary to several categories.

The Multinoulli distribution comes in the picture when the problem is of multi-class classification problem.

4. Multinomial Distribution

Multiple trials of Multinolli distribution constitute a multinomial distribution

*5. Gaussian / Normal Distribution

It is the most commonly used distribution in machine-learning and statistics.

The normalized sum of several independent variable is inclined towards gaussian type of distribution irrespective of the distribution take by the individual variable.

4/3/24

Conditional Probability :-

eg. Suppose we roll 2 dice and suppose that we observe the first dice '4' then given this information what is the probability that sum of the two dice is '6'. The answer to this question requires the concept of conditional probability, where the calculation of present event is based on priorly occured event.

—› If 'E' denotes that the event that the sum of dice is '6' and 'F' becomes an event that the first dice is '4' then probability just obtained is called conditional probability that the Event 'E' occured given that event 'F' has already been occured. The conditional probability of this scenario is given by P(E/F).

—› In this scenario, the event 'F' occurs then in order for 'E' to occur it is necessary for the actual occurence to be a point in both 'E' and 'F', i.e. it must be 'EF' ($E \cap F$)

—› Because we know that the event 'F' already been occured, it follows that the 'F' becomes our new sample space and the probability of 'EF' is relative to the probability of 'F'.

Date
Page

$$P(E/F) = \frac{P(EF)}{P(F)} \quad \text{--- (1)}$$

-4 The eq (1) is only well defined when the probability of 'F' is greater than 0. and hence $P(E|F)$ is only defined when $P(F) > 0$.

eg 1: Suppose cards numbered from 1 to 10 are placed in a box and not in sequence then one of the cards is drawn. If you are told that the number on the drawn card is atleast '5' then what is the probability than it is '10'

-4 E = no. of drawn card is 10. $= 1/10$
F = no. of drawn card is atleast 5. $= 6/10$.

$$\therefore P(E|F) = \frac{P(EF)}{P(F)}$$

$$= \frac{(1/10 \cap 6/10)}{6/10}$$

$$= \frac{1/10}{6/10}$$

$$\therefore \boxed{P(E/F) = \frac{1}{6}}$$

eg2: To calculate what is the probability that both the children are boys in a family having two children.

-4 B = both the children are boys. $= 1/4$
A = at least one of them is a boy $= 3/4$

classmate

Date

Page

$S = \{(b,b), (b,g), (g,b), (g,g)\}$

$\therefore P(B|A) = \frac{P(BA)}{P(A)}$

$= \frac{(1/4 \cap 3/4)}{3/4}$

$= \frac{1/4}{3/4}$

$\therefore \boxed{P(B|A) = \frac{1}{3}}$

Q3. To check whether the email is spam or not.

① There are total 100 emails

② 40 are label spam and 60 are label not spam.

③ The spam or not spam emails are considered based on single feature word 'offer'.

Out of 40 spam 30 contains word offer and out of 60 non-spam mails 5 contains word 'offer'.

$P(S) = 0.4$ $P(NS) = 0.6$

→ 4. $P(O|S) = \frac{30}{40} = 0.75$

$P(O|S) = \frac{30}{40} = 0.75$

$P(O|NS) = \frac{5}{60} = 0.0833$

$P(S|O) = 0.$

$P(S|O) = 0.8571$

$\therefore P(O) = 0.35$

$P(NS|O) = \frac{P(NS) \cap P(O)}{P(O)}$

$= \frac{\frac{5}{100}}{35/100} = \frac{1}{7} = 0.1429$

$\boxed{P(S|O) + P(NS|O) = 1}$

8/3/24

eg: We have a dataset of 300 news articles where 150 are related to politics and 150 are related to sports.

> Presence and absence of certain keywords like "Election and President" for politics and "football and basketball" for sports represents the type of article.

(1) Out of 150 political articles, 120 contains the word "Election" and 80 articles contain the word "President".

(2) Out of 150 sport articles, 60 contains the word "football" and 90 contains the word "basketball".

-4 $P(P) = 0.5$ $\quad P(S) = 0.5$

$P(\text{"Election"}|P) = \frac{120}{150} = 0.8$ $\quad P(\text{"President"}|P) = \frac{80}{150} = 0.53$

$P(\text{"Football"}|S) = \frac{60}{150} = 0.4$ $\quad P(\text{"basketball"}|P) = \frac{90}{150} = 0.6$

• $P(P|\text{"Election"}, \text{"President"}) = \frac{P(\text{"Election"}, \text{"President"}|P)\,P(P)}{P(\text{"Election"}, \text{"President"})}$

• M_{PE}, politics $= 100$

• $M_{PE} = 120$ $\quad = \frac{100/150 \cdot 150/300}{120/300}$

• $M_{total} = 300$

• $M_{politics} = 150$ $\quad = \underline{0.83}$

E[X] : Instead of presenting the entire dataset, a single number is given to summarize the data.

• Instead of giving the entire distribution, we could give a single number that summarizes the distribution eg: let x be the DRV with a range $\{x_1, x_2, \ldots\}$ and PMF (probability Mass Function) 'P' then the expected value of variable 'x' is defined as

$$\therefore E[X] = \sum_{k=1}^{\infty} x_k \, P(x_k) \quad \text{——— (2)}$$

PMF, DRV

Thus the expected value is weighted average of all the possible values of 'x' with the corresponding probabilities as weights

$$\therefore E[X] = \sum_{k=1}^{2} x_k \, P(x_k) \qquad x = \begin{cases} 1 & \text{with } p \text{ odd} \\ -1 & \text{with } p \text{ even} \end{cases}$$

$$= 1 \cdot \frac{18}{38} + (-1) \cdot \frac{20}{38}$$

$$= \frac{18}{38} - \frac{20}{38}$$

$$= \frac{-2}{38}$$

$$= -1/19 = -0.05263158$$

eg: Flip a coin twice and let 'x' denotes the number of heads. Then what will be the range of x and what will be the associated probability.

$$x = \begin{cases} 0, & H \quad 1/4 \\ 1, & H \quad 2/4 \\ 2, & H \quad 1/4 \end{cases}$$

$$\therefore E[X] = \sum_{k=1}^{3} x_k \, P(x_k)$$

$$= 0 \cdot \frac{1}{4} + 1 \cdot \frac{2}{4} + 2 \cdot \frac{1}{4}$$

$$= 1/2 + 1/2 = 1$$

BBB BBG BGB . BGG

GGG GGB GBG GBB.

eg: let x be the no. of daughters in a family with 3 children.

→

$$x = \begin{cases} 0, & 1/8 \\ 1, & 3/8 \\ 2, & 3/8 \\ 3, & 1/8 \end{cases}$$

• Types of DRV:

i) Bernoulli's RV

Suppose that a trial or an experiment whose outcomes can be classified either as a success or a failure is performed. Let 'x' = 1 if the outcome is a success ~~and~~ or 0 when it is a failure, then the PMF is given by

$$P(0) = P\{x=0\} = 1-P$$
$$P(1) = P\{x=1\} = P. \quad \text{—— (3)}$$

A given RV is said to Bernoulli if its PMF is given by the above eq (3) where $P \in \{0, 1\}$

ii) Binomial RV.

Suppose that 'n' independent trials where each of the trial results in a success with a probability 'P' and in a failure with probability '$1-P$' are to be performed. If 'x' represents the no. of success that occurs in 'n' no. of trials then 'x' is said a binomial RV with parameters (n, P).

The PMF of any binomial RV is given by

$$P(P') = \binom{n}{P} P_i (1-P)^{n-i}$$

where $i = 0, 1, 2, \ldots n$.

HHTT HTHT HTTH TTHH ~~HHTH~~

TTHH THTH THHT HHTT ~~THHT~~

~~TTHH~~

$$\binom{n}{i} = \frac{n!}{(n-i)!\,i!}$$

eg: Four fair coins are flipped. ~~with~~ It the outcomes are assumed independent then what is the probability that 2 heads and 2 tails are obtained.

→ $n = 4 \qquad p = 1/2$

$$\therefore P(p') = \binom{n}{i} \cdot p_i (1-p)^{n-i}$$

$$= \binom{4}{1/2} \cdot p_2 (1 - 1/2)^{4-2}$$

$$= 6 \cdot \frac{4}{16} (1 - 0.5)^2$$

$$= \frac{24^{6}}{16} \times \frac{1}{4}$$

$$= \frac{\cancel{6}\,3}{\cancel{16}\,8} = \frac{3}{8}$$

$$\frac{n!}{(n-i)!\,i!} = \frac{4!}{(4-2)!\,2!} = \frac{24}{2!\,2!} = \frac{24}{2 \times 2} = 6.$$

eg: It is known that any item produced by a certain machine will be ~~is~~ defective with a probability of 0.1. What is the probability that in a sample space of 3 items, atmost 1 will be defective. x with parameter (3, 0.1)

→ $P(p') = \binom{3}{0} \cdot p_0 (1 - 0.1)^{3-0} + \binom{3}{1} \cdot p_1 (1 - 0.1)^{3-1}$

$$= 1 \times p^{(i)} \left(1 - \frac{1}{10}\right)^3 + 3 \cdot p^{(i)} \left(\frac{9}{10}\right)^2$$

$$= 1 \times (0.1)^0 (0.9)^3 + 3 \times (0.1)^1 (0.9)^2$$

$$= 0.729 + 0.243$$

$$= 0.972$$

iii) Geometric RV.

Suppose that independent trials where each having probability, p having a success are performed until a success occurs. If 'x' be the no. trials required until the first success then x is said to be a geometric RV with parameter p and its PMF is given by.

$P(x) = P(X=x) = (1-p)^{x-1} \cdot p$ where $x = 1, 2, 3.$

eg: x is no. of coins flipped until first head lets find the probability that we need to flip the coin exactly 3 times before getting the first head.

→ $x = 3.$

$$P(x) = (1-p)^{x-1} \cdot p$$

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

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

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

eg: Suppose you are playing a game where roll a six side fair die repeatedly until you roll a six. What is the probability that you need to roll a die exactly 4 time before rolling a 6 for first time

→ $x = 4$

$$P(x) = \left(1-\frac{1}{6}\right)^{3} \cdot \frac{1}{6}$$

$$= \left(\frac{6-1}{6}\right)^{3} \times \frac{1}{6} = \left(\frac{5}{6}\right)^{3} \times \frac{1}{6} = \frac{125}{216} \times \frac{1}{6} = 0.096$$

classmate

Date
Page

eg: Imagine you are conducting a series of medical test ~~disease~~ to detect a certain disease. Each test has a fixed probability of 0.8 for correctly detecting the disease. What is the probability that the disease is correctly diagonsed with a first 3 test

→ $x = 3$.

$$P(x \leq 3) = 1 - P(x > 3)$$

15/3/24

Continuous Random Variable:-

The continuous Random Variable are the category of RV whose set of possible values is uncountable.

Let some x be RV, we say that x is continuous RV if there exist some non-negative function $f(x)$ which is defined for all real $x \in \{-\infty, \infty\}$ having the property that any set B of real numbers satisfy

$$P(x \in B) = \int_B f(x) . dx$$

The function '$f(x)$' is called the PDF of RV x.

All probabilities statement about the RV 'x' can be answered in terms of '$f(x)$'.

Suppose, $B = [a, b]$

$$\therefore P(a \leq x \leq b) = \int_a^b f(x) . dx$$

* Types of CRV

i) Uniform RV $\qquad F(a) = P$

ii) Exponential RV $\qquad x \in (-\infty, a) = \int_{-\infty}^{a} f(x) \cdot dx$

iii) Gamma RV

iv) Normal RV.

1. The Uniform Random Variable.

A RV is said to be uniformly distributed over the interval $(0,1)$, if its PDF is given by

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

In General, we can say that the variable x is the uniform RV on the interval of (α, β), if its PDF is given by

$$f(x) = \begin{cases} \frac{1}{\beta - \alpha}, & \text{for } \alpha < x < \beta \\ 0, & \text{otherwise.} \end{cases}$$

eg: 1 If the RV 'x' is uniformly distributed over 0 to 10, then calculate the probability that

i) $x < 3$ $\qquad$ ii) $x > 7$ $\qquad$ iii) $1 < x < 6$.

—

(i) $P(a < x < b) = \int_{a}^{b} f(x) \cdot dx$

$$\frac{1}{\beta - \alpha} = \frac{1}{10 - 0} = \frac{1}{10} \qquad \frac{1}{10} \times 3$$

classmate

Date

Page

eg:2 Consider a scenario where you are waiting for a bus to arrive, you are not sure about the exact arrival time. But you know that previous expriences that a bus arrives at any random time between 12:00 to 12:30. A RV x represents the arrival time. $(12 \le x \le 12{:}30)$.

—4

$$f(x) = \begin{cases} \frac{1}{30}, & 12{:}00 < x < 12{:}30 \\ 0 & \text{otherwise} \end{cases}$$

* What is the probability that the arrival of bus is between 12:10 to 12:20

—4 $PDF = \frac{1}{30}$ $\quad \therefore \frac{1}{30} \times 10 = \frac{1}{3}$

2 The Exponential Random

A continuous RV whose PDF is given for some $\lambda > 0$ by

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{cases}$$

This is considered x the exponential RV with the given parameter λ

eg-1: Consider the scenario where you are modelling a time between phone calls received at a customer service center. Suppose a time between consecutive phone calls follows an exponential distribution with an average of 5 mins. 'x' represents the time between phone calls. 'x' follows exponential

$\star\ \lambda = \frac{1}{\mu}$

distribution of $\mu = 5$. And here the PDF is

$f(x;\lambda) = \lambda e^{-\lambda x}$. Then what is the probability that the time between consecutive phone calls is less than 3 minutes.

→

$\therefore F(x) = \int_0^3 \lambda e^{-\lambda x}$ $\quad \because \lambda = \frac{1}{5} = 0.2$

$= \left(\frac{1}{5}\right) e^{-1/5 \times 3} - \left(\frac{1}{5}\right) e^{0}$

$= \frac{1}{5} e^{-3/5} - \frac{1}{5}$

$= \frac{1}{5}\left(e^{-3/5} - 1\right)$

$= \frac{1}{5}\left(9.96 \times 10^{-03} - 1\right)$

3. Gamma Random Variable

The Gamma distribution is used to model waiting time or duration until a certain event occurs.

The Gamma distribution is categorized by two parameters; shape and scale. Then shape is represented by 'K' and scale is 'θ'. and the PDF of gamma distribution f(x).

$\star\ x \geq 0$

$f(x, K, \theta) = \left(\frac{1}{\gamma(K)\,\theta^K}\right) \cdot x^{K-1} \cdot e^{-x/\theta}$; $\gamma(n) = (n-1)!$

eg-1 The lifetime of an electronic component is given. Suppose the life time of a component follows the gamma distribution with $K = 3$ and $\theta = 1000$ hr. 'X' represents the lifetime of component and 'X' follows the gamma distribution with parameter $K = 3$ & $\theta = 1000$ hr ; $x = 1500$ hr.

So, the PDF

$\Gamma(n) = (n-1)!$

$\Gamma(3) = 2.$

$$f(1500, 3, 1000) = \left(\frac{1}{\Gamma(3)\,(1000)^3}\right) \cdot 1500^{(3-1)} \cdot e^{-\cancel{1500}/\cancel{1000}}$$

$$= \frac{1}{2 \times (1000)^3} \cdot (1500)^2 \cdot e^{-15/10}$$

$$= \frac{1500 \times 1500}{2 \times 1000 \times 1000 \times 1000} \cdot \frac{1}{e^{15/10}}$$

$$= \left(\frac{225}{200000} \cdot \frac{1}{e^{15/10}}\right)$$

Cummulative Distribution Function :-

eg: Consider a rolling of fair die. let x be RV represents outcome of roll. The CDF denoted by $F(x)$ gives the probability that outcome is less than or equal to specific 'x'. let $x = 3$ then, $F(3)$ is the probability of rolling a number less than equal to 3.

$[F(3) \approx P(x \leq 3)] \approx P(1), P(2), P(3)$

Date
Page

* Probability Density Function :-

The PDF denoted by $f(x)$ gives the probability density at particular value of x.

Considering a continuous RV 'x' representing the height of particular individual population, let x follows a normal distribution with the mean = 170 cm & Standard Deviation = 10 cm.

9) Normal Distribution :-

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{——} \quad (9)$$

* Probability Mass Function :

PMF gives the probability that a RV or DRV 'x' takes on specific value.

Joint Distribution :-

The Joint Distribution describes the simultaneous behaviour of two or more RVs.

JD specifies the probabilities of various outcomes occurring together for all the variables involved.

There are two ways to represent the JD, first we are using probability mass Function for DRVector $F(x, y)$ & second PDF for continuous Random Vectors.

Let x & y be some random variables the (x, y) is then called a two dimensional random vectors.

The JD or Joint CDF for the (x, y) is represent by

$$F(x, y) = P(X \le x, Y \le y) \text{ for } x, y \in R. \quad — (5)$$

Discrete Random Vector :-

If x and y are discrete RVs then the (x, y) is called discrete Random vectors.

If (x, y) is discrete with the range $\{(x_j, y_k) : j, k = 1, 2, \dots\}$ then probability mass Function represent by.

$$P(x_j, y_k) = P(X = x_j, Y = y_k) \quad — (6)$$

* Jointly Continuous Random Vectors :-

If there exists a function 'f' for pair of RV (x, y) which follows a continuous distribution then, $f(x, y)$ is defined as

$$P((x, y) \in B) = \iint_B f(x, y) \cdot dx\,dy \quad — (7)$$

The function 'f' is called Joint PDF of two dimensional vector x & y

* Conditional Distribution :-

Let x & y be discrete RVs with range $= \{x_1, x_2, \dots\}$ and $\{y_1, y_2, \dots\}$ then, the conditional Probability mass Function of R variable y given $X = x_j$ is defined as

$$P_y(y_k / x_j) = \frac{P(x_j, y_k)}{P_x(x_j)} \quad — (8)$$

Date
Page

5/4/24

The Central limit Thearem (CLT) and Law of large Numbers (LLN)

* Law Of large Numbers (LLN) :~

The LLN has a central role in probability and Statistics. The law states that if you repeat an experiment independently a large number of times (n) and average the result what we obtain is closure to the expected value of that random variable 'x'

There are two types of LLN : ① Weak LLN ② Strong LLN.

* The central limit Thearem (CLT)

We have the average of multiple measurements of the same unknown quantity that tends to give the better estimate then the single measurement.

This is because the random error of each measurement cancels out the average.

The two components that follows this phenomenon in probability are LLN & CLT.

The CLT states that under some finite conditions, the sum of large number of random variables has an approximately, a normal distribution.

" The average of many independent sample is close to the mean of the underline distribution" — Statement Of LLN.

This density of histogram of many independent samples is close to the graph of the density of the underline distribution.

The CLT says that the sum or average of many independents copies of random variable is approx. the normal random variable. The CLT goes to give precise value for the mean and standard deviation of the normal variable.