Random Variables

Statistics is concerned with making inferences about population & population characteristics. Experiments are conducted with results that are subject to chance. The testing of a no. of electronic components is an example of statistical experiment; a term that is used to describe any process by which several chance observations are generated.

It is often important to allocate a numerical description to the outcome. for eg: the sample space giving a detailed description of each possible outcome when three electronic equipments are tested, may be written as

$$S = \{ \overset{0}{NNN}, \overset{1}{NND}, NDN, DNN, \overset{2}{NDD}, DND, \overset{2}{DDN}, \overset{3}{DDD} \}$$

where N denotes nondefective & D denotes defective.

One is naturally concerned with the no. of defectives that occurs. Thus each point in the sample space will be assigned a numerical value 0, 1, 2 or 3. These values are random quantities determined by the outcome of the experiment and are the values assumed by the random variable X, the no. of defective items when three electronic equipment are tested.

Def! A random variable is a function that associates a real no. with each element in the sample space.

X → random variable

& its corresponding small letter x, for one of its values.

In above eg: $X = 2$ for all elts in the subset $E = \{DDN, DND, NDD\}$ of i.e; each possible value of X represents an event that is a subset of the sample space for the given experiment.

Eg 1 Two balls are drawn in succession w/o replacement from an urn containing 4 red balls & 3 black balls. The possible outcomes & the values y of the random variable Y, where y is the no. of red balls, are

discrete r.v

Sample space	y
RR	2
RB	1
BR	1
BB	0

Ex 2 A stockroom clerk returns three safety helmets at random to three steel mill employees who had previously checked them. If Smith, Jones & Brown, in that order, receive one of the three helmets, list the sample pts for the possible orders of returning the helmets, & find the value m of the random variable M that represents the no. of correct matches.

S J B

Sample Space	m
SJB	3
SBJ	1
BJS	1
JSB	1
JBS	0
BSJ	0

In Ex 1 & 2, the sample space contains a finite no. of elements.

On the other hand, when a die is thrown until a 5 occurs, we obtain a sample space with an unending sequence of elements

$$S = \{ F, NF, NNF, NNNF, \ldots \}$$

where F & N represent, resp, the occurrence & non occurrence of a 5.

Ex 3 Components are arriving from the production line & they are stipulated to be defective or not defective
r.v is binary

Define $X = \begin{cases} 1, & \text{if the component is defective} \\ 0, & \text{if the component is not defective} \end{cases}$

The random variable for which 0 or 1 are chosen to describe the two possible values is called a Bernoulli r.v.

Ex 4 Proportion of people who respond to a certain mail order solicitation.
cts r.v

Let X be that proportion. X is a r.v. that takes on all values x for which $0 \le x \le 1$.

Ex 5 Let X be a r.v. defined by the waiting time in hrs, b/w ~~two~~ successive speeders spotted by a radar unit. r.v. X takes on all values x for which $x \ge 0$.
cts r.v.

Def 2 If a sample space contains a finite no. of possibilities with as many elements as there are whole nos., it is called a discrete sample space.

The outcomes of some statistical experiments may be n. finite nor countable.

for eg: when one conducts an investigation measuring the distances that a certain make of automobile will travel over a prescribed test course on 5 ltrs of gasoline. There are infinite no of outcomes in the sample space.

Def3 If the sample space contains an infinite no. of possibilities equal to the no. of pts on a line segment, it is called a continuous sample space.

A r.v. is called discrete r.v. if its set of possible outcomes is countable.

But a r.v. whose set of possible values is an entire interval of nos. is not discrete. When a r.v can take on values on a continuous scale, it is called a cts. r.v.

Cts r.v. represent measured data such as all possible heights, weights, temperature, distance or life periods whereas

discrete r.v. represent count data such as the no. of defectives in a sample of k items or the no. of highway fatalities per year in a given state.

The prob. distribution of a r.v. X is a description of the probabilities associated with the possible values of X.

Discrete Probability distributions

- A discrete r.v assumes each of its values with a certain probability

HHT, HTH, THH

In the case of tossing a coin three times, the variable X representing the no of heads, assumes the value 2 with probability 3/8, since 3 of the 8 equally likely sample points results in 2 heads and one tail.

- If one assumes equal weights for the sample events in Ex 2, the probability that no employee gets back the right helmet, i.e., the probability that M assumes the value 0 is 1/3. The possible values m of M and their probabilities are

m	0	1	3
$P(M=m)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{1}{6}$

the value of m exhaust all possible cases & hence the probability add to 1.

It is convenient to represent all the probabilities of a r.v X by a formula. Such a formula is a function of x, denoted by $f(x)$ etc. $\therefore$ $f(x) = P(X=x)$

The set of ordered pairs $(x, f(x))$ is called the probability function, prob. mass function or prob. distribution of a discrete r.v. X.

Prob. fn → pmf/pdf

dist fn → CDF

Def. 4 The set of ordered pairs $(x, f(x))$ is a prob. prob. mass fn or prob distribution of a discrete if, for each possible outcome x,

(i) $f(x) \geq 0$

(ii) $\sum_x f(x) = 1$

(iii) $P(X = x) = f(x)$

Ex 6 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 prob. distribution for a no. of defectives.

- Let X be a random variable whose values x are the possible no. of defectives computers purchased by the school.
  x can only take the nos 0, 1 & 2.

$$f(0) = P(X=0) = \frac{{}^3C_0 \; {}^{17}C_2}{{}^{20}C_2} = \frac{68}{95}$$

$$f(1) = P(X=1) = \frac{{}^3C_1 \; {}^{17}C_1}{{}^{20}C_2} = \frac{51}{190}$$

$$f(2) = P(X=2) = \frac{{}^3C_2 \; {}^{17}C_0}{{}^{20}C_2} = \frac{3}{190}$$

Thus, the prob. distribution of X is

x	0	1	2
$f(x)$	$\frac{68}{95}$	$\frac{51}{190}$	$\frac{3}{190}$

There are many problems where we may wish to compute the prob that the observed value of a r.v. X will be less than or equal to some real no. x.

$$F(x) = P(X \leq x) \quad \text{for every real no. } x$$

$F(x)$ is defined as cumulative distribution function of a r.v X.

Def 5 The cumulative distribution function $F(x)$ of a discrete r.v. X with prob. distribution $f(x)$ is

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

In helmet eg: for a r.v. M, the no. of correct matches,

$$F(2) = P(M \leq 2) = f(0) + f(1) + f(2)$$
$$= \frac{1}{3} + \frac{1}{2} + 0 = \frac{5}{6}$$

The cumulative distribution function of M is

$$F(m) = \begin{cases} 0 & \text{for } m < 0 \\ \frac{1}{3} & \text{,, } 0 \leq m < 1 \\ \frac{5}{6} & 1 \leq m < 3 \\ 1 & m \geq 3 \end{cases}$$

Note that the cumulative dist. fn is a montone nondecreasing function defined not only for the values assumed by the given r.v. but for all real nos.

Ex 7 If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for a prob distribution of the no. of cars with side airbags among the next 4 cars sold by the agency.

Also, find the cumulative distribution fn. of the r.v.

Using $F(x)$, verify that $f(2) = 3/8$

- Prob of selling an automobile with side airbags is 0. ... is

$2^4 = 16$ pts in the sample space are equally likely to occur.

The event of selling x models with side airbags and $4-x$ models without side airbags can occur in ${}^4C_x$ ways where x can be 0, 1, 2, 3 or 4.

Thus the prob. distribution $f(x) = P(X=x)$ is

$$f(x) = \frac{1}{16}\, {}^4C_x, \quad x = 0, 1, 2 \ldots 4$$

$\therefore f(0) = \frac{1}{16}, \quad f(1) = \frac{1}{4}, \quad f(2) = \frac{1}{16}\, {}^4C_2 = \cancel{\frac{6}{16}}\ \frac{3}{8}.$

$f(3) = \frac{1}{4}, \quad f(4) = \frac{1}{16}$

$$\therefore F(0) = f(0) = \frac{1}{16}$$

$$F(1) = f(0) + f(1) = \frac{5}{16}$$

$$F(2) = f(0) + f(1) + f(2) = \frac{11}{16}$$

$$F(3) = f(0) + f(1) + f(2) + f(3) = \frac{15}{16}$$

$$F(4) = \quad " \quad\quad " \quad + f(4) = 1$$

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

$\frac{4!}{2!\,2!} = \frac{4 \cdot 3}{2}$

$\sim \quad f(2) = F(2) - F(1) = \frac{11}{16} - \frac{5}{16} = \frac{3}{8}$

It is helpful to look at a prob. distribution in graphic form. One might plot the pts $(x, f(x))$.
By joining the pts to the ~~axes~~ x-axis either with a dashed or with a solid line, we obtain pmf plot.
Fig 1 makes it easy to see what values of X are most likely to occur and it also indicates a perfectly symmetric situation in this case.

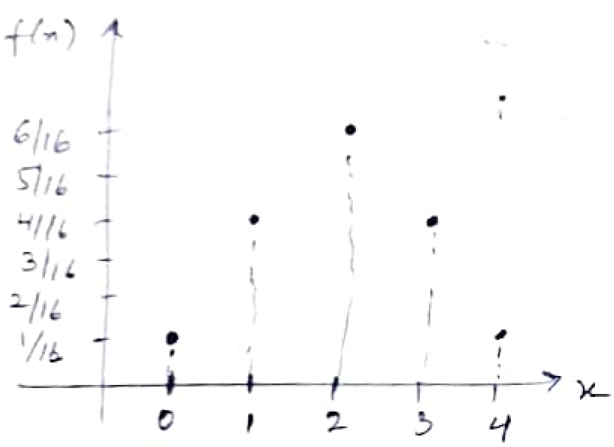


Fig 1 Pmf plot

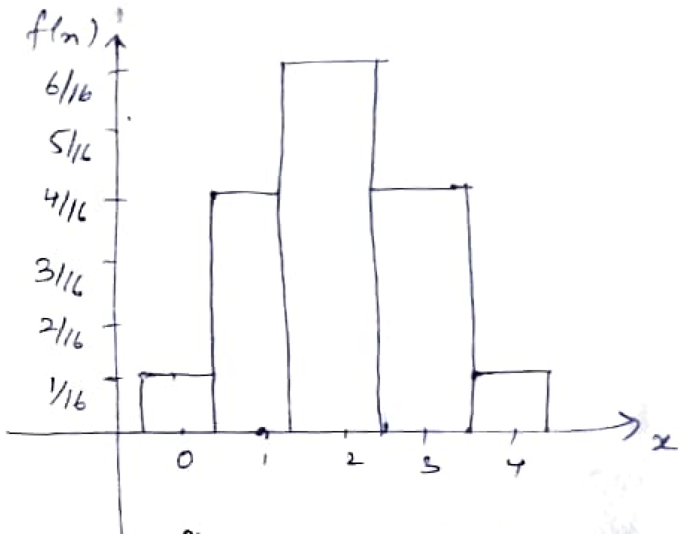


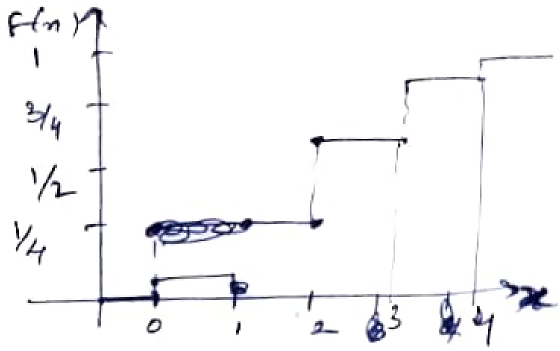
Fig 2 Prob. Histogram

Instead of plotting the pts $(x, f(x))$, we can construct rectangles as in fig 2. Here the rectangles are constructed so that their bases of equal width are centered at each value x and their heights are equal to the corresponding prob.-ies given by $f(x)$. The bases are constructed so as to leave no space b/w the rectangles. fig 2 is called a prob. histogram.

Since each base has unit width, $P(X=x)$ is equal the area of the rectangle centered at x.

This concept of using areas to represent probabilities is necessary for our consideration of the prob. distribution of a continuous r.v.

The graph of cumulative dist fn. of Ex , appears as a step function in Fig , is obtained by plotting the pts $(x, F(x))$.



Certain prob. distributions are applicable to more than one physical situation.

Read P107

Continuous Probability distribution

A continuous r.v. has a probability of 0 of assuming exactly one of its values. ∴ its probability distribution cannot be given in tabular form.

for eg: A r.v. whose values are the heights of all people over 21 years of age. Between any two values say 163.5 & 164.5 cm or even 163.99 and 164.01 cms, there are an infinite no. of heights, one of which is 164 cm.

The prob of selecting a person at random who is exactly 164 cm tall & not one of the infinitely large sets of heights so close to 164 cms that you cannot humanly measure the difference is remote, & thus we assign a prob of 0 to the event.

This is not the case, however, if talk of prob of selecting a person who is atleast 163 cm but not more than 165 cms tall. interval not point of a r.v.

$$P(a < X \leq b) = P(a < X < b) + P(X = b)$$
$$= P(a < X < b)$$

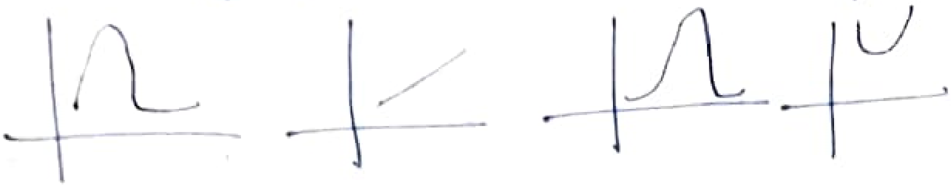
ie. it does not matter whether we include an endpoint of the interval or not.

The prob. dist. of a cts. r.v can be stated by a formula.

$f(x)$ is called the prob. density function or density fn, of, X.

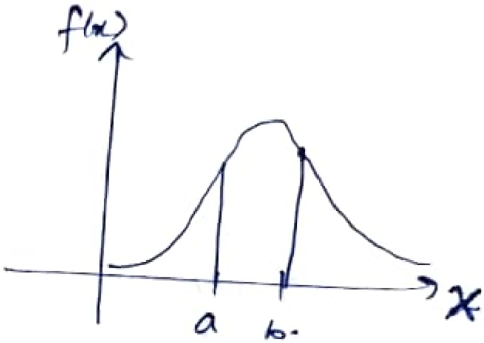
Since X is defined over a cts. sample space, it is possible for $f(x)$ to have a finite no. of discontinuities. verify find

Since areas will be used to represent probabilities and probabilities are positive ~~real no~~ numerical values, the density function must lie entirely above the x-axis



A prob. 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 for which $f(x)$ is defined.

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



Def 6 The function $f(x)$ is a prob. density function (pdf) for the continuous r.v. X, defined over the set of real nos, if

(i) $f(x) \geq 0, \quad x \in \mathbb{R}$

(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

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

Ex 8 Suppose that the error in the reaction temp in °C for a controlled lab experiment is a cts r.v. X having the pdf

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

(i) Verify that $f(x)$ is a density fn
(ii) find $P(0 < X \leq 1)$

- Obviously $f(x) \geq 0$

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-1}^{2} \frac{x^2}{3}\,dx = \frac{x^3}{9}\Big|_{-1}^{2} = \frac{8}{9} + \frac{1}{9} = 1$$

$$P(0 < X \leq 1) = \int_0^1 f(x)\,dx = \int_0^1 \frac{x^2}{3}\,dx = \frac{1}{9}$$

Def 7 The cumulative distribution function $F(x)$ of a continuous r.v. X with density function $f(x)$ is

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

$\therefore$ $P(a < X < b) = F(b) - F(a)$ and $f(x) = \frac{d}{dx}F(x)$ if the derivative exist.

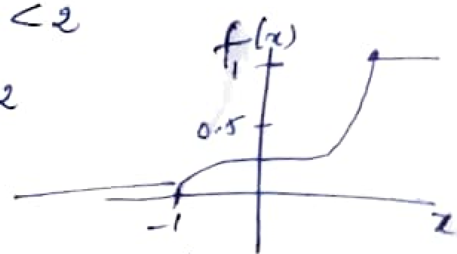
Ex 9 For the density fn of Ex 8, find $F(x)$ & use it to evaluate $P(0 < X < 1)$

- for $-1 < x < 2$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{-1}^{x} \frac{t^2}{3}\,dt = \frac{t^3}{9}\Big|_{-1}^{x}$$

$$= \frac{x^3 + 1}{9}$$

$$\therefore F(x) = \begin{cases} 0 & x < -1 \\ (x^3+1)/9 & -1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$



The cumulative dist fn $F(x)$

$$P(0 < X \leq 1) = F(1) - F(0) = \frac{2}{9} - \frac{1}{9} = \frac{1}{9}.$$

operties of CDF $F(x)$

① $F(x)$ is a non-decreasing function of x ie. if $x_1 < x_2$ then $F(x_1) \le F(x_2)$

② $F(-\infty)=0$ and $F(\infty)=1$

③ If X is a discrete r.v. taking values $x_1, x_2, \ldots$, where $x_1 < x_2 < x_3 < \ldots < x_{i-1} < x_i < \ldots$

then $P(X=x_i) = F(x_i) - F(x_{i-1})$

④ If X is a continuous r.v., then $\frac{d}{dx} F(x) = f(x)$, at all pts where $F(x)$ is differentiable.

Ex1 If the r.v. takes the values 1, 2, 3 and 4 s.t.

$2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4)$

find the prob. distribution and the cumulative distribution function of X.

let $2P(X=1) = 3P(X=2) = P(X=3) = 5P(X=4) = k$

i.e. $P(X=1) = k/2$, $P(X=2) = k/3$,

$P(X=3) = k$, $P(X=4) = k/5$

$\sum_{i} p_i = 1 \Rightarrow \frac{k}{2} + \frac{k}{3} + k + \frac{k}{5} = 1$

$15k + 10k + 30k + 6k = 30$

$k = 30/61$

Probability distribution of X is given by

x_i	1	2	3	4
$P(x_i)$	15/61	10/61	30/61	6/61
$F(x)$	15/61	25/61	55/61	61/61 = 1

CDF $F(x) = P(X \le x)$

when, $x < 1$ $\quad F(x) = 0$

when $1 \le x < 2$, $\quad F(x) = 15/61$

$2 \le x < 3$, $\quad F(x) = 25/61$

$3 \le x < 4$, $\quad F(x) = 55/61$

$x \ge 4$, $\quad F(x) = 1$

Ex 2 A r.v. X has the following probability function

Values of X	0	1	2	3	4	5	6	7	8
$P(X=x)$	a	3a	5a	7a	9a	11a	13a	15a	17a

(i) find the value of a.

(ii) find $P(X<3)$, $P(0<X<3)$, $P(X \ge 3)$

(iii) find the distribution function of X.

Soln (i) $\sum_i p_i = 1$

$a + 3a + 5a + \dots + 17a = 1$

$81a = 1 \Rightarrow a = 1/81$

(ii) $P(X<3) = P(X=0) + P(X=1) + P(X=2)$

$a + 3a + 5a = 9a = \frac{9}{81} = \frac{1}{9}$

(iii) $P(0<X<3) = P(X=1) + P(X=2) = 3a + 5a = 8a = \frac{8}{81}$

(iv) $P(X \ge 3) = 1 - P(X<3) = 1 - \frac{1}{9} = 8/9$

(V) Distribution function of X, F(x)

X: 0 1 2 3 4 5 6 7 8 ~~9~~

a 4a 9a 16a 25a 36a 49a 64a 81a

$F(X) = P(X \le x) = \frac{1}{81}, \frac{4}{81}, \frac{9}{81}, \frac{16}{81}, \frac{25}{81}, \frac{36}{81}, \frac{49}{81}, \frac{64}{81}, 1$

$f(x) = 6x(1-x), \quad 0 \le x \le 1$

Check if $f(x)$ is probability density function.

Soln for pdf $\int_{-\infty}^{\infty} f(x) = 1$

$$\int_{-\infty}^{0} f(x)\,dx + \int_0^1 f(x)\,dx + \int_1^{\infty} f(x)\,dx$$

$$\int_0^1 6x(1-x)\,dx$$

$$\left. 6\frac{x^2}{2} - \frac{6x^3}{3} \right)_0^1$$

$f(x) = 6x(1-x) \ge 0$ in $[0,1]$

$3 - 2 = 1$

$\Rightarrow f(x)$ is pdf.

Ex 4 find the value of k for the pdf

$$f(x) = \begin{cases} kx^2, & 0 \le x \le 3 \\ 0, & \text{o.w} \end{cases}$$

Also compute $P(1 \le x \le 2)$ and the distribution function.

Soln Since $f(x)$ is a pdf,

$$\int_{-\infty}^{\infty} f(x) = 1 \quad \text{or} \quad k\int_0^3 x^2\,dx = 1$$

$$\left. k\frac{x^3}{3} \right)_0^3 = 1 \quad \text{or} \quad k = 1/9$$

$$P(1 \le x \le 2) = \int_1^2 f(x)\,dx = \frac{1}{9}\int_1^2 x^2\,dx = \left. \frac{1}{9}\frac{x^3}{3} \right)_1^2$$

$$= \frac{1}{27}(8-1) = \frac{7}{27}$$

$$F(x) = \int_{-\infty}^{x} f(x)\,dx$$

$P(X \leq x)$

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{x} f(x)\,dx$$

$$k\int_{0}^{x} x^2\,dx = k\frac{x^3}{3}\Big)_0^x = \frac{1}{9} \times \frac{x^3}{3} = \frac{x^3}{27}$$

$$F(x) = \begin{cases} 0 & x < 0 \\ x^3/27 & 0 \leq x \leq 3 \\ 1 & x > 3 \end{cases}$$

Ex 5 If the density function of a contenuous r.v. is given by

$$f(x) = \begin{cases} ax & 0 \leq x \leq 1 \\ a & 1 \leq x \leq 2 \\ 3a - ax & 2 \leq x \leq 3 \\ 0 & e.w \end{cases}$$

(i) find the value of 'a'.

(ii) find the cdf of X.

Soln Since $f(x)$ is a pdf,

$$\int_{-\infty}^{\infty} f(x) = 1$$

$$\int_0^1 ax\,dx + a\int_1^2 dx + \int_2^3 (3a - ax)\,dx = 1$$

$$\frac{a}{2} + a + a\left(3x - \frac{x^2}{2}\right)_2^3 = 1$$

$$\frac{3a}{2} + a\left(9 - \frac{9}{2} - 6 + 2\right) = 1 \quad \Rightarrow \quad a = 1/2$$

cdf of x

$F(x) = P(X \le x) = 0$ when $x < 0$

$F(x) = \int_0^x f(x)\,dx$, $0 \le x < 1$

$= \int_0^x \frac{x}{2}\,dx = \frac{x^2}{4}$, $0 < x < 1$

when $1 \le x \le 2$

$F(x) = \int_0^1 \frac{x}{2}\,dx + \int_1^x \frac{1}{2}\,dx = \frac{x}{2} - 1/4$

when $2 \le x < 3$

$F(x) = \int_0^1 \frac{x}{2}\,dx + \int_1^2 \frac{1}{2}\,dx + \int_2^x \left(\frac{3}{2} - \frac{x}{2}\right) dx$

$= \frac{3}{2}x - \frac{x^2}{4} - \frac{5}{4}$

for $x > 3$, $F(x) = 1$

Ex6 Suppose that X is a cts r.v. whose probability density function is given by

$$f(x) = \begin{cases} c(4x - 2x^2), & 0 < x < 2 \\ 0, & \text{o.w} \end{cases}$$

(i) What is the value of c?

(ii) find $P(X > 1)$.

Soln

$\int_{-\infty}^{\infty} f(x) = 1 \Rightarrow c\int_0^2 (4x - 2x^2)\,dx = 1$

$c\,\frac{8}{3} = 1$ or $c = 3/8$

$P(X > 1) = \int_1^{\infty} f(x)\,dx = \frac{3}{8}\int_1^2 (4x - 2x^2)\,dx = 1/2$