**Stats 2 (AQA) Formula**

(everything in blue in formula book)

**Expectation** (mean) **& Variance of Random Variable** from Probability Distribution

( ≠ .  is *theoretical mean*,  is *sample mean*.)

**Poisson distribution** (a discrete distribution)

|  |  |  |
| --- | --- | --- |
|  |  | , |

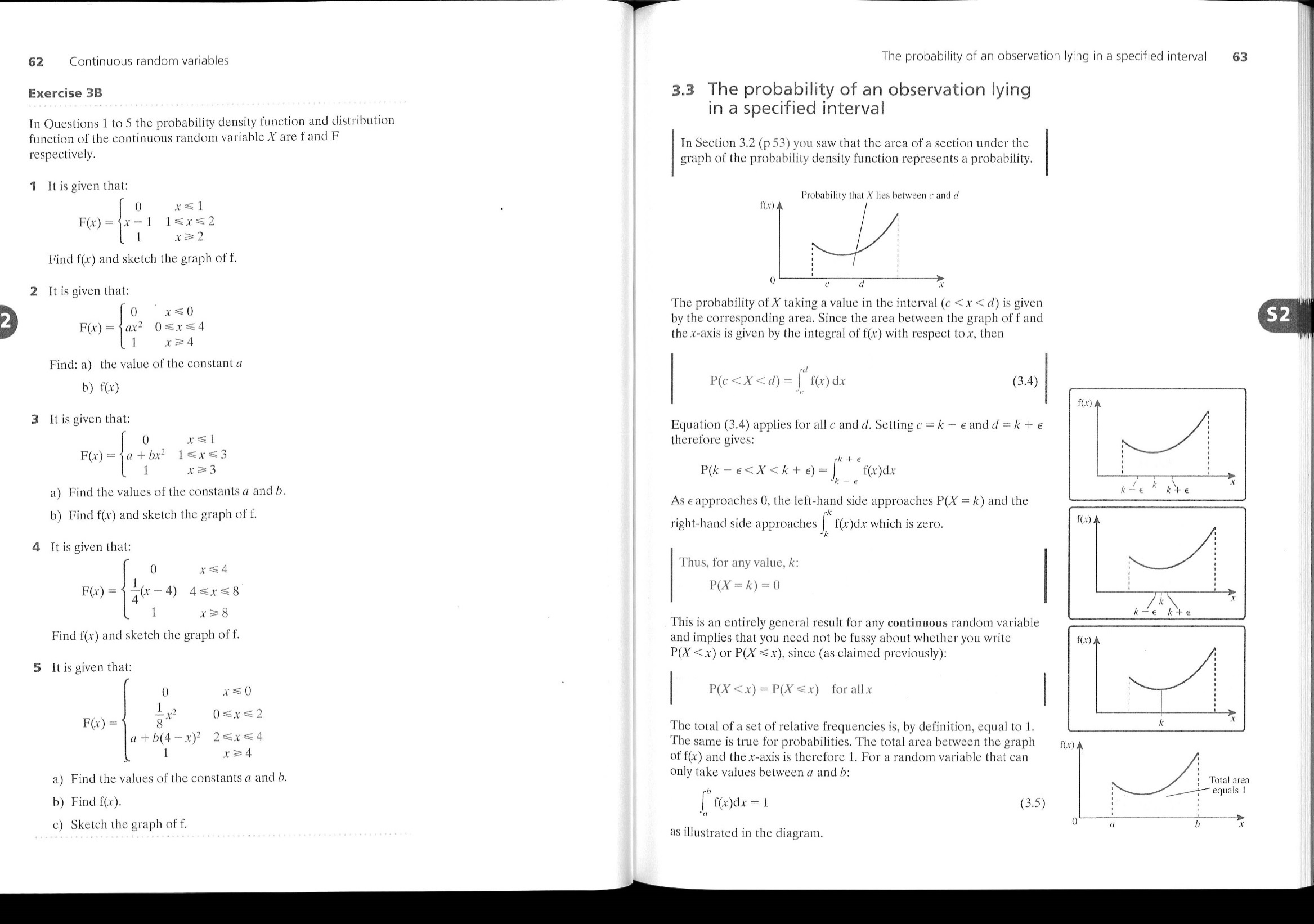
If and , then the distribution of .

**Continuous Random Variables**

|  |  |
| --- | --- |
| Probability Density Function (pdf) = *f* | Cumulative Distribution Function (cdf) = *F* |
|  |  |
|  |  |
|  |  |
|  | (replace lower limit with lower limit of function when calculating) |

Three methods to find specific probabilities of *X*:

1. (ie integrate to find area between two values)



1. Having integrated, evaluate between limits (similar to finding normal distribution probabilities).
2. Use geometry of graph of (where possible, splitting it into triangles, trapezia etc).

To find median value, *m*, solve for *m* either:

or

To find, for example, the value at 95th percentile, solve for d;

or

Mean and variance…

|  |  |  |
| --- | --- | --- |
|  |  |  |

If distribution is symmetrical then

|  |  |
| --- | --- |
| Positive Skew | Negative Skew |
| Right skewed | Left skewed |
| Mode < Median < Mean | Mean < Median < Mode |

Rectangular (Uniform) Distribution

|  |  |  |
| --- | --- | --- |
|  |  |  |

|  |  |
| --- | --- |
|  |  |

**Estimation**

The t distribution (two random variables, and , i.e. unknown population variance).

or

(parameter pronounced ‘nu’)

Confidence intervals given by

|  |  |
| --- | --- |
|  |  |

**Hypothesis Testing**

|  |  |
| --- | --- |
| Null Hypothesis | Alternative Hypothesis |
|  | |  |  | | --- | --- | |  |  | | One tailed | Two tailed | |

*Acceptance of null hypothesis does not mean it is true, rather that the data provides no evidence to prefer the alternative.*

Type 1 error – to reject H0 (and accept H1) when H0 is actually true.

Type 2 error – to reject H1 (and accept H0) when H1 is actually true.

Test Procedure:

1. Write down the two hypotheses.
2. Identify an appropriate test statistic and the distribution of the corresponding random variable.
3. Identify the significance level (usually given). This is also .
4. Determine the critical region (should be done before collecting data).
5. Calculate the value of the test statistic.
6. Determine *and clarify in context* the outcome of the test.

**Contingency Tables Tests**

**Part A – The Distribution**

|  |  |
| --- | --- |
|  |  |

(where )

(where = number of different possible outcomes)

All expected frequencies must be greater than 5

**Part B – Contingency Tables**

Associated vs independent

To calculate expected frequencies from observed frequencies use

Yates’s correction for a 2x2 table ( ):

Where…

|  |  |  |
| --- | --- | --- |
| a | b | m |
| c | d | n |
| r | s | N |