MCA Sem. I Core

Mathematical Techniques for Computer Applications(MCAC 103) L 4

Random Variables

14 Jan 2021

Vasudha Bhatnagar
Department of Computer Science,University of Delhi
Delhi, India.

Outline

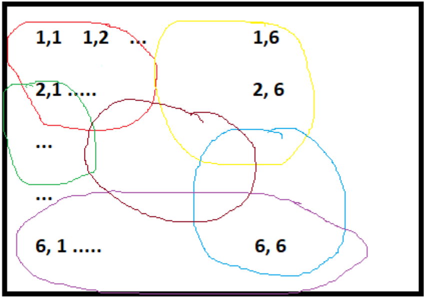
1. **Random Variables**
2. **Probability Distribution**
3. **Probability Distribution Function**
4. **Probability Density Function**
5. **Joint Probability Distribution**
6. **Marginal Probability Distribution**

1,1	1,2	...			1,6
2,1					2, 6
...		4,3	4,4	4,5	4,6
...					
6, 1					6, 6

Sample space for 2-dice rolling experiment
(|S| = 36)

1. **Probability measure assigns probability** $p(o_i)$ **to each outcome** o_i **, such that** $\forall o_i \in S, 0 \leq p(o_i) \leq 1$ **and** $\Sigma p(o_i) = 1$
2. **If all outcomes are mutually exclusive and equaliy likely then** $\forall i, j, p(o_i) = p(o_j) = 1/n$**, where** n **is the size of** S

Multiple events can be mapped to S



Two dice rolling experiment - sample space of size 36

1. **Event E_1: Sum of two numbers greater than 8**
2. **Event E_2: First number bigger than second**
3. **Event E_3: Max of the two numbers is divisible by 2**

1. Event E_1: {4,5; 5,4; 3,6; 6,3,; ... 6,6 }
2. Event E_2: {2,1; 3,2; ....}
3. Event E_3: {1,2; 1, 4; 1, 6; 2, 4, ...}

Suppose we are interested in more general questions about the outcomes of the experiment.

1. What is the sum of the two numbers? {2, ... 12}
2. Which is the bigger of the two numbers? {2, 3, ...6}
3. Sum of the numbers, when one number is divisible by the other (not considering divisibility by 1). { ???}

Define a function that maps all outcomes in S to a set of values

$$\boxed{f : S \longrightarrow R}$$

$\boxed{\text{Several random variables can be defined for a set of outcomes.}}$

1. Experiment of tossing four coins
   X1 counts # of Heads, X2 count# of Tails, X3 denotes the case when all faces are same
2. Experiment of drawing three balls with replacement from an urn containing 4 red and 3 black balls
   X1 counts # of Red balls, X2 denotes the case when # of Red balls is more than black
3. Set of students of MCA 2022 Admission
   X1 counts # students with UG degree from DU, X2 denotes the native state of the student, X3 denotes grade of the student, X4 denotes height of the student, X5 denotes rank of the student in entrance exam

Definitions:

Definition 1: A random variable is a function from a sample space S to the real numbers. Coventionally, random variables are denoted with capital letters, e.g., $X : S \longrightarrow R$

Definition 2: A discrete random variable is a random variable that takes a finite or countably infinite (integers or whole numbers) number of possible values.

Example: X is a rv denoting count of heads in toss of four coins. X takes values 0, 1, 2, 3, 4

*Example:*Y is a rv denoting the # rolls of dice till a six is obtained. Y takes value 0, 1, 2, ...

Definition 3: A continuous random variable is a random variable with infinitely many possible values (in an interval of real numbers).

Example: X is a rv that denotes height of a student in mts. X takes values in interval [1.5, 1.9]

Example: Y is a rv that denotes the weight of a pet dog in kgs. Y takes values in interval [10, 50]

Discrete Vs. Continuous Random Variable

Object	Discrete rv	Continuous rv
Flowers	Color, Size, Smell	Length of petals, Length of sepals, Height of plant
Students	College of UG Native state	Distance of home Height and weight
States	State animal # of Districts # of Universities	GDP Population density Area under forest(Sq. Km)

Probability Distribution

1,1 1,2 ... 1,6

2,1 ..... 2, 6

... 4,3 4,4 4,5 4,6

...

6, 1 ..... 6, 6

Sample space for 2-dice rolling experiment ($|S| = 36$)

1. **Probabilites for random variables can be computed from the probability measure defined on underlying sample space.**
2. **Probability that a discrete random variable takes a specific value is denoted by $P_X(X = x)$**
3. **Probability mass function (PMF) of rv X tells probability for each value that X takes.**

1. **Probabilites for random variables can be computed from the probability measure defined on underlying sample space.**
2. **Probability that a discrete random variable takes a specific value is denoted by $P_X(X = x)$**
3. **Probability mass function (PMF) of rv X tells probability for each value that X takes.**

Experiment: 4 fair coins tossed

Elements of S	**Prob**	**X**
TTTT	**1/16**	**0**
TTTH	**1/16**	**1**
...		
HHHT	**1/16**	**3**
HHHH	**1/16**	**4**

RV - X	0	1	2	3	4
Prob	1/16	4/16	6/16	4/16	1/16

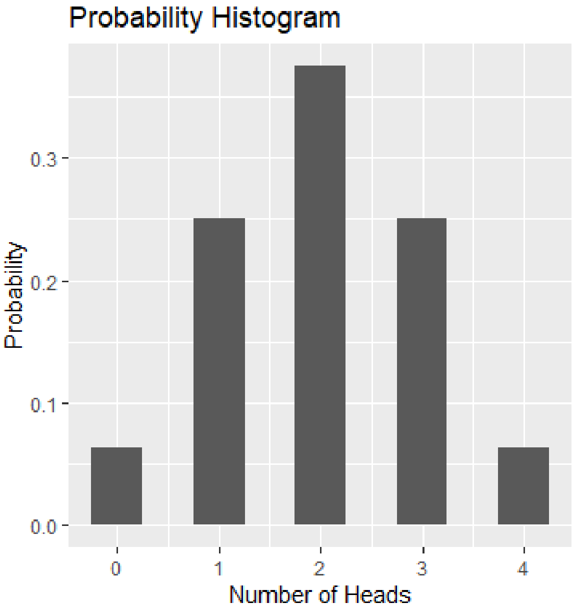
Formally,

Definition: The probability mass function of a discrete random variable assigns probabilities to all possible values of the random variable. If $x_1, x_2, \ldots$ denote the possible values of a random variable X , then the probability mass function is denoted as $p(x_i) = P_X(X = x_i)$,
such that $\forall x_i, 0 \leq p(x_i) \leq 1$ and $\Sigma_x P_X(x) = 1$

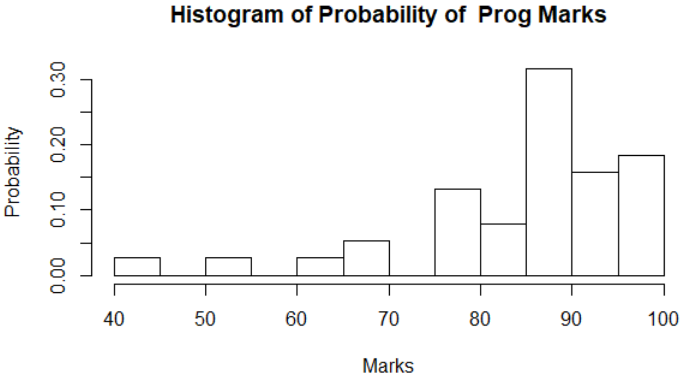
PMF follows axioms of probability.

PMF can be represented numerically as a table, graphically as a histogram, or analytically as a formula

Theorem: A function can serve as a probability distribution of a discrete random variable X iff $p(x)$ satisties following conditions.
i. $p(x) \geq 0$ for each values of X
ii. $\Sigma_x p(x) = 1$


Probability Histogram
Probability
0.0
0.1
0.2
0.3
0
1
2
3
4
Number of Heads

X is a rv that denotes marks in programming. Grouping in class of size 5, probabilities are computed and plotted as hitogram.



Practice Exercise

1. Find the probability distribution for rv X that denotes the product of the numbers that show on rolling of two fair dice.
2. Suppose that we toss a coin having a probability p of coming up heads, until the first head appears. Letting X denote the number of flips required. Assuming that the outcome of successive flips are independent, find out the probability distribution of X.
3. Check whether the probability function $p(x) = \frac{x+2}{25}$ for x= 1, 2, 3, 4, 5 can serve as probability distribution of a discrete random variable
4. Suppose that independent trials, each of which results in any of m possible outcomes with respective probabilities $p_1, ..., p_m, \Sigma_i p_i = 1$, are continually performed. Let X denote the number of trials needed until each outcome has occurred at least once. Find probability distribution of X.

1. **What is the probability that student gets let than pass marks?** $P(X < 40)$
2. **What is the probability that student gets distinction?** $P(X \geq 75)$ **or** $1 - P(X \leq 74)$
3. **What is the probability that I have to wait for not more than 10 minutes?** $P(X \leq 10)$

We are interested in finding probability for an interval

Cumulative Distribution Function (CDF)

Definition: If X is a discrete rv, the CDF of X is computed as follows
$F(x) = P_X(X \leq x) = \Sigma_i^x p(i)$
CDF of a random variable is defined for all real numbers(unlike the pmf of a discrete random variable)

X	**0**	**1**	**2**	**3**	**4**
p(x)	**1/16**	**4/16**	**6/16**	**4/16**	**1/16**
$F(x)$	**1/16**	**5/16**	**11/16**	**15/16**	**16/16**

$$F(x) = \begin{cases} 0 & \textbf{for } x < 0 \\ 1/16 & \textbf{for } 0 \leq x < 1 \\ 5/16 & \textbf{for } 1 \leq x < 2 \\ 11/16 & \textbf{for } 2 \leq x < 3 \\ 15/16 & \textbf{for } 3 \leq x < 4 \\ 16/16 & \textbf{for } x \geq 4 \end{cases}$$

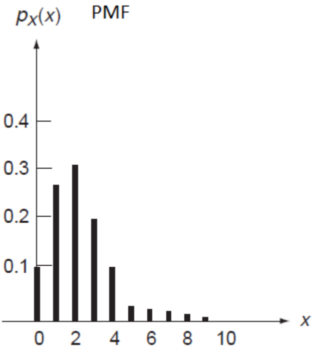
X	0	1	2	3	4
p(x)	**1/16**	**4/16**	**6/16**	**4/16**	**1/16**
$F(x)$	**1/16**	**5/16**	**11/16**	**15/16**	**16/16**

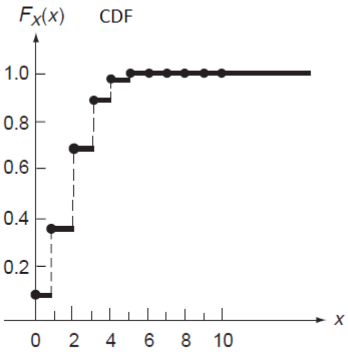
$$F(x) = \begin{cases} 0 & \textbf{for } x < 0 \\ 1/16 & \textbf{for } 0 \le x < 1 \\ 5/16 & \textbf{for } 1 \le x < 2 \\ 11/16 & \textbf{for } 2 \le x < 3 \\ 15/16 & \textbf{for } 3 \le x < 4 \\ 16/16 & \textbf{for } x \ge 4 \end{cases}$$

Theorem: CDF of a discrete rv satisfies the following properties.

1. $F(-\infty) = 0$ and $F(\infty) = 1$
2. If $a < b$, then $F(a) \le F(b)$ for all real numbers a and b

If CDF of a discrete rv is known, its pmf can be computed


$p_X(x)$
PMF
0.4
0.3
0.2
0.1
0
2
4
6
8
10
x


$F_X(x)$
CDF
1.0
0.8
0.6
0.4
0.2
0
2
4
6
8
10
x

Practice Exercise

1. Find the distribution function of a rv with probability distribution $f(x) = x/15, x = 1, 2, 3, 4, 5$
2. Find $p(2 \leq X \leq 4)$, if X is the number of heads when five coins are tossed.
3. Verify that $P(X > x_i) = 1 - F(x_i)$ and $P(X \geq x_i) = 1 - F(x_{i-1})$

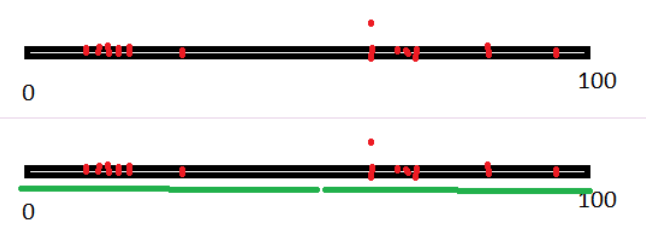
Probability Density Function

Consider an 100 Km long expressway. Count the number of accidents in week, and note locations.

We are interested in finding the probability that accident takes place at a particular stretch

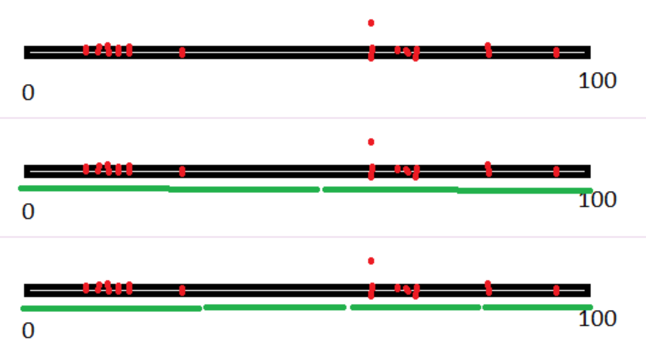
Probability Density Function

Consider an 100 Km long expressway. Count the number of accidents in week, and note locations.
We are interested in finding the probability that accident takes place at a particular stretch



Probability Density Function

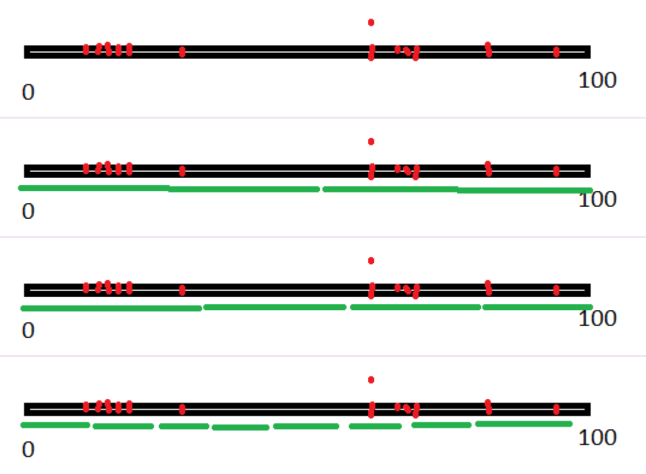
Consider an 100 Km long expressway. Count the number of accidents in week, and note locations.
We are interested in finding the probability that accident takes place at a particular stretch



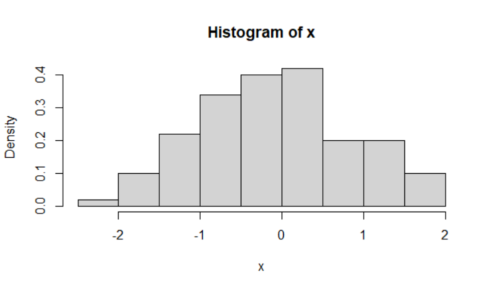
Probability Density Function

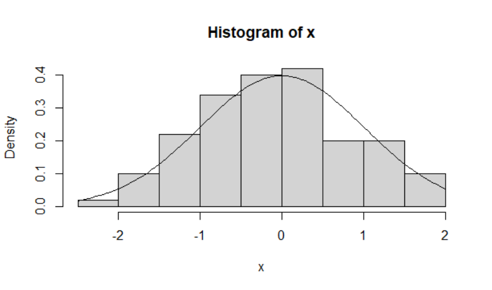
Consider an 100 Km long expressway. Count the number of accidents in week, and note locations.

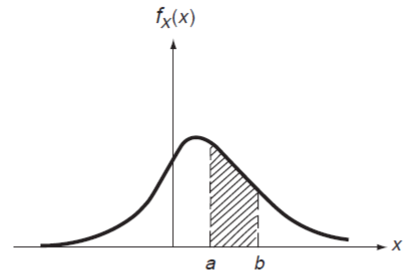
We are interested in finding the probability that accident takes place at a particular stretch

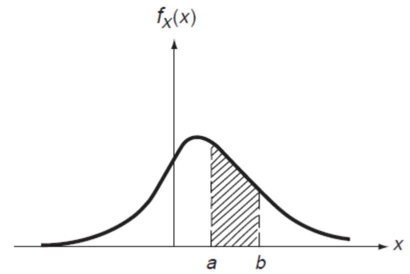


Coffee Vending machine manufactured to dispense 150 ml of coffee : X measures the volume of coffee dispensed.

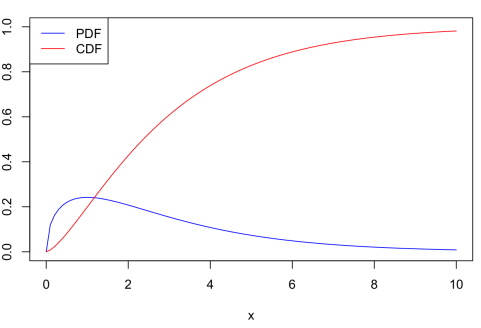





$f_X(x)$
x
a
b



$f_X(x)$
x
a
b



PDF
CDF
0.0
0.2
0.4
0.6
0.8
1.0
0
2
4
6
8
10
x

Definition: A real valued function $f(x)$ is called a probability density function of the continuous random variable X iff

$$P(a \leq X \leq b) = \int_a^b f(x)dx \tag{1}$$

for any real constants a and b with $a \leq b$

1. Theorem: A function can serve as a probability density function of a continuous random variable X if its values $f(x)$ satisfy the conditions
    1. $f(x) \geq 0$ for $-\infty < x < \infty$
    2. $\int_{-\infty}^{\infty} f(x)dx = 1$
2. Theorem: If X is a continuous random variable and a and b are real constants with $a \leq b$, then

$$P(a \leq X \leq b) = P(a < X \leq b) = P(a \leq X < b) = P(a < X < b) \tag{2}$$

(Cumulative) Distribution Function

1. Definition: If X is a continuous random variable with pdf $f(x)$, then the distribution function of X is given by

$$F_X(x) = F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)dt \tag{3}$$

2. Theorem: If $f(x)$ and $F(x)$ are the values of probability density and the distribution function of X at x, then for any real constants a and b with $a \leq b$

$$P(a \leq X \leq b) = F(b) - F(a) \text{ and } f(x) = \frac{dF(x)}{dx} \tag{4}$$

where the derivative of $F(x)$ exists

Practice Exercise

1. If X has the pdf $f(x) = ke^{-3x}$ for $x > 0$, find k.
2. Find distribution function of the above, and find $P(0.5 \leq X \leq 1)$
3. Find pdf of a rv X, whose distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } 0 < x < 1 \\ 1 & \text{for } x \geq 1 \end{cases}$$

Multivariate Data: Multiple random variables associated with single experiment/outcome

Joint probability - Statement of the probability of two or more rvs taking values
Bivariate - $P(X = x, Y = y)$ or $P(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2)$
Multivariate - $P(X_1 = x_1, \ldots, X_k = x_k)$ or
$P(x_{11} \leq X_1 \leq x_{12}, \ldots, x_{k1} \leq X_k \leq x_{k2})$

Joint Probability Distribution
Discrete random variables X and Y, is given by $p_{XY}(X = x, Y = y)$ for each pair of values (x, y) within the range of X and Y

Joint probability density function of continuous random variable X and Y is given by
$f(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{x_i}^{x_2} \int_{y_i}^{y_2} f(x, y) dxdy$

Theorem: A Bivariate function can serve as the joint probability distribution for a pair of discrete random variables X and Y if and only if its value $p(x, y)$ satisfy the conditions:
i. $p(x, y) \geq 0$ **for each pair** (x, y) **within the domain**
ii. $\Sigma_x \Sigma_y p(x, y) = 1$
where double summation extends over all possible pairs (x, y) **within the domain**

Thoerem: A Bivariate function can serve as the joint density function for a pair of continuous random variables X and Y if and only if its value $f(x, y)$ satisfy the conditions:
i. $f(x, y) \geq 0$ for $-\infty < x, y < \infty$
ii. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy = 1$

Joint Distribution Function: Cumulative probability of two or more rvs

If X and Y are discrete random variables, the joint distribution function of X and Y is given by

$$F(x,y) = P(X \leq x, Y \leq y) = \Sigma_s^x \Sigma_t^y p(s,t) \text{ for } -\infty < x, y < \infty \quad (5)$$

where p(s,t) is the value of the joint probability function of X and Y at (s,t)

If X and Y are continuous random variables, the joint distribution function of X and Y is given by

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t) ds dt, \text{ for } -\infty < x, y < \infty \quad (6)$$

where f(s,t) is the value of the joint density of X and Y at (s,t)

Practice Exercise

1. *Find pmf :* Two balls are selected at random from an urn containing 3 Red 2 white and 4 Black balls. if X and Y respectively are the numbers of of red and white balls, find the probabilities associated with all possible pairs of X and Y values (joint pmf of X and Y).
2. *Find if valid pmf :* Determine the value of k for which the function given by $f(x, y) = kxy$, for $x, y = 1, 2, 3$ can serve as a probability distribution.
3. *Find prob from pdf :* Given the joint probability density function

$$f(x,y) = \begin{cases} \frac{3}{5}x(x+y) & \text{for } 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{otherwise} \end{cases}$$

of two random variables X and Y, find $P(X, Y)$ in the region $0 \le x \le 1/2, 1 \le y \le 2$.

Practice Exercise - pdf and DF

1. Find *joint DF from pdf :* If the joint probability density of X and Y is given by

$$f(x,y) = \begin{cases} (x+y) & \text{for } 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0 & \text{elsewhere} \end{cases}$$

   Find the joint distribution function of these random variable.

2. Find *joint pdf from DF :* Find joint pdf of X and Y, whose DF is given as follows:

$$F(x,y) = \begin{cases} (1-e^{-}x)(1-e^{-}y) & \text{for } x, y > 0 \\ 0 & \text{elsewhere} \end{cases}$$

   And determine $P(1 < X < 3, 1 < Y < 2)$

Marginal Probability Distribution

If X and Y are discrete rvs and $f(x, y)$ is the joint probability distribution at (x, y), then $g(x) = \Sigma_y f(x, y)$ is the marginal probability distribution of X for each value of x in range of X. Similarly, $h(y) = \Sigma_x f(x, y)$ is the marginal probability distribution of Y for each value of y in range of Y

If X and Y are continuous rvs and $f(x, y)$ is the joint probability density at (x, y), then $g(x) = \int_{-\infty}^{\infty} f(x, y)dy$ for $-\infty \leq x \leq \infty$ is the marginal density of X.
Similarly, $h(y) = \int_{-\infty}^{\infty} f(x, y)dx$ for $-\infty \leq y \leq \infty 1$ is the marginal density of Y.

Conditional Probability Distributions

$P(A|B) = P(A \cap B)/P(B)$

Conditional Probability Distributions

$P(A|B) = P(A \cap B)/P(B)$

$P(X = x|Y = y) = P(X = x \cap Y = y)/P(Y = y) =$

Conditional Probability Distributions

$P(A|B) = P(A \cap B)/P(B)$

$$P(X = x|Y = y) = P(X = x \cap Y = y)/P(Y = y) = f(x, y)/h(y)$$

Conditional Probability Distributions

$P(A|B) = P(A \cap B)/P(B)$

$$P(X = x|Y = y) = P(X = x \cap Y = y)/P(Y = y) = f(x, y)/h(y)$$

If $f(x, y)$ is the value of the joint probability distribution of the discrete random variables X and Y at (x, y) and $h(y)$ is the value of the marginal distribution of Y at y,

$$f_{X|Y}(x|y) = f(x, y)/h(y), h(y) \neq 0$$

for each x within the range of X, is called the conditional distribution of X given $Y = y$.

Correspondingly, it is possible to define the conditional distribution of Y given $X = x$

$$w_{Y|X}(y|x) = f(x, y)/g(x), g(x) \neq 0$$

where $g(x)$ is the value of the marginal distribution of X at x

Conditional density

If $f(x, y)$ is the value of the joint density of the continuous random variables X and Y at (x, y) and $h(y)$ is the value of the marginal distribution of Y at y,

$$f_{X|Y}(x|y) = f(x, y)/h(y), h(y) \neq 0$$

for each $-\infty \leq x \leq \infty$ is called the conditional density of X given Y.

Practice Exercise -Marginal and Conditional distributions

1. Consider the joint pmf for Q1 on slide 28. Find marginal distributions for all values of X and Y.
2. From the above, find f(0|1), f(1|1), f(2|1)
3. Given the joint pdf

$$f(x, y) = 2/3(x + 2y), \text{ for } 0 < x < 1, 0 < y < 1$$

   find marginal densities of X and Y
4. From the above, find conditional density of X given $Y = y$, and use it to evaluate $P(X \leq 1/2 | Y = 1/2)$