BACHELOR OF SCIENCE IN INFORMATION TECHNOLOGY

KIBABII UNIVERSITY

STA 205

Probability and Statistics

Lecture notes

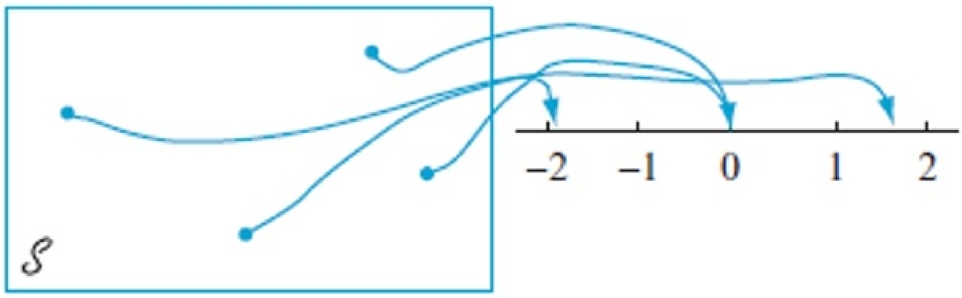
THE CONCEPT OF A RANDOM VARIABLE

Introduction

Whether an experiment yields qualitative or quantitative outcomes, methods of statistical analysis require that we focus on certain numerical aspects of the data (such as a sample proportion $\frac{x}{n}$, mean $\bar{x}$, or standard deviation s). The concept of a random variable allows us to pass from the experimental outcomes themselves to a numerical function of the outcomes. There are two fundamentally different types of random variables — discrete random variables and continuous random variables.

Random Variable

Each outcome of an experiment can be associated with a number by specifying a rule of association. Such a rule of association is called a random variable – a variable because different numerical values are possible and random because the observed value depends on which of the possible experimental outcomes results. The figure below depicts a random variable.



A random variable is a way of mapping outcomes of random processes to numbers (quantify outcomes) e.g. Let X be the outcome when you toss a coin,

$$X = \begin{cases} 1, & \text{if heads} \\ 0, & \text{if tails} \end{cases}$$

or let Y be the sum of the uppermost faces when two die are rolled.

When you quantify outcomes you can do more mathematics on the outcomes and equally more mathematical notations on the outcome. The probability that the sum of the uppermost faces is less than or equal to 12 is denoted as $p(Y \leqslant 12)$.

Capital letters, X, are used to denote the random variables, while small letters x, denote a particular value (or realized value) of the random variable.

For a given sample space S of some experiment, a random variable is any rule that associates a number with each outcome in S. In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers. Random variables are denoted using capital letters e.g. X while the realized value (or a particular value) of the random variable is denoted using small letters e.g. x.

Notation

- $X(s) = x$ means that x is the value associated with the outcome s by the random variable X.
- $p(x) = p(X = x)$ means the probability that the random variable X is equal to the particular value x.

Example 1

When a student attempts to connect to a university computer system, either there is a failure (F), or there is a success (S). With $S = \{S, F\}$ define an random variable X by $X(S) = 1$, $X(F) = 0$. The random variable X indicates whether (1) or not (0) the student can connect.

Sometimes the possible values are many and would be tedious to list; hence a description as in Example 2 can be used.

Example 2

Consider the experiment in which a telephone number in a certain area code is dialed using a random number dialer and define a random variable Y by

$$Y = \begin{cases} 1, & \text{if the selected number is unlisted} \\ 0, & \text{if the selected number is listed in a directory} \end{cases}$$

For example, if 5282966 appears in the telephone directory, then $Y(5282966) = 0$, whereas $Y(7727350) = 1$ tells us that the number 7727350 is unlisted. A word description of this sort is more economical than a complete listing, so we will use such a description whenever possible.

In Examples 1 and 2, the only possible values of the random variable were 0 and 1. Such a random variable arises frequently enough to be given a special name, after the individual who first studied it.

Definition

Any random variable whose only possible values are 0 and 1 is called a Bernoulli random variable.

Class Exercise

Three fair coins are tossed. Let X = the number of heads counted. List each outcome in the sample space along with the associated value of X.

TWO TYPES OF RANDOM VARIABLES

In your MAS 101: Descriptive Statistics Course you distinguished between data resulting from observations on a counting (discrete) variable and data obtained by observing values of a measurement (continuous) variable. A slightly more formal distinction characterizes two different types of random variables.

A discrete random variable is a random variable whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on.

A random variable is continuous if both of the following apply:

i) Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g. from $-\infty$ to $+\infty$) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).

ii) No possible value of the variable has positive probability, that is, $p(X = c) = 0$ for any possible value c.

NB

1. To study basic properties of discrete random variables, only the tools of discrete mathematics - summation and differences - are required.
2. The study of continuous variables requires the continuous mathematics of the calculus - integrals and derivatives.

Class Exercise

For each random variable defined here, describe the set of possible values for the variable, and state whether the variable is discrete or continuous.

a) X = the number of unbroken eggs in a randomly chosen standard egg carton.

b) Y = the number of students on a class list for a particular course who are absent on the first day of classes.

c) U = the number of times a duffer has to swing at a golf ball before hitting it.

d) X = the length of a randomly selected rattlesnake.

e) Z = the amount of royalties earned from the sale of a first edition of 10,000 textbooks.

f) Y = the pH of a randomly chosen soil sample.

g) X = the tension (psi) at which a randomly selected tennis racket has been strung.

h) X = the total number of coin tosses required for three individuals to obtain a match (HHH or TTT).

PROBABILITY FUNCTIONS

A DISCRETE RANDOM VARIABLE

The probability distribution of a random variable X says how the total probability of 1 is distributed among (allocated to) the various possible values of the random variable X.

Example 1

Six lots of components are ready to be shipped by a supplier. The number of defective components in each lot is as follows:

Lot	1	2	3	4	5	6
Number of Defectives	0	2	0	1	2	0

Determine the probability distribution of X.

Solution

Let X be the number of defectives in the selected lot. The three possible X values are 0, 1, and 2. Of the six equally likely simple events, three result in $X = 0$, one in $X = 1$ and the other two in $X = 2$. Let $p(x) = p(X = x)$ then $p(0) = \frac{1}{2}$,$p(1) = \frac{1}{6}$ and $p(2) = \frac{2}{6}$. The values of X along with their probabilities collectively specify the probability distribution or probability mass function of X.

The probability mass function can be specified in a table:

x	0	1	2
$p(x) = p(X = x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{2}{6}$

or as a function

$$p(x) = p(X = x) = \begin{cases} \frac{1}{2}, & x = 0 \\ \frac{1}{6}, & x = 1 \\ \frac{2}{6}, & x = 2 \end{cases}$$

This means that if this experiment were repeated over and over again, in the long run $X = 0$ would occur one-half of the time, $X = 1$ one-sixth of the time, and $X = 2$ one-third of the time.

Properties of probability mass functions

A function is a probability mass function if it satisfies the following conditions

1. $p(x) \geqslant 0)$; probability exists

2. $\sum_{\forall x} p(x) = 1$; total probability $= 1$

Class Exercise 1

1) A tetrahedron die has the numbers 1, 2, 3 and 4 on its faces. The die is biased in a way that the probability of the die landing on any number x is $\frac{k}{x}$ where k is a constant. Find the probability mass function of X, the number the die lands on after a single roll.

2) A discrete random variable X has probability mass function

x	0	1	2
$p(x) = p(X = x)$	$\frac{1}{4} - a$	a	$\frac{7}{12} + a$

Determine a.

3) The random variable Y has probability function

$$p(y) = \begin{cases} ky, & y = 1, 3 \\ k(y-1), & y = 2, 4 \\ 0, & \text{otherwise} \end{cases}$$

Find k and the probability mass function of Y.

4) A discrete random variable X has probability mass function

x	1	2	3	4	5	6
$p(x)$	0.1	0.2	0.3	0.25	0.1	0.05

Find $p(1 < X < 5)$; $p(2 \leqslant X \leqslant 4)$; $p(3 < X \leqslant 6)$ and $p(X \leqslant 3)$.

You must have realized that for a discrete random variable

$$p(X \leqslant a) \neq p(X < a) \quad \text{unless} \quad p(X = a) = 0$$

Cumulative Distribution Function

The cumulative distribution function (cdf) $F(x)$ of a discrete random variable X with probability mass function $p(x)$ is defined for every number x by

$$\begin{aligned} F(x) &= p(X \leqslant x) \\ &= \sum_{y:y \leqslant x} p(y) \end{aligned}$$

For any number x, $F(x)$ is the probability that the observed value of X will be at most x.

Example 2

A store carries flash drives with either 1, 2, 4, 8, or 16 GB of memory. The accompanying table gives the distribution of Y = the amount of memory in a purchased drive:

y	1	2	4	8	16
$p(y)$	0.05	0.10	0.35	0.40	0.10

Determine the cumulative distribution function, $F(y)$, for the five possible values of Y.

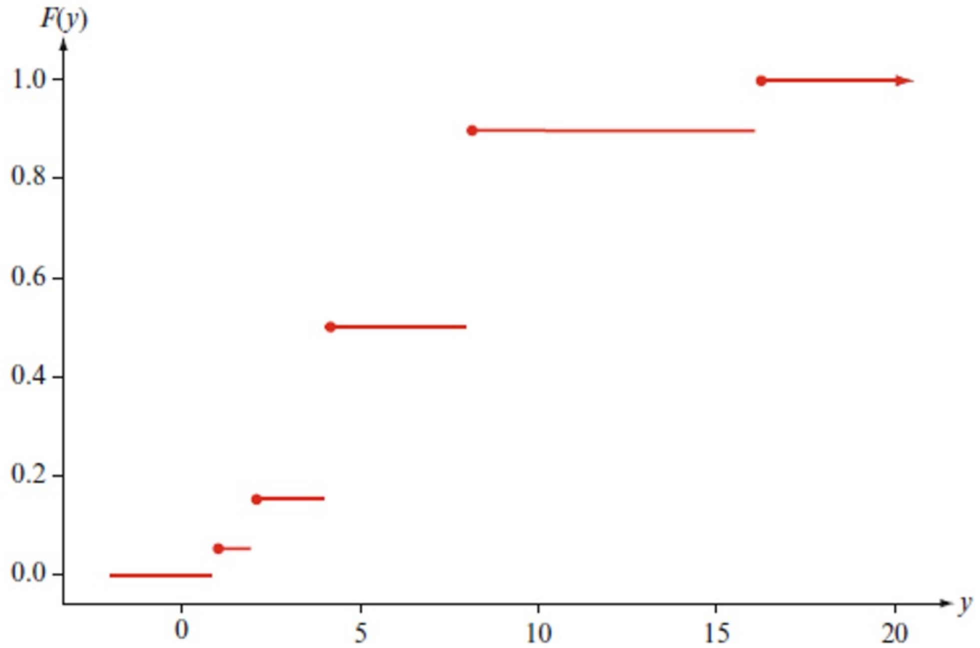
Solution

y	1	2	4	8	16
$p(y)$	0.05	0.10	0.35	0.40	0.10
$F(y)$	0.05	0.15	0.5	0.9	1

or expressed as a function

$$F(y) = \begin{cases} 0, & y < 1 \\ 0.05, & 1 \leqslant y < 2 \\ 0.15, & 2 \leqslant y < 4 \\ 0.5, & 4 \leqslant y < 8 \\ 0.9, & 8 \leqslant y < 16 \\ 1, & 16 \leqslant y \end{cases}$$

The graph of the cumulative distribution function is given by: For X a discrete random variable,



the graph of $F(x)$ will have a jump at every possible value of X and will be flat between possible values. Such a graph is called a step function.

For any other number y, $F(y)$ will equal the value of F at the closest possible value of Y to the left of y.

For example,

$F(2.7) = p(Y \leqslant 2.7) = p(Y \leqslant 2) = F(2) = 0.15.$

$F(7.999) = p(Y \leqslant 7.999) = p(Y \leqslant 4) = F(4) = 0.5.$

$F(25) = p(Y \leqslant 25) = p(Y \leqslant 16) = F(16) = 1.$

Class Exercise 2

1) Any positive integer is a possible X value, and the probability mass function is

$$p(x) = \begin{cases} (1-p)^{x-1}, & x = 1, 2, 3, \cdots \\ 0, & \text{otherwise} \end{cases}$$

Determine the cumulative distribution function of X.

2) The discrete random variable X has cumulative distribution function $F(x)$ defined by

$$F(x) = \begin{cases} \dfrac{x+k}{8}, & x = 1, 2, 3. \\ 0, & \text{otherwise} \end{cases}$$

i) Find the value of k.

ii) Draw the distribution table for the cumulative distribution function of X.

iii) Find $F(2.6)$.

iv) Find $p(x)$ the probability distribution of X.

3) The discrete random variable X has cumulative distribution function $F(x)$ defined by

$$F(x) = \begin{cases} 0, & x = 0. \\ \dfrac{1+x}{6}, & x = 1, 2, 3, 4, 5. \\ 1, & x > 5 \end{cases}$$

i) Find $F(4)$.

ii) Show that $p(4) = \frac{1}{6}$.

iii) Find $p(x)$.

PROPERTIES OF CDFs (for a discrete random variable)

Suppose X represents the numbers of defective components in a shipment consisting of six components, so that possible X values are $0, 1, ..., 6$. Then

$$\begin{aligned} p(3) = p(X = 3) &= [p(0) + p(1) + p(2) + p(3)] - [p(0) + p(1) + p(2)] \\ &= p(X \leqslant 3) - p(X \leqslant 2) \\ &= F(3) - F(2) \end{aligned}$$

More generally, the probability that X falls in a specified interval is easily obtained from the cumulative distribution function. For example,

$$\begin{aligned} p(2 \leqslant X \leqslant 4) &= p(2) + p(3) + p(4) \\ &= [p(0) + p(1) + p(2) + p(3) + p(4)] - [p(0) + p(1)] \\ &= p(X \leqslant 4) - p(X \leqslant 1) \\ &= F(4) - F(1) \end{aligned}$$

Notice that $p(2 \leqslant X \leqslant 4) \neq F(4) - F(2)$. This is because the X value 2 is included in $p(2 \leqslant X \leqslant 4)$, so we do not want to subtract out its probability. However,

$$p(2 < X \leqslant 4) = F(4) - F(2)$$

because $X = 2$ is not included in the interval $(2 < X \leqslant 4)$.

For any two numbers a and b with $a \leqslant b$,

$$p(a \leqslant X \leqslant b) = F(b) - F(a-)$$

$F(a-)$ represents the maximum of $F(x)$ values to the left of a.

Equivalently, if a is the limit of values of x approaching from the left, then $F(a-)$ is the limiting value of $F(x)$. In particular, if the only possible values are integers and if a and b are integers, then

$$\begin{aligned} p(a \leqslant X \leqslant b) &= p(X = a \quad \text{or} \quad a+1 \quad \text{or} \ldots \text{or} \quad b) \\ &= F(b) - F(a-1) \end{aligned} \tag{1}$$

Taking $a = b$ in (1) yields $p(X = a) = F(a) - F(a-1)$ in this case. The reason for subtracting $F(a-)$ rather than $F(a)$ is that we want to include $p(X = a)$; $F(b) - F(a)$ gives $p(a < X \leqslant b)$. This proposition will be used extensively when computing binomial and Poisson probabilities.

PROBABILITY FUNCTIONS

A CONTINUOUS RANDOM VARIABLE

As mentioned earlier, the two important types of random variables are discrete and continuous. In this sub-topic, we study the continuous random variable that arises in many applied problems.

A random variable X is continuous if

1. possible values comprise either a single interval on the number line (for some $A < B$, any number x between A and B is a possible value) or a union of disjoint intervals, and

2. $p(X = c) = 0$ for any number c that is a possible value of X.

Example 1

If in the study of the ecology of a lake, we make depth measurements at randomly chosen locations, then $X =$ the depth at such a location is a continuous random variable. Here A is the minimum depth in the region being sampled, and B is the maximum depth.

Example 2

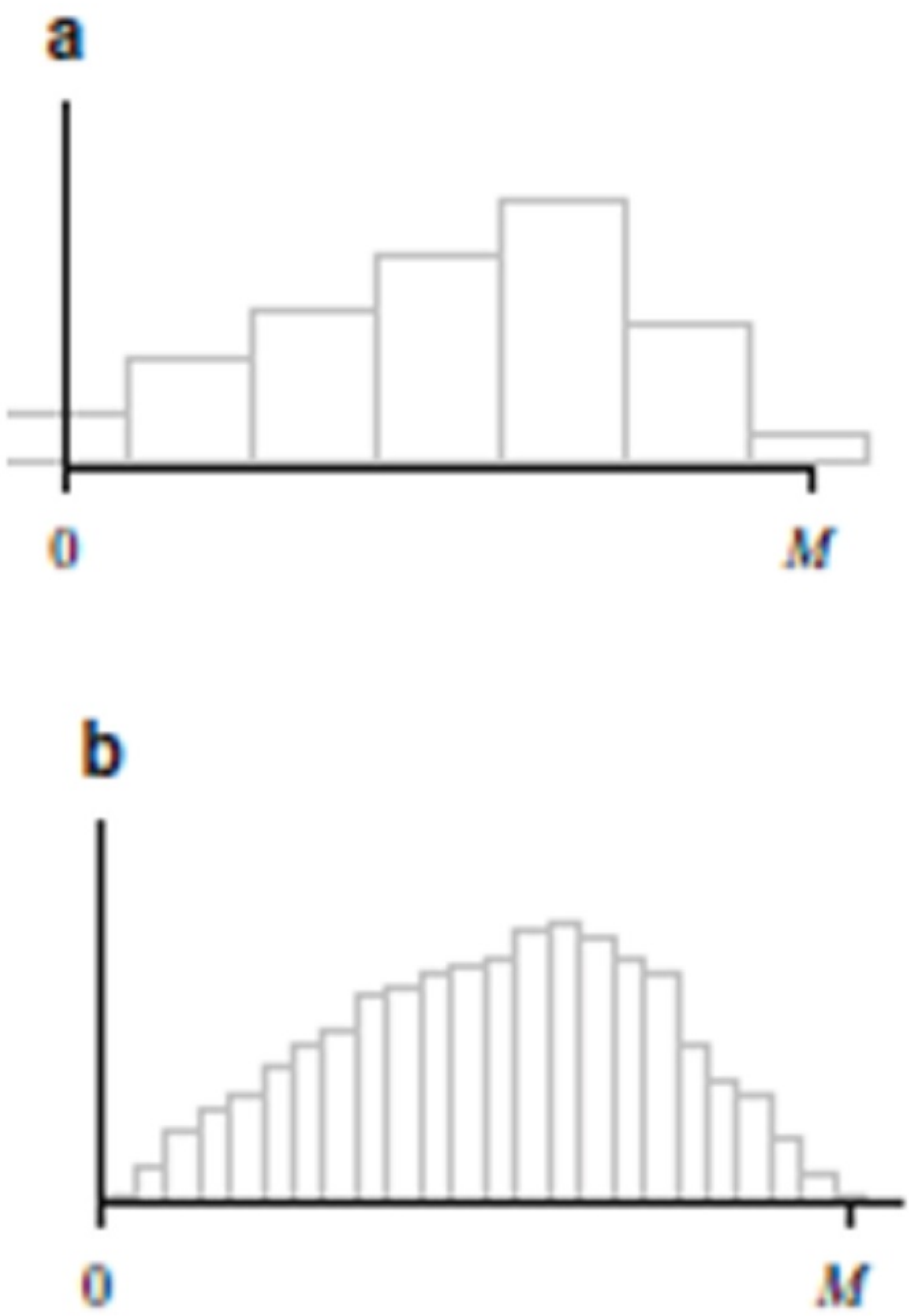
If a chemical compound is randomly selected and its pH X is determined, then X is a continuous random variable because any pH value between 0 and 14 is possible. If more is known about the compound selected for analysis, then the set of possible values might be a sub-interval of [0,14], such as $5.5 \leqslant X \leqslant 6.5$, but X would still be continuous.

PROBABILITY DENSITY FUNCTIONS

Suppose the variable X of interest is the depth of a lake at a randomly chosen point on the surface. Let $M =$ the maximum depth (in meters), so that any number in the interval [0, M] is a possible value of X. If we 'discretize' X by measuring depth to the nearest meter, then possible values are non-negative integers less than or equal to M. The resulting discrete distribution of depth can be pictured using a probability histogram. If we draw the histogram so that the area of the rectangle above any possible integer k is the proportion of the lake whose depth is (to the nearest meter) k, then the total area of all rectangles is 1. A possible histogram appears in the figures below.

If depth is measured much more accurately and the same measurement axis as in the figure above is used, each rectangle in the resulting probability histogram is much narrower, although the total area of all rectangles is still 1. A possible histogram is pictured in the next figure; it has a much smoother appearance than the first histogram.

If we continue in this way to measure depth more and more finely, the resulting sequence of histograms approaches a smooth curve, as pictured in our last figure below. Because for each histogram the total area of all rectangles equals 1, the total area under the smooth curve is also 1. The probability that the depth at a randomly chosen point is between a and b is just



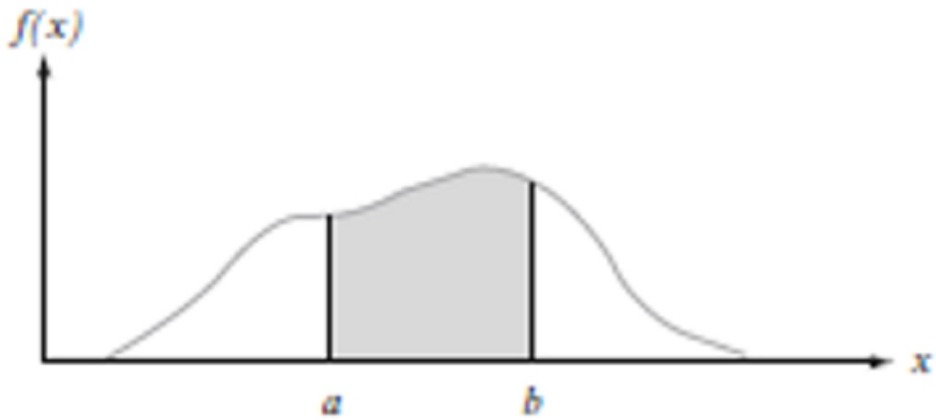
the area under the smooth curve between a and b. It is exactly a smooth curve of the type pictured in the following figure that specifies a continuous probability distribution.

Let X be a continuous random variable. Then a probability distribution or probability density function of X is a function $f(x)$ such that for any two numbers a and b with $a \leqslant b$,

$$p(a \leqslant X \leqslant b) = \int_a^b f(x)dx$$

That is, the probability that X takes on a value in the interval $[a, b]$ is the area above this

interval and under the graph of the density function, as illustrated in the figure below. The graph of $f(x)$ is often referred to as the density curve.



For $f(x)$ to be a legitimate probability density function it must satisfy two conditions namely:

1. $f(x) \geqslant 0$; probability exists.

2. $\int_{-\infty}^{\infty} f(x)dx = 1$; total probability

Also always remember that for a continuous random variable, $p(X = x) = 0$ hence $p(a \leqslant X \leqslant b) = p(a \leqslant X < b) = p(a < X \leqslant b) = p(a < X < b)$.

Recall: How to integrate

1. $\int x^n dx = \dfrac{x^{n+1}}{n+1} + c.$

2. $\int \dfrac{1}{x} dx = \ln(x) + c.$

3. $\int ax^n dx = a\dfrac{x^{n+1}}{n+1} + c.$

4. $\int e^x dx = e^x + c.$

5. $\int e^{ax} dx = \dfrac{1}{a} e^{ax} + c.$

Class Exercise 1

1. The random variable X has probability density function

$$f(x) = \begin{cases} kx(4-x), & 2 \leqslant x \leqslant 4 \\ 0, & \text{otherwise} \end{cases}$$

   Determine the value of k and sketch the probability density function of X.

2. The random variable Y has probability density function

$$f(y) = \begin{cases} k, & 1 < y < 2 \\ k(y-1), & 2 \leqslant y \leqslant 4 \\ 0, & \text{elsewhere} \end{cases}$$

   Determine the value of k and sketch the probability density function of Y.

3. The random variable X has probability density function

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

   Determine the value of k.

4. The random variable Z has probability density function

$$f(z) = \begin{cases} k, & 0 \leqslant z < 2 \\ k(2z-3), & 2 \leqslant z \leqslant 3 \\ 0, & \text{elsewhere} \end{cases}$$

Determine the value of k and sketch the probability density function of Z for all values of Z.

5. The direction of an imperfection with respect to a reference line on a circular object such as a tire, brake rotor, or flywheel is, in general, subject to uncertainty. Consider the reference line connecting the valve stem on a tire to the center point, and let X be the angle measured clockwise to the location of an imperfection. One possible probability density function for X is

$$f(x) = \begin{cases} \frac{1}{300}, & 0 \leqslant z \leqslant 360 \\ 0, & \text{elsewhere} \end{cases}$$

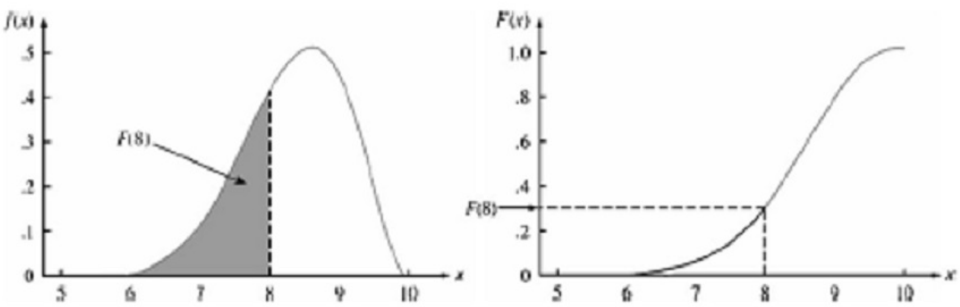
Determine the probability that the angle is between 90 and 180 i.e. $p(90 \leqslant X \leqslant 180)$.

Cumulative Distribution Function

The cumulative distribution function $F(x)$ for a continuous random variable X is defined for every number x by

$$F(x) = p(X \leqslant x) = \int_{-\infty}^{x} f(x)dx$$

For each x, $F(x)$ is the area under the density curve to the left of x as illustrated in the figure below. $F(x)$ increases smoothly as x increases.



Example

The random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{4}x, & 1 \leqslant x \leqslant 3 \\ 0, & \text{elsewhere} \end{cases}$$

Determine $F(x)$

Solution

For $x < 1$, $F(x) = \int_{-\infty}^{x} 0dx = 0$

For $1 \leqslant x \leqslant 3$, $F(x) = \int_{-\infty}^{1} 0dx + \int_{1}^{x} \frac{1}{4}xdx = \frac{x^2-1}{8}$

For $x > 3$, $F(x) = \int_{-\infty}^{1} 0dx + \int_{1}^{3} \frac{1}{4}xdx + \int_{3}^{x} 0dx = 1$

Hence

$$F(x) = \begin{cases} 0, & x < 1 \\ \frac{x^2-1}{8}, & 1 \leqslant x \leqslant 3 \\ 1, & x > 3 \end{cases}$$

Properties of CDFs for continuous random variables

$p(a < X \leqslant b) = p(X \leqslant b) - p(X \leqslant a) = F(b) - F(a)$

$p(a \leqslant X \leqslant b) = p(a < X \leqslant b) + p(X = a) = F(b) - F(a) + f(a)$

$p(a < X < b) = p(a < X \leqslant b) - p(X = b) = F(b) - F(a) - f(b)$

However for a continuous random variable $p(X \leqslant x) = 0$ and

$$f(x) = \frac{d}{dx}F(x)$$

Class Exercise 2

1. The random variable X has probability density function

$$f(x) = \begin{cases} \frac{1}{5}, & 1 < x < 2 \\ \frac{1}{5}(x-1), & 2 \leqslant x \leqslant 4 \\ 0, & \text{elsewhere} \end{cases}$$

   Find $F(x)$.

2. The random variable X has cumulative distribution function given by

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{5}x + \frac{3}{20}x^2, & 0 \leqslant x \leqslant 2 \\ 1, & x > 2 \end{cases}$$

   Find

   i) $p(X \leqslant 1.5)$.

   ii) $p(0.5 \leqslant X \leqslant 1.5)$.

iii) $p(X = 1)$.

iv) $f(x)$.

3. Suppose that the time in minutes that a person has to wait at a certain station for a train is seen to be a random phenomenon and has a probability function specified by the cumulative distribution function

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{2}, & 0 \leqslant x < 1 \\ \frac{1}{2}, & 1 \leqslant x < 2 \\ \frac{x}{4}, & 2 \leqslant x < 4 \\ 1, & x > 4 \end{cases}$$

What is the probability that a person will have to wait

a) more than three minutes?

b) less than three minutes?

c) between one and three minutes?

SOME IMPORTANT RESULTS

This section will provide nine results which you will use henceforth without prove. You will meet and prove most of them in your Mathematics courses.

1. Summation and its properties, $\sum$ read 'sigma', is the summation notation.

a) $x_1 + x_2 + \cdots + x_n = \sum_{i=1}^{n} x_i$

b) If a and b are fixed numbers then
$\sum_{i=1}^{n} bx_i = b\sum_{i=1}^{n} x_i$
$\sum_{i=1}^{n} (bx_i + a) = b\sum_{i=1}^{n} x_i + an$

2. $(a+b)^n = \binom{n}{0}a^n b^0 + \binom{n}{1}a^{n-1}b^1 + \binom{n}{2}a^{n-2}b^2 + \cdots + \binom{n}{n}a^0 b^n$

3. $e^x = 1 + \dfrac{x}{1!} + \dfrac{x^2}{2!} + \dfrac{x^3}{3!} + \cdots = \sum_{k=0}^{\infty} \dfrac{x^k}{k!}$

4. $\sum_{l=0}^{\infty} ar^l = a + ar + ar^2 + \cdots = \dfrac{a}{(1-r)}, |r| < 1$

5. $1^2 + 2^2 + 3^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$

6. $\sum_{k=0}^{\infty} kx^{(k-1)} = 1 + 2x + 3x^2 + 4x^3 + \cdots = \dfrac{1}{(1-x)^2}, |x| < 1$

7. $\sum_{k=0}^{\infty} k^2x^{(k-1)} = 1 + 4x + 9x^2 + 16x^3 + \cdots = \dfrac{(1+x)}{(1-x)^3}, |x| < 1$

8. $1 + rx + \dfrac{r(r+1)}{2!}x^2 + \dfrac{r(r+1)(r+2)}{3!}x^3 + \cdots = (1-x)^{-r}$

9. $\lim_{n\to\infty} \left(1 - \dfrac{a}{n}\right)^n = e^{-a}$

EXPECTATIONS AND MOMENTS OF RANDOM VARIABLES

Another view of probability mass functions

It is often helpful to think of a probability mass function as specifying a mathematical model for a discrete population. Consider selecting at random a student who is among the 15000 registered for the current semester at Maseno University. Let X be the number of courses for which the selected student, is registered for and suppose that X has the following probability mass function:

x	1	2	3	4	5	6	7
$p(x) = p(X = x)$	0.01	0.03	0.13	0.25	0.39	0.17	0.02

Since $p(1) = 0.01$, we know that $0.01 \times 15000 = 150$ of the students are registered for one course, and similarly for the other x values.

x	1	2	3	4	5	6	7
$p(x) = p(X = x)$	0.01	0.03	0.13	0.25	0.39	0.17	0.02
Number of Students	150	450	1950	3750	5850	2550	300

To compute the average number of courses per student, or the average value of X in the population, we should calculate the total number of courses and divide by the total number of students. Since each of 150 students is taking one course, these 150 contribute 150 courses to the total. Similarly, 450 students contribute 2(450) courses, and so on. The population average value of X is then

$$\frac{1(150)+2(450)+3(1950)+\cdots+7(300)}{15000}=4.57 \tag{2}$$

Since $\frac{150}{15000}=.01=p(1)$, $\frac{450}{15000}=.03=p(2)$, and so on, an alternative expression for the population average value of X is

$$1p(1)+2p(2)+3p(3)+\cdots+7p(7)=\sum_{\forall x} xp(x) \tag{3}$$

This shows that to compute the population average value of X, we need only the possible values of X along with their probabilities (proportions).

In particular, the population size is irrelevant as long as the probability mass function is given. The average or mean value of X is then a weighted average of the possible values $1,\cdots,7$, where the weights are the probabilities of those values.

NB: The mean value of a random variable X is called the **expected value** of (or **expectation** of) the random variable X and is denoted $E(X)$.

EXPECTATION OF A RANDOM VARIABLE

Let $g(x)$ denote any function of X; then we define the expected value of $g(x)$, $E[g(x)]$ as

$$E[g(x)]=\begin{cases}\sum_{\forall x} g(x)p(x), & \text{for } x \text{ discrete}\\ \int_{-\infty}^{\infty} g(x)f(x), & \text{for } x \text{ continuous}\end{cases}$$

Example 1

A coin is thrown 3 times. X denotes the number of heads recorded. Let $g(x)=2x+1$. Find $E[g(x)]$.

Solution

x	0	1	2	3
$p(x)=p(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$
$g(x)=2x+1$	1	3	5	7
$g(x)p(x)$	$\frac{1}{8}$	$\frac{9}{8}$	$\frac{15}{8}$	$\frac{7}{8}$

$E[g(x)]=\frac{1}{8}+\frac{9}{8}+\frac{15}{8}+\frac{7}{8}=4$

Example 2

A random variable Y has probability function

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

Determine $E[2y + 3]$.

Solution

$$\begin{aligned} E[2y+3] &= \int_0^1 (2y+3)2y\,dy \\ &= \int_0^1 (4y^2 + 6y)\,dy \\ &= \left[\frac{4y^3}{3} + 3y^2\right]_0^1 \\ &= \left(\frac{4}{3} + 3\right) - (0+0) \\ &= \frac{13}{3} \end{aligned}$$

Special case

Let $g(x) = x$ then

$$E[x] = \begin{cases} \sum_{\forall x} xp(x), & \text{for } x \text{ discrete} \\ \int_{-\infty}^{\infty} xf(x), & \text{for } x \text{ continuous} \end{cases}$$

This special type of expectation is called the mean of the random variable and is denoted by μ. Hence $\mu = E[X]$. It is also referred to as the 1^{st} moment of the random variable X about the origin.

$E[X^r]$ is the r^{th} moment of the random variable X about the origin.

The mean of a random variable X is it's 1^{st} moment about the origin.

In Example 1, the mean of the random variable X is given by

$$\mu = E[X] = \sum_{\forall x} xp(x) = \left(0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8}\right) = 1.5$$

Properties of the Mean

Theorem 1

Let X be a random variable and k a real number then:

1.) $E[kX] = kE[X]$

2.) $E[X + k] = E[X] + k$

Class Exercise 1

Prove Theorem 1 for X discrete and continuous.

Special Type of Function

Let $g(x) = (x - \mu)^2$ then where μ is the mean of the random variable X then

$$E[(x-\mu)^2] = \begin{cases} \sum_{\forall x}(x-\mu)^2 p(x), & \text{for } x \text{ discrete} \\ \int_{-\infty}^{\infty}(x-\mu)^2 f(x), & \text{for } x \text{ continuous} \end{cases}$$

This special type of expectation is called the variance of the random variable and is denoted by $\sigma^2 = Var(X)$. It is also referred to as the 2^{nd} moment of the random variable X about the mean.

Class Exercise 2

Show that the 1^{st} moment of a random variable X about its mean is 0.

In general $E[(X - A)^r]$ refers to the r^{th} moment of the random variable X about A.

In Example 1, the variance of the random variable X is given by

$$\begin{aligned} \sigma^2 &= E[(X-\mu)^2] \\ &= E[(X-1.5)^2] \\ &= \sum_{\forall x}(X-1.5)^2 p(x) \\ &= \left[\left(\frac{9}{4}\times\frac{1}{8}\right) + \left(\frac{1}{4}\times\frac{3}{8}\right) + \left(\frac{1}{4}\times\frac{3}{8}\right) + \left(\frac{9}{4}\times\frac{1}{8}\right)\right] \\ &= 0.75 \end{aligned}$$

Theorem 2

Given that X is a random variable

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E(X^2) - 2\mu Ex + E[\mu^2] \\ &= E(X^2) - 2\mu^2 + \mu^2 \\ &= E(X^2) - \mu^2 \\ &= E(X^2) - [E(X)]^2 \end{aligned}$$

The variance is the 2^{nd} moment about the origin minus the 1^{st} moment about the origin squared.

Properties of the Variance

Theorem 3

Let X be a random variable and k a real number then:

1.) $Var[kX] = k^2 Var[X]$

2.) $Var[X + k] = Var[X]$

Class Exercise 3

Prove Theorem 3 for X discrete and continuous.

Example 3

A random variable Y has probability density function

$$f(y) = \begin{cases} \frac{y}{4}, & 1 \leqslant y \leqslant 3 \\ 0, & \text{elsewhere} \end{cases}$$

Find: $E(y)$, $E(2y-3)$, $Var(y)$ and $Var(2y-3)$.

Solution

$$\begin{aligned}
E(y) &= \int_1^3 \frac{y^2}{4} dy \\
&= \left[\frac{y^3}{12}\right]_1^3 \\
&= \frac{13}{6}
\end{aligned}$$

$$E(2y-3) = 2E(y) - 3 = \frac{13}{3} - 3 = \frac{4}{3}$$

$$\begin{aligned}
Var(y) &= E(y^2) - [E(y)]^2 \\
&= \int_1^3 \frac{y^3}{4} dy - \frac{169}{36} \\
&= \left[\frac{y^4}{16}\right]_1^3 - \frac{169}{36} \\
&= \frac{80}{16} - \frac{169}{36} \\
&- \frac{11}{36}
\end{aligned}$$

$$Var(2y-3) = 2^2 Var(y) = 4 \times \frac{11}{36} = \frac{11}{9}.$$

Class Exercise 4

1. Telephone calls arriving at a company are referred immediately by the receptionist to other people working in the company. The time a call takes in minutes is modeled by the continuous random variable T having the probability density function

$$f(t) = \begin{cases} kt^2, & 0 \leqslant t \leqslant 10 \\ 0, & \text{elsewhere} \end{cases}$$

   a) Find k, $E(t)$ and $Var(t)$.

   b) Find the probability of a call lasting between 7 and 9 minutes.

   c) Sketch the probability density function of T.

2. Two fair ten shilling coins are tossed. The random variable X represents the total value of the coin that lands heads up.

   a) Find $E(X)$ and $Var(X)$.

   b) Random variables S and T are defined as follows: $S = X - 10$ and $T = \frac{1}{2}X - 5$. Show that $E(S) = E(T)$.

   c) Find $Var(S)$ and $Var(T)$.