S1 Chapter 6: Statistical Distributions

Probability Distributions

This Chapter Overview

1 :: General Probability Distributions

"Given that $P(X = x) = \frac{k}{x}$, find the value of k."

2 :: Binomial Distribution

"I toss an unfair coin, with probability heads of 0.6, 10 times. What's the probability I see 5 heads?"

3 :: Cumulative Binomial Probabilities

"I toss an unfair coin, with probability heads of 0.6, 10 times. What's the probability I see at most 3 heads?"

Probability distributions

You are already familiar with the concept of **variable** in statistics: a collection of values (e.g. favourite colour of students in the room):

$\boldsymbol{x}$	red	green	blue	orange
$\boldsymbol{P(X = x)}$	0.3	0.4	0.1	0.2

If each is assigned a probability of occurring, it becomes a **random variable**.

✏ A **random variable X represents a single experiment/trial**. It consists of outcomes with a probability for each.

i.e. X is a random variable (capital letter), but x is a particular outcome.

$$P(X = x)$$

"The probability that…

…the outcome of the random variable X…

…was the specific outcome x"

A shorthand for $P(X = x)$ is ✏ $p(x)$ (note the lowercase p).
It's like saying "the probability that the outcome of my coin throw was heads" ($P(X = heads)$) vs "the probability of heads" ($p(heads)$). In the latter the coin throw was implicit, so we can skip the '$X =$ '.

Probability Distributions vs Probability Functions

There are two ways to write the mapping from outcomes to probabilities:

The "{" means we have a 'piecewise function'. This just simply means we choose the function from a list depending on the input.

As a function:

$$p(x) = \begin{cases} 0.1x, & x = 1,2,3,4 \\ 0, & otherwise \end{cases}$$

e.g. if $x = 2$, then the probability is $0.1 \times 2 = 0.2$

Advantages of functional form:

?

As a table:

x	1	2	3	4
$p(x)$	?			

The table form that you know and love.

Advantages of table form:

?

Probability Distributions vs Probability Functions

There are two ways to write the mapping from outcomes to probabilities:

The "{" means we have a 'piecewise function'. This just simply means we choose the function from a list depending on the input.

As a function:

$$p(x) = \begin{cases} 0.1x, & x = 1,2,3,4 \\ 0, & otherwise \end{cases}$$

e.g. if $x = 2$, then the probability is $0.1 \times 2 = 0.2$

Advantages of functional form:
Can have a rule/expression based on the outcome. Particularly for continuous random variables (in Yr2), it would be impossible to list the probability for every outcome. More compact.

As a table:

x	1	2	3	4
$p(x)$	0.1	0.2	0.3	0.4

The table form that you know and love.

Advantages of table form:
Probability for each outcome more explicit.

Example

The random variable X represents the **number of heads when three coins are tossed**.

Underlying Sample Space	Distribution as a Table
?	?
	Distribution as a Function
	?

Example

The random variable X represents the **number of heads when three coins are tossed**.

Underlying Sample Space

{ HHH,
HHT,
HTT,
HTH,
THH,
THT,
TTH,
TTT }

Distribution as a Table

Num heads x	0	1	2	3
$P(X = x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

Distribution as a Function

$$P(X = x) = \begin{cases} \frac{1}{8} & x = 0,3 \\ \frac{3}{8} & x = 1,2 \\ 0 & otherwise \end{cases}$$

Example Exam Question

(Hint: Use your knowledge that $\Sigma p(\ldots) = 1$)

Edexcel S1 May 2012

1. A discrete random variable X has the probability function

$$P(X = x) = \begin{cases} k(1-x)^2 & x = -1, 0, 1 \text{ and } 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $k = \frac{1}{6}$.

(3)

x	-1	0	1	2
$p(x)$	?	?	?	?

?

Example Exam Question

(Hint: Use your knowledge that $\Sigma p(\ldots) = 1$)

Edexcel S1 May 2012

1. A discrete random variable X has the probability function

$$P(X = x) = \begin{cases} k(1-x)^2 & x = -1, 0, 1 \text{ and } 2 \\ 0 & \text{otherwise.} \end{cases}$$

(a) Show that $k = \frac{1}{6}$.

(3)

$\boldsymbol{x}$	-1	0	1	2
$\boldsymbol{p(x)}$	$k(1--1)^2 = 4k$	k	0	k

Since $\Sigma p(x) = 1$, $\quad 4k + k + 0 + k = 1$

$$6k = 1 \quad \rightarrow \quad k = \frac{1}{6}$$

Probability of a Range

x	2	3	4	5
$p(x)$	0.1	0.3	0.2	0.4

Determine:

$P(X > 3) =$?

$P(2 \leq X < 4) =$?

$P(2X + 1 \geq 6) =$?

Probability of a Range

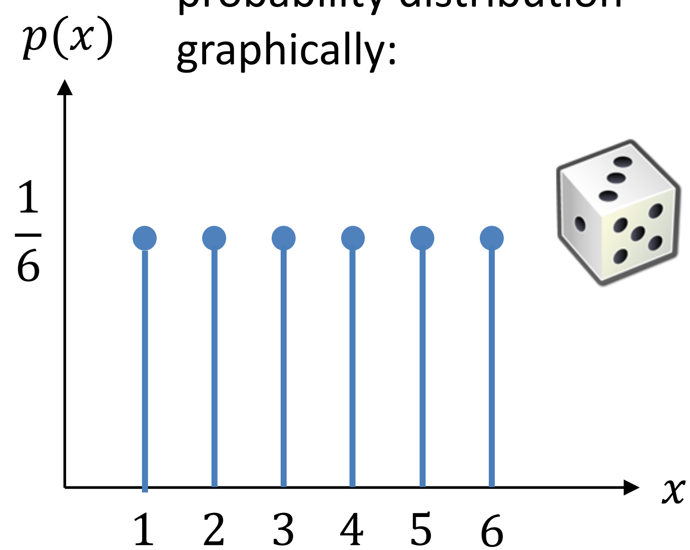
$\boldsymbol{x}$	2	3	4	5
$\boldsymbol{p}(\boldsymbol{x})$	0.1	0.3	0.2	0.4

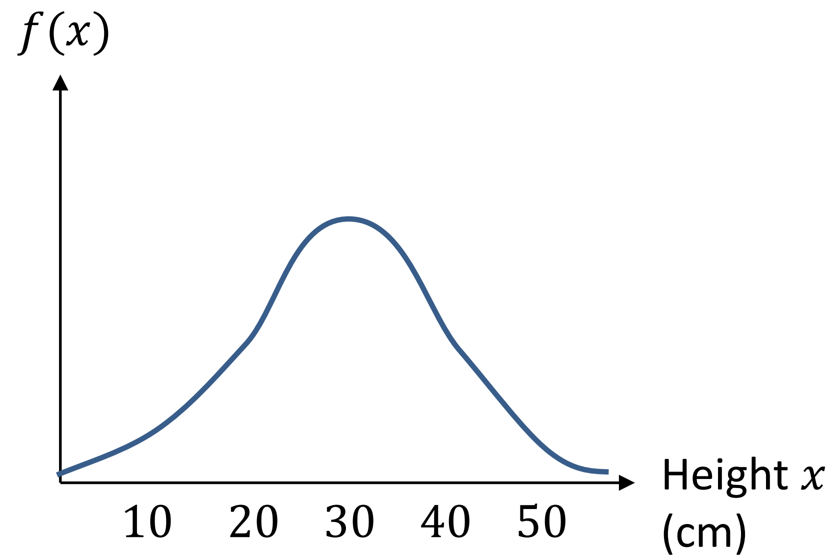
Determine:

$$P(X > 3) = \mathbf{0.2 + 0.4 = 0.6}$$
$$P(2 \leq X < 4) = \mathbf{0.1 + 0.3 = 0.4}$$
$$P(2X + 1 \geq 6) = P(X \geq 2.5) = \mathbf{0.3 + 0.2 + 0.4 = 0.9}$$

A few last things...

We can also represent a probability distribution graphically:





✐ The throw of a die is an example of a **discrete uniform distribution** because the probability of each outcome is the same.

$p(x)$ for discrete random variables is known as a **probability mass function**, because the probability of each outcome represents an actual 'amount' (i.e. mass) of probability.

We can also have probability distributions for **continuous** variables, e.g. height.
However, the probability that something has a height of say **exactly** 30cm, is infinitely small (effectively 0).
$p(x)$ (written $f(x)$) for continuous random variables is known as a **probability density function**. $p(30)$ wouldn't give us the probability of being 30cm tall, but the amount of probability **per unit height**, i.e. the density. This is similar to histograms where frequency density is the "frequency per unit value". Just as an area in a histogram would then give a frequency, and area under a probability density graph would give a probability (mass).

You will encounter the **Normal Distribution** in Year 2, which is an example of a continuous probability distribution.

Exercise 6.1

Pearson Applied Year 1/AS
Pages 39-40

Homework Exercise

1 Write down whether or not each of the following is a discrete random variable.
Give a reason for your answer.

a The height, X cm, of a seedling chosen randomly from a group of plants.

b The number of times, R, a six is rolled when a fair dice is rolled 100 times.

c The number of days, W, in a given week.

2 A fair dice is thrown four times and the number of times it falls with a 6 on the top, Y, is noted.
Write down the sample space of Y.

3 A bag contains two discs with the number 2 on them and two discs with the number 3 on them.
A disc is drawn at random from the bag and the number noted. The disc is returned to the bag.
A second disc is then drawn from the bag and the number noted.

a Write down all the possible outcomes of this experiment.

The discrete random variable X is defined as the sum of the two numbers.

b Write down the probability distribution of X as:

i a table **ii** a probability mass function.

4 A discrete random variable X has the probability distribution shown in the table.
Find the value of k.

x	1	2	3	4
$\mathrm{P}(X = x)$	$\frac{1}{3}$	$\frac{1}{3}$	k	$\frac{1}{4}$

5 The random variable X has a probability function

$$\mathrm{P}(X = x) = kx \qquad x = 1, 2, 3, 4.$$

Show that $k = \frac{1}{10}$. **(2 marks)**

Homework Exercise

6 The random variable X has a probability function

$$P(X = x) = \begin{cases} kx & x = 1, 3 \\ k(x-1) & x = 2, 4 \end{cases}$$

where k is a constant.

a Find the value of k. **(2 marks)**

b Find $P(X > 1)$. **(2 marks)**

7 The discrete random variable X has a probability function

$$P(X = x) = \begin{cases} 0.1 & x = -2, -1 \\ \beta & x = 0, 1 \\ 0.2 & x = 2 \end{cases}$$

a Find the value of β.

b Construct a table giving the probability distribution of X.

c Find $P(-1 \leqslant X < 2)$.

8 A discrete random variable has a probability distribution shown in the table.

Find the value of a.

x	0	1	2
$\mathbf{P(X = x)}$	$\frac{1}{4} - a$	a	$\frac{1}{2} + a$

9 The random variable X can take any integer value from 1 to 50. Given that X has a discrete uniform distribution, find:

a $P(X = 1)$

b $P(X \geqslant 28)$

c $P(13 < X < 42)$

Homework Exercise

10 A discrete random variable X has the probability distribution shown in this table.

x	0	1	2	3
P($X = x$)	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{8}$

Find:

a $P(1 < X \leqslant 3)$ **(1 mark)**

b $P(X < 2)$ **(1 mark)**

c $P(X > 3)$ **(1 mark)**

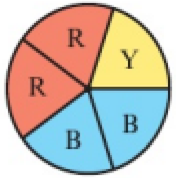
11 A biased coin is tossed until a head appears or it is tossed four times.

If $P(\text{Head}) = \frac{2}{3}$:

a Write down the probability distribution of S, the number of tosses, in table form. **(4 marks)**

b Find $P(S > 2)$. **(1 mark)**

12 A fair five-sided spinner is spun.



Given that the spinner is spun five times, write down, in table form, the probability distributions of the following random variables:

a X, the number of times red appears

b Y, the number of times yellow appears.

The spinner is now spun until it lands on blue, or until it has been spun five times. The random variable Z is defined as the number of spins in this experiment.

c Find the probability distribution of Z.

Homework Exercise

13 Marie says that a random variable X has a probability distribution defined by the following probability mass function:

$$P(X = x) = \frac{2}{x^2}, \quad x = 2, 3, 4$$

a Explain how you know that Marie's function does not describe a probability distribution. **(2 marks)**

b Given that the correct probability mass function is in the form $P(X = x) = \frac{k}{x^2}, \quad x = 2, 3, 4$ where k is a constant, find the exact value of k. **(2 marks)**

Challenge

The independent random variables X and Y have probability distributions

$$P(X = x) = \frac{1}{8}, \quad x = 1, 2, 3, 4, 5, 6, 7, 8 \qquad P(Y = y) = \frac{1}{y}, \quad y = 2, 3, 6$$

Find $P(X > Y)$.

Hint X and Y are independent so the value taken by one does not affect the probabilities for the other.

Homework Answers

1 **a** This is not a discrete random variable, since height is continuous quantity.
b This is a discrete random variable, since it is always a whole number and it can vary.
c This is not a discrete random variable, since the number of days in a given week is always 7.

2 0, 1, 2, 3, 4

3 **a** (2, 2) (2, 3) (3, 2) (3, 3)
b **i**

x	4	5	6
$P(X = x)$	0.25	0.5	0.25

ii
$$P(X = x) = \begin{cases} 0.25, & x = 4, 6 \\ 0.5, & x = 5 \end{cases}$$

4 $\frac{1}{12}$

5 $k + 2k + 3k + 4k = 1$,
so $10k = 1$, so $k = \frac{1}{10}$.

6 **a** 0.125 **b** 0.875

7 **a** 0.3
b

x	−2	−1	0	1	2
$P(X = x)$	0.1	0.1	0.3	0.3	0.2

c 0.7

8 0.25

9 **a** 0.02 **b** 0.46 **c** 0.56

10 **a** 0.625 **b** 0.375 **c** 0

11 **a**

s	1	2	3	4
$P(S = s)$	$\frac{2}{3}$	$\frac{2}{9}$	$\frac{2}{27}$	$\frac{1}{27}$

b $\frac{1}{9}$

12 **a**

x	$P(X = x)$
0	0.07776
1	0.2592
2	0.3456
3	0.2304
4	0.0768
5	0.01024

b

y	$P(X = y)$
0	0.32768
1	0.4096
2	0.2048
3	0.0512
4	0.0064
5	0.00032

c

z	$P(X = z)$
1	0.4
2	0.24
3	0.144
4	0.0864
5	0.1296

13 **a** The sum of the probabilities is not 1.
b $2\frac{22}{61}$

Challenge
0.625