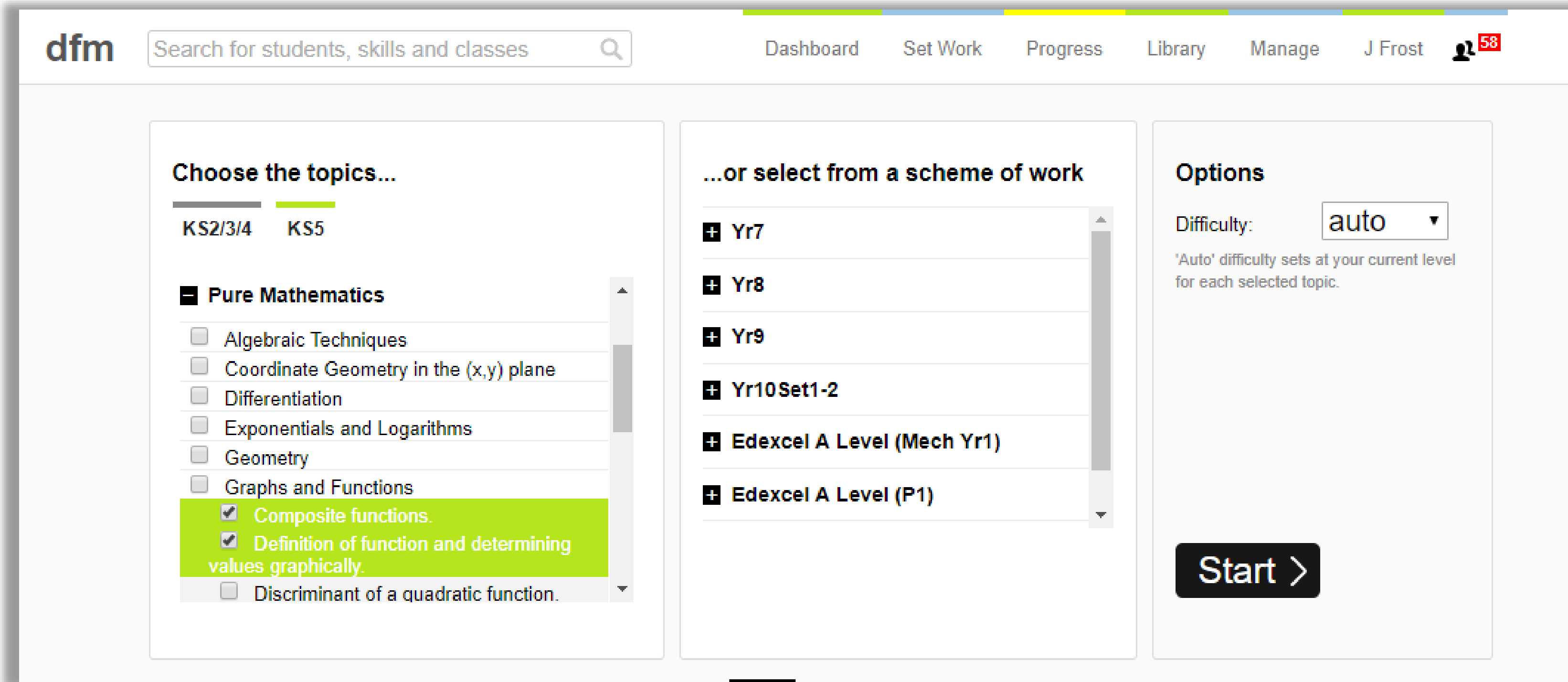
Stats1 Chapter 6 :: Statistical Distributions

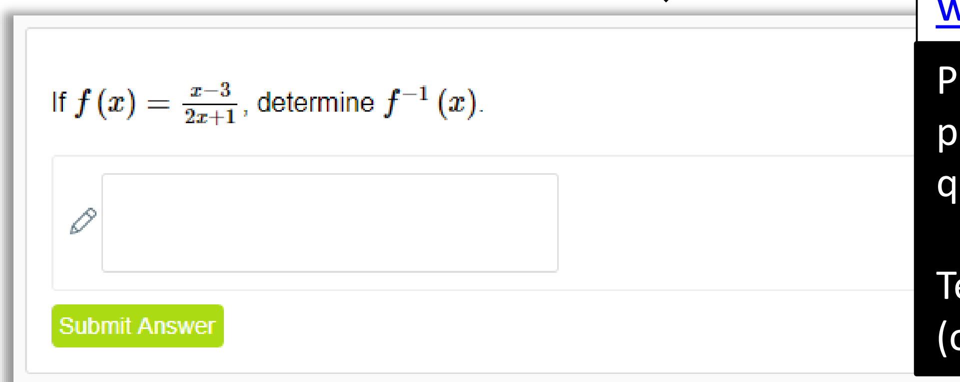
jfrost@tiffin.kingston.sch.uk

www.drfrostmaths.com
@DrFrostMaths

Last modified: 17th February 2018

Use of DrFrostMaths for practice





Register for **free** at:
www.drfrostmaths.com/homework

Practise questions by chapter, including past paper Edexcel questions and extension questions (e.g. MAT).

Teachers: you can create student accounts (or students can register themselves).

Experimental

i.e. Dealing with collected data.

Chp1: Data Collection

Methods of sampling, types of data, and populations vs samples.

Chp2: Measures of Location/Spread

Statistics used to summarise data, including mean, standard deviation, quartiles, percentiles. Use of linear interpolation for estimating medians/quartiles.

Chp3: Representation of Data

Producing and interpreting visual representations of data, including box plots and histograms.

Chp4: Correlation

Measuring how related two variables are, and using linear regression to predict values.

Theoretical

Deal with probabilities and modelling to make inferences about what we 'expect' to see or make predictions, often using this to reason about/contrast with experimentally collected data.

Chp5: Probability

Venn Diagrams, mutually exclusive + independent events, tree diagrams.

Chp6: Statistical Distributions

Common distributions used to easily find probabilities under certain modelling conditions, e.g. binomial distribution.

Chp7: Hypothesis Testing

Determining how likely observed data would have happened 'by chance', and making subsequent deductions.

This Chapter Overview

This chapter is a recap of the concepts you learnt at GCSE.

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?"

Changes since the old 'S1' syllabus:
You are no longer required to find the expected value ($E(X)$) or variance of a random variable, or find the cumulative distribution function of a probability mass function.

The Binomial distribution has been moved from S2 to this module.

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:

?

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 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
$p(x)$	?	?	?	?

?

Probability of a Range

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

Determine:

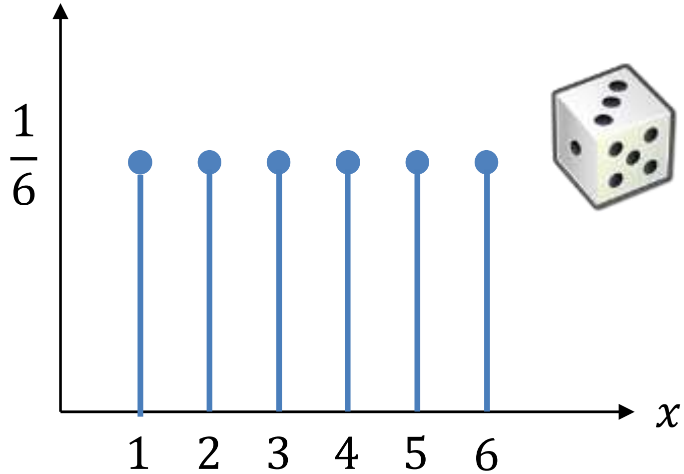
$P(X > 3) =$?

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

$P(2X + 1 \geq 6) = P(X \geq 2.5) =$?

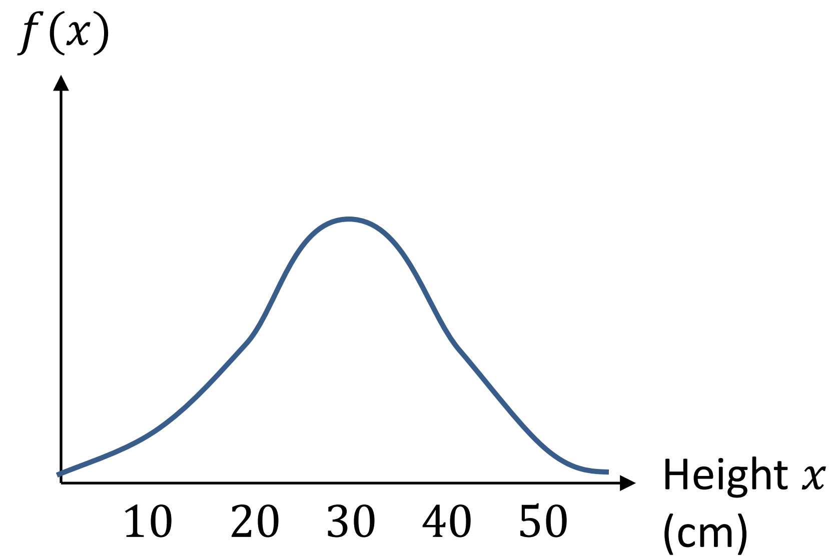
A few last things…

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).

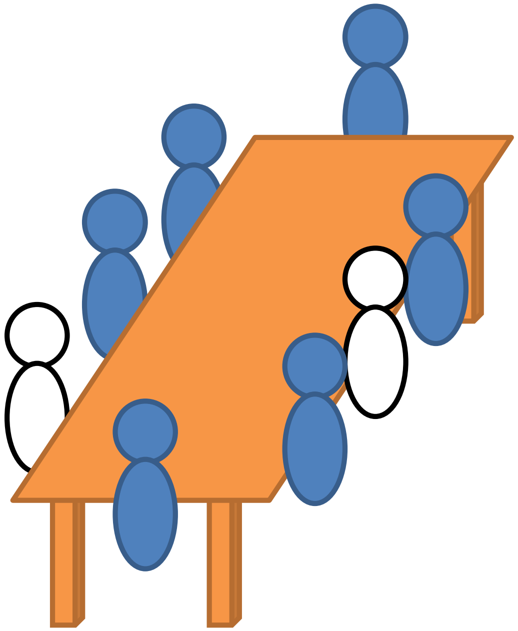
You will encounter the **Normal Distribution** in Year 2, which is an example of a continuous probability distribution.

Exercise 6A

Pearson Pure Mathematics Year 1/AS
Pages 86-88

Frost True Stories

Back in 2010 I was on holiday in Hawaii and visited the family of a friend. We noticed that at the dinner table that out of the **8 of us, 6 of us were left-handed** (including myself). One of them commented, “The chances of that must be very low”.
“CHALLENGE ACCEPTED”.



Leftie Example

Let's simplify the problem by using just 3 people:

The probability a randomly chosen person is left-handed is 0.1. If there is a group of 3 people, what is the probability that:

a) All 3 are left-handed.
b) 0 are left-handed.
c) 1 person is left-handed.
d) 2 people are left-handed.

d ?

a ?

b ?

c ?

Let's try to generalise!
If there were x 'lefties' out of 3, then we can see, using the examples, that the probability of a single matching outcome is $0.1^x \times 0.9^{3-x}$. How many rows did we have each time? In a sequence of three L's and R's, there are "3 choose x", i.e. $\binom{3}{x}$ ways of choosing x of the 3 letters to be L's. Therefore the probability of x out of 3 people being left handed is:

$$\binom{3}{x} 0.1^x \, 0.9^{3-x}$$

The Binomial Distribution

✎ You can model a random variable X with a binomial distribution $B(n, p)$ if
- there are a fixed number of trials, n,
- there are two possible outcomes: 'success' and 'failure',
- there is a fixed probability of success, p
- the trials are independent of each other

If $X \sim B(n, p)$ then:

$$P(X = r) = \binom{n}{r} p^r (1-p)^{n-r}$$

In our example, 'success' was 'leftie'.

r is the number of successes out of n.

"$\sim$" means "has the distribution"

On a table of 8 people, 6 people are left handed.

a) Suggest a suitable model for a random variable X: the number of left-handed people in a group of 8, where the probability of being left-handed is 0.1.
b) Find the probability 6 people are left handed.
c) Suggest why the chosen model may not have been appropriate.

a ?

b ?

c ?

In general, choosing a well-known model, such as a Binomial distribution, makes certain **simplifying assumptions**. Such assumptions simplifies the maths involved, but potentially at the expense of not adequately modelling the situation.

Binomial Coefficient

Binomial Coefficient

For non-negative integers n and r, with $r \leq n$,

$$ {}^{n}C_{r} = \binom{n}{r} = \frac{n!}{r!(n-r)!} $$

Further Examples

The random variable $X \sim B\left(12, \frac{1}{6}\right)$. Find:

a) $P(X = 2)$
b) $P(X = 9)$
c) $P(X \leq 1)$

a ?

b ?

c ?

Tip: The two powers add up to n.

Tip: Remember the two 'edge cases':

$$P(X = 0) = (1 - p)^n$$
$$P(X = n) = p^n$$

Edexcel S2 June 2010 Q6

A company claims that a quarter of the bolts sent to them are faulty. To test this claim the number of faulty bolts in a random sample of 50 is recorded.

(a) Give two reasons why a binomial distribution may be a suitable model for the number of faulty bolts in the sample. (2)

?

Test Your Understanding

1 $X \sim B(6, 0.2)$
What is $P(X = 2)$?

?

What is $P(X \geq 5)$?

?

2 I have a bag of 2 red and 8 white balls. X represents the number of red balls I chose after 5 selections (with replacement).

a How is X distributed?

?

b Determine the probability that I chose 3 red balls.

?

More Challenging Example

Q

An awkward Tiffin boy ventures into Tiffin Girls. He asks 20 girls out on the date. The probability that each girl says yes is 0.3.
Determine the probability that he will end up with:

a) Less than 6 girls on his next date.
b) At least 9 girls on his next date.

The boy considers the evening a success if he dated at least 9 girls that evening. He repeats this process across 5 evenings.
c) Calculate the probability that he had at least 4 successful evenings.
(Note: You won't be able to use your table for (c) as p is not a nice round number – calculate prob directly)

a

?

b

?

c

?

Exercise 6B

Pearson Pure Mathematics Year 1/AS
Pages 90-91

Cumulative Probabilities

Often we wish to find the probability of a range of values.
For a Binomial distribution, this was relatively easy if the range was narrow, e.g. $P(X \leq 1) = P(X = 0) + P(X = 1)$, but would be much more computationally expensive if we wanted say $P(X \leq 6)$.

If $X \sim B(10, 0.3)$, find $P(X \leq 6)$.

How to calculate on your ClassWiz:

Press Menu then 'Distributions'.
Choose "Binomial CD" (the C stands for 'Cumulative').
Choose 'Variable'.

$$x = 6$$
$$N = 10$$
$$p = 0.3$$

Pressing = gives the desired value.

Using tables (e.g. Page 204 of textbook)

Look up $n = 10$ and the column $p = 0.3$.
Then look up the row $x = 6$.
The value should be 0.9894.

Important Note: The tables only have limited values of p. You may have to use your calculator.
You will need to use your calculator in the exam anyway.

Cumulative Probabilities

The random variable $X \sim B(20, 0.4)$. Find:

$P(X \leq 7) =$?

$P(X < 6) =$?

$P(X \geq 15) =$?

Look up $n = 20, p = 0.4, x = 7$

Note that the table requires $\leq$

To get this right, just say in your head *"What's the opposite of 'at least 15'?"*. Hopefully you can see it's *'at most 14'*.

Given that $X \sim B(25, 0.25)$

$P(X = 6) =$?

$P(X > 20) =$?

$P(6 < X \leq 10) =$?

X can be 7 to 10. So we want up to 10, with everything up to 6 excluded.

Quickfire Questions

Write the following in terms of cumulative probabilities, e.g. $P(X < 7) = P(X \leq 6)$

$P(X < 5) =$?

$P(X \geq 7) =$?

$P(X > 7) =$?

$P(10 \leq X < 20) =$?

$P(10 \leq X \leq 20) =$?

$P(X = 100) =$?

$P(20 < X < 30) =$?

"at least 30" = ?

"greater than 30" = ?

Dealing with Probability Ranges

Q

A spinner is designed so that probability it lands on red is 0.3. Jane has 12 spins.

a) Find the probability that Jane obtains at least 5 reds.

Jane decides to use this spinner for a class competition. She wants the probability of winning a prize to be < 0.05. Each member of the class will have 12 spins and the number of reds will be recorded.

b) Find how many reds are needed to win the prize.

?

Test Your Understanding

Q

At Camford University, students have 20 exams at the end of the year. All students pass each individual exam with probability 0.45. Students are only allowed to continue into the next year if they pass some minimum of exams out of the 20.
What do the university administrators set this minimum number such that the probability of continuing to next year is at least 90%?

?

Exercise 6C

Pearson Pure Mathematics Year 1/AS
Pages 93-94