

“JUST THE MATHS”

SLIDES NUMBER

19.3

**PROBABILITY 3
(Random variable)**

by

A.J.Hobson

19.3.1 Defining random variables

**19.3.2 Probability distribution and
probability density functions**

UNIT 19.3 - RANDOM VARIABLES

19.3.1 DEFINING RANDOM VARIABLES

(i) The theory of probability usually discusses **“random experiments”**.

For example, in throwing an unbiased die, it is just as likely to show one face as any other.

Similarly, drawing 6 lottery numbers out of 45 is a random experiment provided it is just as likely for one number to be drawn as any other.

An experiment is a random experiment if there is more than one possible outcome (or event) and any one of those possible outcomes may occur.

We assume that the outcomes are **“mutually exclusive”**.

The probabilities of the possible outcomes of a random experiment form a collection called the **“probability distribution”** of the experiment.

These probabilities need not be the same as one another.

The complete list of possible outcomes is called the **“sample space”** of the experiment.

(ii) In a random experiment, each outcome may be associated with a numerical value called a **“random variable”**

This variable, x , makes it possible to refer to an outcome without having to use a complete description of it.

In tossing a coin, we might associate a head with the number 1 and a tail with the number 0.

Then the probabilities of either a head or a tail are given by the formulae

$$P(x = 0) = 0.5 \quad \text{and} \quad P(x = 1) = 0.5$$

Note:

There is no restriction on the way we define the values of x ; it would have been just as correct to associate a head with -1 and a tail with 1 .

It is customary to assign the values of random variables as logically as possible.

For example, in discussing the probability that two 6's would be obtained in 5 throws of a dice, we could sensibly use $x = 1, 2, 3, 4, 5$ and 6 respectively for the results that a 1,2,3,4,5 and 6 would be thrown.

(iii) It is necessary to distinguish between “**discrete**” and “**continuous**” random variables.

Discrete random variables may take only certain specified values.

Continuous random variables may take any value within a certain specified range.

Examples of discrete random variables include those associated with the tossing of coins, the throwing of dice and numbers of defective components in a batch from a production line.

Examples of continuous variables include those associated with persons’ height or weight, lifetimes of manufactured components and rainfall figures.

Note:

For a random variable, x , the associated probabilities form a function, $P(x)$, called the “**probability function**”

19.3.2 PROBABILITY DISTRIBUTION AND PROBABILITY DENSITY FUNCTIONS

(a) Probability Distribution Functions

A “**probability distribution function**”, denoted here by $F(x)$, is the relationship between a random variable, x , and the probability of obtaining any outcome for which the random variable can take values up to and including x .

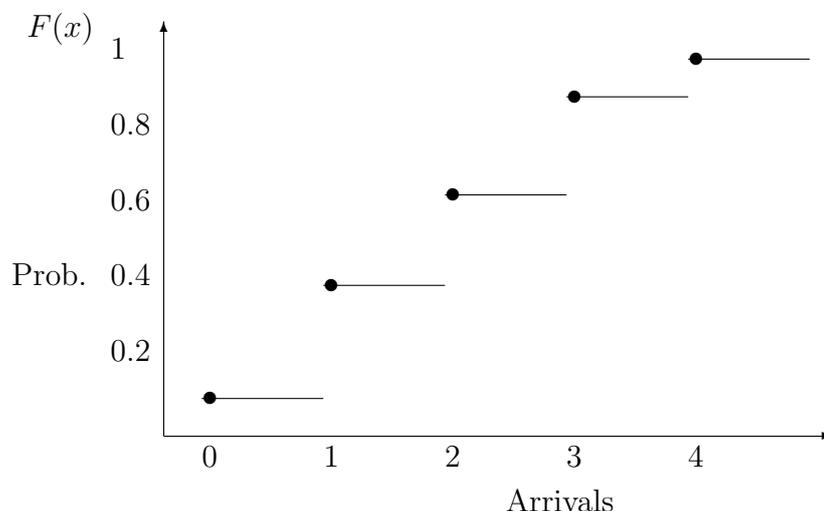
It is the probability, $P(\leq x)$, that the random variable for the outcome is less than or equal to x .

(i) Probability distribution functions for discrete variables

Suppose that the number of ships arriving at a container terminal during any one day can be 0,1,2,3 or 4, with respective probabilities 0.1, 0.3, 0.35, 0.2 and 0.05

The probabilities for other outcomes is taken to be zero.

The graph of the probability distribution function is as follows:



The value of the probability distribution function at a value, x , of the random variable is the sum of the probabilities to the left of, and including, x .

The “bullet” marks to indicate which end of each horizontal line belongs to the graph.

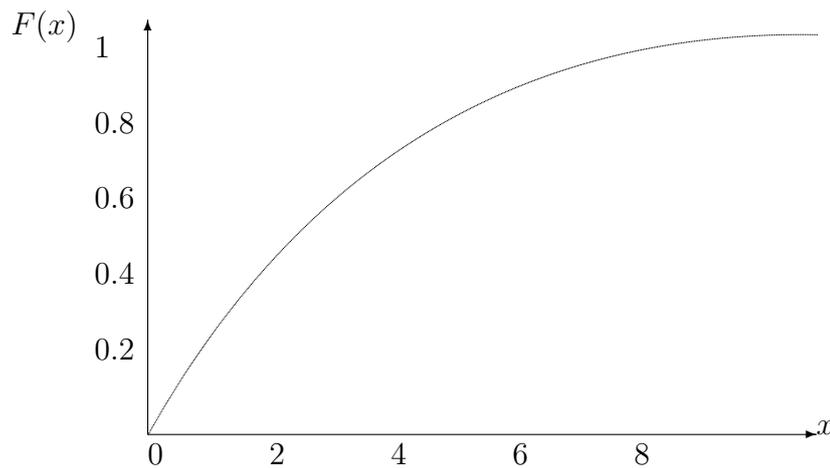
(ii) Probability distribution functions for continuous variables

For a continuous random variable, the probability distribution function is also measures the probability that the value of the random variable is less than or equal to x .

We illustrate with the an **“exponential distribution”**, in which the lifetime of a certain electronic component (in thousands of hours) is represented by a probability distribution function.

$$F(x) \equiv \begin{cases} 1 - e^{-\frac{x}{2}} & \text{when } x \geq 0; \\ 0 & \text{when } x < 0. \end{cases}$$

The graph of the probability distribution function is as follows:



(b) Probability Density Functions

For continuous random variables, the **“probability density function”**, $f(x)$, is defined by

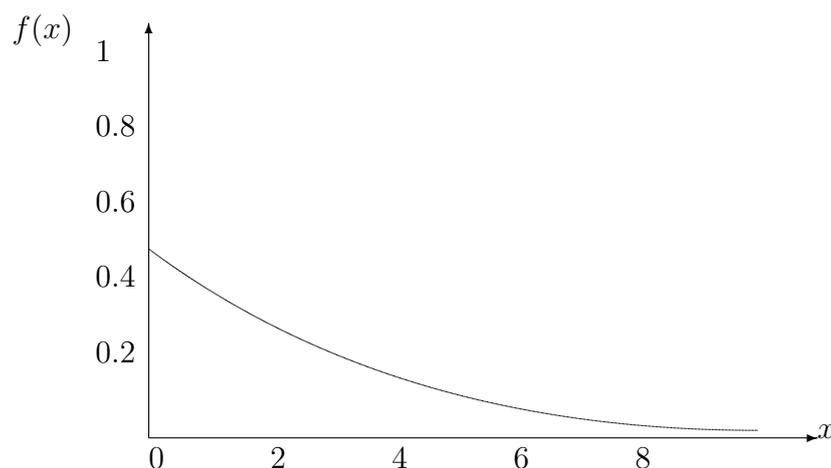
$$f(x) \equiv \frac{d}{dx}[F(x)].$$

The probability density function measures the **concentration** of possible values of x .

In the previous example, the probability density function is given by

$$f(x) \equiv \begin{cases} \frac{1}{2}e^{-\frac{x}{2}} & \text{when } x \geq 0; \\ 0 & \text{when } x < 0 \end{cases}$$

The graph of the probability density function is as follows:



Here, most components have short lifetimes, while a small number can survive for much longer.

(c) Properties of probability distribution and probability density functions

(i)

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{and} \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

Proof:

It is impossible for a random variable to have a value less than $-\infty$ and it is certain to have a value less than ∞ .

(ii) If $x_1 < x_2$, then $F(x_1) \leq F(x_2)$.

Proof:

The outcomes of an experiment with random variable values up to and including x_2 includes those outcomes with random variable values up to and including x_1 so that $F(x_2)$ is at least as great as $F(x_1)$.

Note:

Results (i) and (ii) imply that, for any value of x , the probability distribution function is either constant or increasing between 0 and 1.

(iii) The probability that an outcome will have a random variable value, x , within the range $x_1 < x \leq x_2$ is given by the expression

$$F(x_2) - F(x_1)$$

Proof:

From, the outcomes of an experiment with random variable values up to and including x_2 , suppose we exclude those outcomes with random variable values up to and including x_1 .

The residue will be those outcomes with random variable values which lie within the range $x_1 < x \leq x_2$.

Note:

For a continuous random variable, this is the area under the graph of the probability density function between the two given points because $f(x) \equiv \frac{d}{dx}[F(x)]$.

That is,

$$\int_{x_1}^{x_2} f(x) \, dx.$$

(iv)

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

Proof:

The total area under the probability density function must be 1 since the random variable must have a value somewhere.

EXAMPLE

For the distribution of component lifetimes (in thousands of hours) given by

$$F(x) \equiv \begin{cases} 1 - e^{-\frac{x}{2}} & \text{when } x \geq 0; \\ 0 & \text{when } x < 0, \end{cases}$$

determine the proportion of components which last longer than 3000 hours but less than or equal to 6000 hours.

Solution

The probability that components have lifetimes up to and including 3000 hours is $F(3) = 1 - e^{-\frac{3}{2}}$.

The probability that components have lifetimes up to and including 600 hours is $F(6) = 1 - e^{-\frac{6}{2}}$.

The probability that components last longer than 3000 hours but less than or equal to 6000 hours is

$$F(6) - F(3) = e^{-\frac{3}{2}} - e^{-3} \simeq 0.173$$