Discrete & Continuous Random Variables

(Probability distributions, PMF, PDF, CDF, JPDF,JPMF)

Random Variable

> A **random variable** is a function that associates a real number with each element in the sample space.

- Two balls are drawn in succession without replacement from an urn containing 4 red balls and 3 black balls. The possible outcomes and the values y of the random variable Y , where Y is the number of red balls, are:
- Suppose an experiment consists of tossing a coin two times & we are interested in the number of Heads (X):
- Interest centers around the proportion of people who respond to a certain mail order solicitation. Let X be that proportion. X is a random variable that takes on all values x for which $0 \leq x \leq 1$.

Discrete & Continuous Random Variable

- A random variable is called a **discrete random variable** if its set of possible outcomes is countable.
- When a random variable can take on values on a continuous scale, it is called a **continuous random variable**

Discrete Probability Distribution

1) Tossing a coin 3 times and X = No. of heads
2) A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Cumulative Distribution Function

The **cumulative distribution function** $F(x)$ of a discrete random variable X with probability distribution $f(x)$ is

$$F(x) = P(X \leq x) = \sum_{t \leq x} f(t), \quad \text{for } -\infty < x < \infty.$$

- Suppose a coin is tossed 3 times and X = No. of heads.

Example # 01

- Consider the following pmf: $f(x) = (x/6)$, $x = 1, 2, 3$, zero elsewhere

(i) Find distribution function and its graph.

(ii) Calculate $P(1.5 < x \leq 4.5)$

Example # 02

- Find P (x = 2) = f(2)

$$F(0) = f(0) = \frac{1}{16},$$
$$F(1) = f(0) + f(1) = \frac{5}{16},$$
$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16},$$
$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16},$$
$$F(4) = f(0) + f(1) + f(2) + f(3) + f(4) = 1.$$

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

Example # 03 & 04

(3) Consider the following functions

(i) $f(x) = (x + 2) / 5$ for $x = 1, 2, 3, 4, 5$.

(ii) $f(x) = (4Cx) / (2^5)$, for $x = 0, 1, 2, 3$, and 4.

and check whether the functions can serve as a pmf ?

(4) A coin is biased so that a head occurs 3 times of tail. If the coin is tossed 3 times, find the probability distribution for the number of heads and also find $P(1 \le x \le 3)$.

Example # 05 & 06

- If two dice are rolled once, find the pmf of the sum of points on two dice and also find c. d. f (also its graph).
- A fair coin is tossed until a “Head” appears for the first time. Find

    (a) p. m. f

    (b) distribution function

    (c) F(4)

Example # 07

- The distribution function for a discrete random variable x is given as:

$F(x) = 1 – (1/2)^{(x+1)}$, for $x = 0, 1, 2, \ldots..$

Find:

(a) $P (X = 3)$

(b) $P (7 \leq x < 10)$

(c) Probability Mass Function.

Example # 08

3.5 Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:

(a) $f(x) = c(x^2 + 4)$, for $x = 0, 1, 2, 3$;

Joint Probability Distribution

The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum_A f(x, y)$.

Example # 09

- Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find
- (a) the joint probability function $f(x, y)$
- (b) $P[(X, Y) \in A]$, where A is the region $\{(x, y) | y \leq 1\}$

The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_{y} f(x, y) \quad \text{and} \quad h(y) = \sum_{x} f(x, y)$$

Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

$$f(y|x) = \frac{f(x,y)}{g(x)}, \text{ provided } g(x) > 0.$$

Similarly, the conditional distribution of X given that $Y = y$ is

$$f(x|y) = \frac{f(x,y)}{h(y)}, \text{ provided } h(y) > 0.$$

Example 09 (Contd.)

- find the conditional distribution of X, given that $Y = 1$, and use it to determine $P(X = 0 \mid Y = 1)$.

$f(x,y)$		$x=0$	$x=1$	$x=2$	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Statistical independence

The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

- Show that the random variables of Example 09 are not statistically independent for the point (0, 1).

Example # 10

If the joint probability distribution of X and Y is given by

$$f(x, y) = \frac{x+y}{30}, \quad \text{for } x = 0, 1, 2, 3;\ y = 0, 1, 2,$$

find

(a) $P(X \leq 2, Y = 1)$;
(b) $P(X > 2, Y \leq 1)$;
(c) $P(X > Y)$;
(d) $P(X + Y = 4)$.

$y \backslash x$	0	1	2	3
0	0	1/30	2/30	3/30
1	1/30	2/30	3/30	4/30
2	2/30	3/30	4/30	5/30

Let X denote the number of times a certain numerical control machine will malfunction: 1, 2, or 3 times on any given day. Let Y denote the number of times a technician is called on an emergency call. Their joint probability distribution is given as

$f(x,y)$		x		
		1	2	3
	1	0.05	0.05	0.10
	3	0.05	0.10	0.35
y	5	0.00	0.20	0.10

(a) Evaluate the marginal distribution of X.

(b) Evaluate the marginal distribution of Y.

(c) Find $P(Y = 3 \mid X = 2)$.

Let the number of phone calls received by a switchboard during a 5-minute interval be a random variable X with probability function

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

(a) Determine the probability that X equals 0, 1, 2, 3, 4, 5, and 6.

(b) Graph the probability mass function for these values of x.

(c) Determine the cumulative distribution function for these values of X.

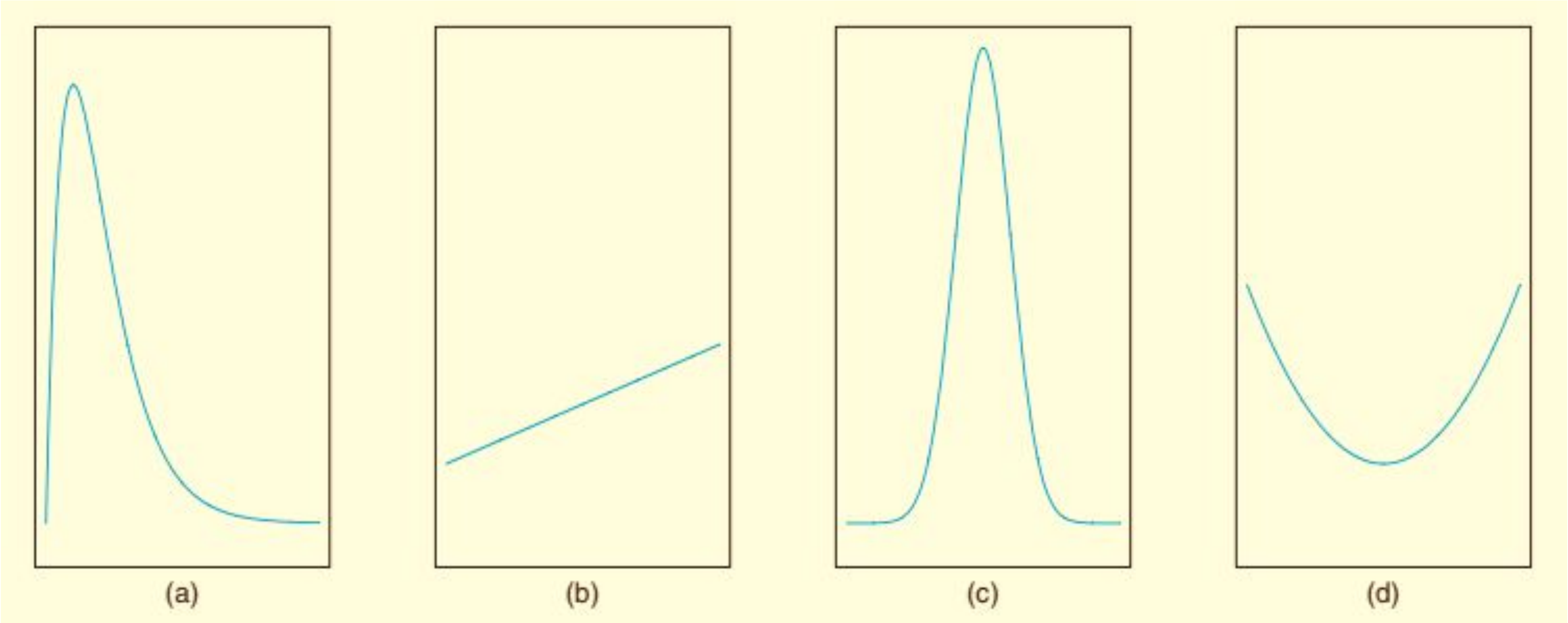
Continuous Probability Distributions

- A continuous random variable has a probability of 0 of assuming *exactly* any of its values.
- Because we are dealing with an interval rather than a point value of our random variable.
- It does not matter whether we include an endpoint of the interval or not. For example:

$$P(a < X \le b) = P(a < X < b) + P(X = b) = P(a < X < b).$$

- Its probability distribution cannot be given in tabular form.
- *f*(*x*) is usually called the **probability density function**.
- Areas will be used to represent probabilities.

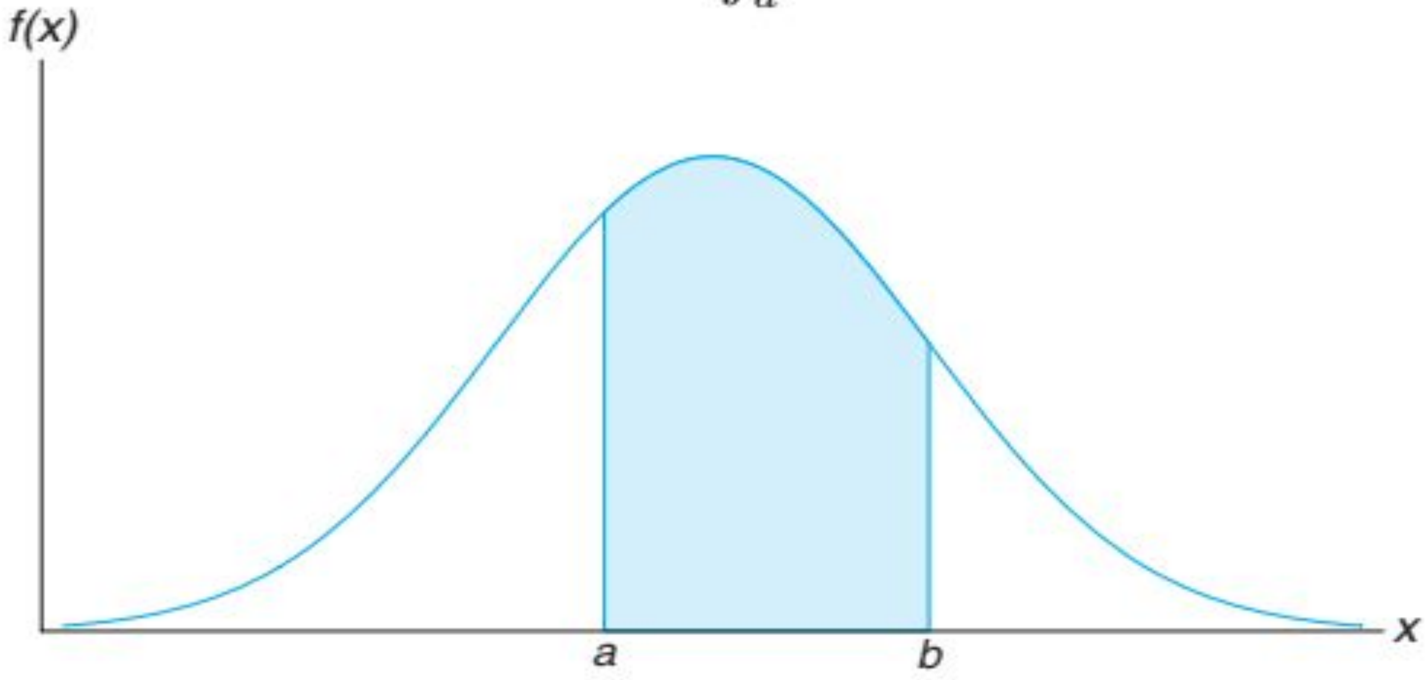
Some graphs of Density Functions



- **A probability density function is constructed so that the area under its curve bounded by the *x* axis is equal to 1 when computed over the range of *X*.**

The probability that *X* assumes a value between *a* and *b* is equal to the shaded area

$$P(a < X < b) = \int_a^b f(x)\,dx.$$



Function to be PDF

The function $f(x)$ is a **probability density function** (pdf) for the continuous random variable X, defined over the set of real numbers, if

1. $f(x) \geq 0$, for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_{a}^{b} f(x)\, dx$.

Example # 01

Suppose that the error in the reaction temperature, in °C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

.

(a) Verify that $f(x)$ is a density function.

(b) Find $P(0 < X \leq 1)$.

CDF for Continuous RV

The **cumulative distribution function** $F(x)$ of a continuous random variable X with density function $f(x)$ is

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

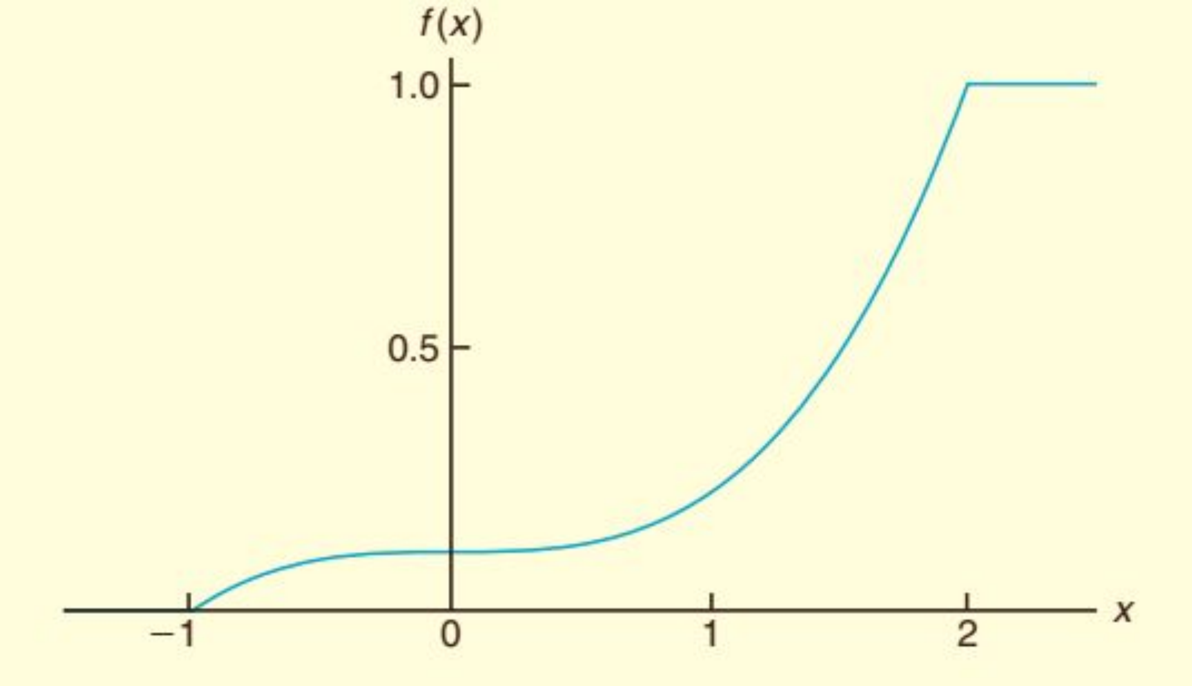
- As an immediate consequence of the above Definition, one can write the two results:

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

Example # 02

- For the density function of previous example , find $F(x)$, and use it to evaluate $P(0 < X \leq 1)$.

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$



Example # 03

- The Department of Energy (DOE) puts projects out on bid and generally estimates what a reasonable bid should be. Call the estimate *b*. The DOE has determined that the density function of the winning (low) bid is:

$$f(y) = \begin{cases} \frac{5}{8b}, & \frac{2}{5}b \leq y \leq 2b, \\ 0, & \text{elsewhere.} \end{cases}$$

- Find *F*(*y*) and use it to determine the probability that the winning bid is less than the DOE's preliminary estimate *b*.

Example # 04

The shelf life, in days, for bottles of a certain prescribed medicine is a random variable having the density function

$$f(x) = \begin{cases} \frac{20{,}000}{(x+100)^3}, & x > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that a bottle of this medicine will have a shell life of

(a) at least 200 days;

(b) anywhere from 80 to 120 days.

Example # 05

The total number of hours, measured in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a continuous random variable X that has the density function

$$f(x) = \begin{cases} x, & 0 < x < 1, \\ 2 - x, & 1 \le x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that over a period of one year, a family runs their vacuum cleaner

(a) less than 120 hours;

(b) between 50 and 100 hours.

Example # 06

The waiting time, in hours, between successive speeders spotted by a radar unit is a continuous random variable with cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0, \\ 1 - e^{-8x}, & x \geq 0. \end{cases}$$

Find the probability of waiting less than 12 minutes between successive speeders

(a) using the cumulative distribution function of X;

(b) using the probability density function of X.

Example # 07

Measurements of scientific systems are always subject to variation, some more than others. There are many structures for measurement error, and statisticians spend a great deal of time modeling these errors. Suppose the measurement error X of a certain physical quantity is decided by the density function

$$f(x) = \begin{cases} k(3 - x^2), & -1 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine k that renders $f(x)$ a valid density function.

(b) Find the probability that a random error in measurement is less than $1/2$.

(c) For this particular measurement, it is undesirable if the *magnitude* of the error (i.e., $|x|$) exceeds 0.8. What is the probability that this occurs?

Example # 08

An important factor in solid missile fuel is the particle size distribution. Significant problems occur if the particle sizes are too large. From production data in the past, it has been determined that the particle size (in micrometers) distribution is characterized by

$$f(x) = \begin{cases} 3x^{-4}, & x > 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify that this is a valid density function.

(b) Evaluate $F(x)$.

(c) What is the probability that a random particle from the manufactured fuel exceeds 4 micrometers?

Example # 09

The proportion of people who respond to a certain mail-order solicitation is a continuous random variable X that has the density function

$$f(x) = \begin{cases} \frac{2(x+2)}{5}, & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Show that $P(0 < X < 1) = 1$.

(b) Find the probability that more than 1/4 but fewer than 1/2 of the people contacted will respond to this type of solicitation.

Example 10

A continuous random variable X that can assume values between $x = 1$ and $x = 3$ has a density function given by $f(x) = 1/2$.

(a) Show that the area under the curve is equal to 1.

(b) Find $P(2 < X < 2.5)$.

(c) Find $P(X \leq 1.6)$.

Joint Density Function for Continuous variable

The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y)\, dx\, dy$, for any region A in the xy plane.

Example # 09

- A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let *X* and *Y* , respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x, y) = \begin{cases} \frac{2}{5}(2x + 3y), & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- (a) Verify for PDF?
  (b) Find $P[(X, Y) \in A]$, where $A = \{(x, y) \mid 0 < x < 1/2, 1/4 < y < 1/2\}$.

Marginal distribution for Continuous Variable

The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_y f(x,y) \quad \text{and} \quad h(y) = \sum_x f(x,y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x,y)\,dx$$

for the continuous case.

Example # 11

The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$f(x, y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal densities $g(x)$, $h(y)$, and the conditional density $f(y|x)$.

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

(c) Calculate: E(x), E(Y), E(xy)

- **take limit of x (0 to y) & limit of y (0 to 1) for verification of JPDF**

Example # 12

Given the joint density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere}, \end{cases}$$

find $g(x)$, $h(y)$, $f(x|y)$, and evaluate $P(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3})$.

Example # 10

- Find $g(x)$ and $h(y)$ for the joint density function of previous Example?

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

Example # 13

The amount of kerosene, in thousands of liters, in a tank at the beginning of any day is a random amount Y from which a random amount X is sold during that day. Suppose that the tank is not resupplied during the day so that $x \leq y$, and assume that the joint density function of these variables is

$$f(x,y) = \begin{cases} 2, & 0 < x \leq y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Determine if X and Y are independent.

(b) Find $P(1/4 < X < 1/2 \mid Y = 3/4)$.

3.39 From a sack of fruit containing 3 oranges, 2 apples, and 3 bananas, a random sample of 4 pieces of fruit is selected. If X is the number of oranges and Y is the number of apples in the sample, find

(a) the joint probability distribution of X and Y;

(b) $P[(X, Y) \in A]$, where A is the region that is given by $\{(x, y) \mid x + y \leq 2\}$.

3.40 A fast-food restaurant operates both a drive-through facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-through and walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x, y) = \begin{cases} \frac{2}{3}(x + 2y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal density of X.

(b) Find the marginal density of Y.

(c) Find the probability that the drive-through facility is busy less than one-half of the time.

Example # 14

Let X, Y, and Z have the joint probability density function

$$f(x, y, z) = \begin{cases} kxy^2 z, & 0 < x, y < 1,\ 0 < z < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find k.

(b) Find $P(X < \frac{1}{4}, Y > \frac{1}{2}, 1 < Z < 2)$.

- G(x,y), G(x,z), G(y,z), G(x), E(x) =?

Example # 16

A tobacco company produces blends of tobacco, with each blend containing various proportions of Turkish, domestic, and other tobaccos. The proportions of Turkish and domestic in a blend are random variables with joint density function (X = Turkish and Y = domestic)

$$f(x,y) = \begin{cases} 24xy, & 0 \le x, y \le 1, \; x + y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the probability that in a given box the Turkish tobacco accounts for over half the blend.

(b) Find the marginal density function for the proportion of the domestic tobacco.

(c) Find the probability that the proportion of Turkish tobacco is less than 1/8 if it is known that the blend contains 3/4 domestic tobacco.

Example # 17

- A salesperson for a medical device company has two appointments on a given day. At the first appointment, he believes that he has a 70% chance to make the deal, from which he can earn $1000 commission if successful. On the other hand, he thinks he only has a 40% chance to make the deal at the second appointment, from which, if successful, he can make $1500. What is his expected commission based on his own probability belief? Assume that the appointment results are independent of each other.

Example # 18

- Let *X* and *Y* be the random variables with joint probability distribution indicated in Table below. Calculate E(XY).
- Also calculate E(XY), E(x), E(y), V(x), V(y), Covariance (X, Y), correlation.

	$f(x,y)$	x = 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Example # 19

Find $E(Y/X)$ for the density function

$$f(x,y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

- Also calculate E(XY), E(x), E(y), V(x), V(y), Covariance (X, Y), Correlation.