MathEVmatics

Supercharge your learning

A2 Applied Mathematics for WJEC

Unit 2a:

Review of Discrete Distributions;

Introduction to Continuous Distributions

Examples and Practice Exercise

Table of Contents

Where it fits:

Specification Reference: 2.4.2 (Statistical Distributions)

Topics	Guidance
2.4.2 Statistical distributions	
Understand and use the continuous uniform distribution and Normal distributions as models. Find probabilities using the Normal distribution. Link to histograms, mean, standard deviation, points of inflection and the binomial distribution.	Use of calculator / tables to find probabilities. Linear interpolation in tables will **not** be required.
Select an appropriate probability distribution for a context, with appropriate reasoning, including recognising when the continuous uniform or Normal model may not be appropriate.	The distributions from which the selection can be made are: Discrete: binomial, Poisson, uniform Continuous: Normal, uniform

Learning Objectives

- To recap on the discrete distributions met in AS Unit 2;
- To understand the idea of a continuous probability distribution, and the idea of a probability density function (PDF);
- To understand the contexts in which a Continuous Uniform Distribution is appropriate, and to be able to find probabilities;
- To understand and use the formulae for the mean and variance of the uniform distribution.

1. Discrete Distributions – Recap

In Unit 2, you met the idea of discrete statistical distributions. As well as gaining a general understanding of discrete distributions, you met three specific distributions. We will briefly review these below and connect them with our work on conditional probability from last week.

For Unit 4, it is crucial to recognise when it is appropriate to use each of these distributions, selecting a suitable distribution for a given context.

The "Big Three" Discrete Distributions:

We will review the three main discrete distributions:

1. **Discrete Uniform Distribution**
2. **Binomial Distribution**
3. **Poisson Distribution**

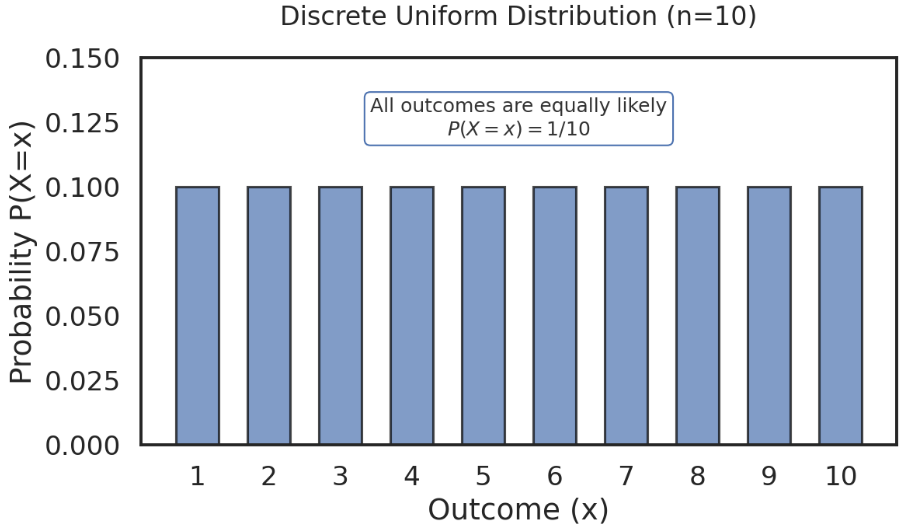
1.1 The Discrete Uniform Distribution

Definition

This distribution is used when a random variable X has a finite number of distinct outcomes, and every outcome is **equally likely**.

If the variable X can take values $\{x_1, x_2, \dots, x_n\}$, then:

$$P(X = x) = \frac{1}{n}$$



Example 1:

A bag contains 10 balls, identical apart from their colour. They are numbered 1,2,3, … ,10. Let X be the number on a ball selected at random.

(a) State the distribution of X.

Since the balls are identical, selecting any specific number is equally likely.

$$P(X = x) = \frac{1}{10} \text{ for } x = 1,2, \dots, 10$$

(b) Find $P(X$ is even$)$.

The even numbers are $\{2, 4, 6, 8, 10\}$. There are 5 outcomes.

$$P(X \text{ is even}) = \frac{5}{10} = 0.5$$

(c) Find $P(X < 4)$.

For discrete variables, the inequality type matters greatly. "Less than 4" does not include 4. The outcomes are $\{1, 2, 3\}$.

$$P(X < 4) = \frac{3}{10} = 0.3$$

(d) Find $P(3 \leq X < 7)$.

This includes 3, but does not include 7. The outcomes are $\{3, 4, 5, 6\}$.

$$P(3 \leq X < 7) = \frac{4}{10} = 0.4$$

(e) Find $P(X < 8|X \geq 2)$

Using our formula for conditional probability,

$$P(X < 8|X \geq 2) = \frac{P(X < 8 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(2 \leq X \leq 8)}{P(X \geq 2)}$$

$$= \frac{\frac{6}{10}}{\frac{9}{10}} = \frac{2}{3}$$

1.2 The Binomial Distribution

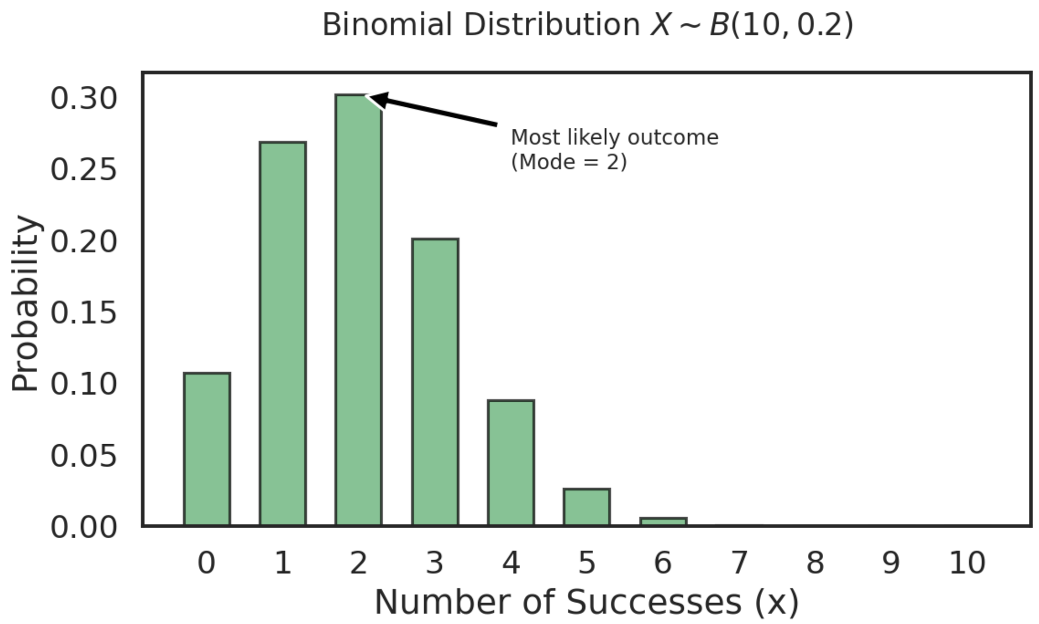
Definition

The Binomial distribution describes the number of "successes" in a fixed number of trials. It is denoted $X \sim B(n, p)$.

Assumptions (Conditions)

To use the Binomial model, **four** conditions must be met:

1. **Fixed number of trials (n):** The experiment is repeated a set number of times.
2. **Two outcomes:** Success or Failure.
3. **Fixed probability (p):** The probability of success remains constant for every trial.
4. **Independence:** The outcome of one trial does not affect the outcome of another.



Example 2:

A dart player throws a dart at a bullseye. She throws the dart 12 times. Based on her historical stats, the probability of her hitting the bullseye on any single throw is 0.2. Let X be the number of times she hits the bullseye.

(a) Suggest a distribution for X.
Assuming her throws are independent and her skill level doesn't change (probability remains constant):

$$X \sim B(12, 0.2)$$

(b) Find the probability she hits the bullseye exactly 3 times.
Using the formula

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

or the calculator **Binomial PD** mode:

$$P(X=3)=\binom{12}{3}(0.2)^3(0.8)^9=0.2362 \text{ (4 d.p.)}$$

(c) Find the probability she hits the bullseye fewer than 4 times.

"Fewer than 4" means 3, 2, 1, or 0. This is the cumulative probability up to 3.

$$P(X<4)=P(X\leq 3)$$

Using calculator **Binomial CD**:

$$P(X\leq 3)=0.7946$$

(d) Find the probability she hits the bullseye at least 5 times.

Tables and calculators usually calculate $P(X\leq x)$. We must use the complement rule.

$$P(X\geq 5)=1-P(X\leq 4)$$
$$P(X\geq 5)=1-0.9274=0.0726$$

1.3 The Poisson Distribution

Definition

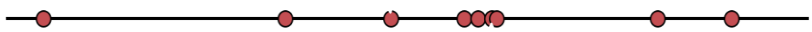
The Poisson distribution models the number of events occurring in a fixed interval (time, space, or volume). It is denoted $X \sim Po(\lambda)$, where is the mean average rate.

Assumptions (Conditions)

Think **SIRC**:

1. **Singly:** Events occur one at a time (they cannot occur simultaneously).
2. **Independently:** The occurrence of one event does not affect another.
3. **Randomly:** There is no pattern to when the events occur.
4. **Constant Rate:** The average number of events per interval (λ) remains constant.

Visualisation



The diagram represents a timeline, with the dots showing a random variable X occurring at random intervals along the line. Events occur randomly, independently and singly in time.

Example 3:

Patients arrive at a triage desk in a medical centre at a mean rate of 6 per hour. Let X be the number of patients arriving in a randomly selected hour.

(a) Suggest a distribution for X.

Assuming patients arrive independently and singly at a constant average rate:

$$X \sim Po(6)$$

(b) Find the probability that exactly 5 patients arrive.

Using the formula

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

or calculator **Poisson PD**:

$$P(X = 5) = \frac{e^{-6}6^5}{5!} = 0.1606$$

(c) Find the probability that more than 8 patients arrive.

$$P(X > 8) = 1 - P(X \leq 8)$$

Using calculator **Poisson CD**:

$$P(X > 8) = 1 - 0.8472 = 0.1528$$

(d) Find the probability that between 3 and 7 patients (inclusive) arrive.

We want the probability including 3 and including 7.

$$P(3 \leq X \leq 7) = P(X \leq 7) - P(X \leq 2)$$

Note: We subtract $P(X \leq 2)$ because we want to keep $X = 3$.

$$P(3 \leq X \leq 7) = 0.7440 - 0.0620 = 0.6820$$

(e) Find the probability that at least 4 patients arrive given that no more than 8 patients arrived.

We want the conditional probability

$$P(X \geq 4 | X \leq 8)$$

Using our formula for conditional probability,

$$P(X \geq 4 | X \leq 8) = \frac{P(X \geq 4 \cap X \leq 8)}{P(X \leq 8)}$$

Further, it should be easy to understand that $P(X \geq 4 \cap X \leq 8)$ is simply $P(4 \leq X \leq 8)$.

Therefore,

$$P(X \geq 4 | X \leq 8) = \frac{P(4 \leq X \leq 8)}{P(X \leq 8)}$$

$$P(4 \leq X \leq 8) = P(X \leq 8) - P(X \leq 4) = 0.8472 - 0.2851 = 0.5621$$

$$\therefore P(X \geq 4 | X \leq 8) = \frac{0.5621}{0.8472} = 0.6635$$

1.4 Discrete Inequalities Cheatsheet

Because discrete distributions count integers $(0,1,2, \ldots)$, strict inequalities ($<$) and inclusive inequalities ($\leq$) result in different calculations.

Calculators and Tables usually give **Cumulative Probability**:

$$P(X \leq x).$$

Desired Probability	**Calculation required**	**Logic**
"Less than x"		
$P(X < 5)$	$P(X \leq 4)$	5 is not included, so we stop at 4.
"At most x"		
$P(X \leq 5)$	$P(X \leq 5)$	5 is included. Use directly.
"More than x"		
$P(X > 5)$	$1 - P(X \leq 5)$	We want 6, 7, 8... remove 0 to 5.
"At least x"		
$P(X \geq 5)$	$1 - P(X \leq 4)$	We want 5, 6, 7... remove 0 to 4.
"Between a and b inclusive"		
$P(3 \leq X \leq 7)$	$P(X \leq 7) - P(X \leq 2)$	Take total up to 7, remove everything *before* 3.

---

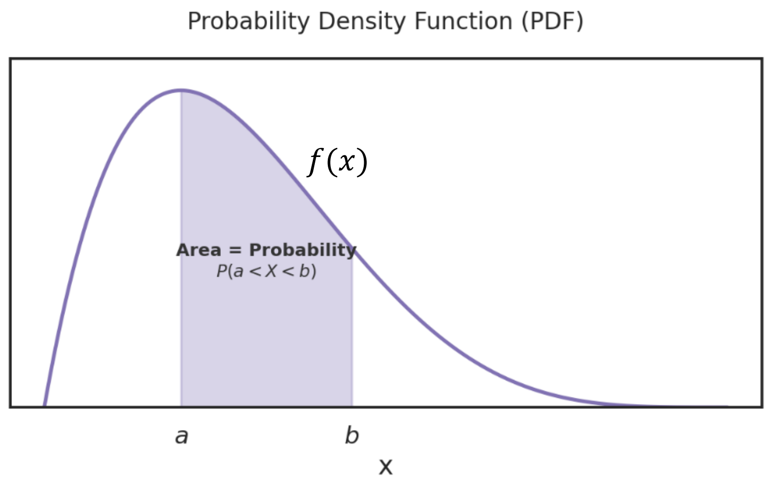
2. Continuous Distributions – Introduction

The Key Difference

In the discrete section above, we counted specific items (people, goals, hits).
In **Continuous Distributions**, we measure quantities (time, weight, height, speed). A continuous variable can take **any** value within an interval.

Probability Density Function (PDF)

A continuous distribution is defined by a function $f(x)$, called the Probability Density Function.



Key properties:

1. **Probability is Area:** The probability $P(a < X < b)$ is the area under the curve between a and b.
2. **Total Area is 1:** The total probability of all possible outcomes must sum to 1.

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

3. **The "Zero Probability" Concept**

This is the most important distinction for Year 13 students.

For a continuous random variable, the probability of obtaining any **exact** single value is zero.
e.g. $P(X = 5) = 0$

(Think about it: To measure exactly 5.000000... cm with infinite precision is impossible).

Consequence:
Strict or non-strict inequality symbols do not change the value of the probability for continuous variables.

$$P(X < 5) = P(X \leq 5)$$

$$P(2 < X < 5) = P(2 \leq X \leq 5)$$

Note: This is completely different from Discrete distributions where

$$P(X < 5)$$

and

$$P(X \leq 5)$$

are different values.

Example 4:

Suppose that T is a continuous random variable.

(a) Write down the value of $P(T = k)$ where k is a constant.
(b) Write down three probability expressions equivalent to $P(T \geq 6)$
(c) Write down five probability expressions equivalent to $P(3 \leq T \leq 7)$

(a) $P(T = k) = 0$

(b) Three expressions equivalent to $P(T \geq 6)$: (Remember that $P(T = 6) = 0$)

Strict inequality: $P(T > 6)$
Complement using strict inequality: $1 - P(T < 6)$
Complement using non-strict inequality: $1 - P(T \leq 6)$

(c) Five expressions equivalent to $P(3 \leq T \leq 7)$ – in fact here are six of them (more are possible)!

$$P(3 < T \leq 7)$$
$$P(3 \leq T < 7)$$
$$P(3 < T < 7)$$
$$P(T \leq 7) - P(T \leq 3)$$
$$P(T < 7) - P(T < 3)$$
$$1 - [P(T < 3) + P(T > 7)]$$

3. Exercise: Statistical Distributions Recap

Q1. A fair 12-sided die (dodecahedron) has faces numbered 1 to 12. Let X be the score obtained when the die is thrown once.
(a) State the distribution of X.
(b) Find $P(X > 8)$.
(c) Find $P(3 < X \leq 6)$.

Q2. A raffle ticket is selected at random from a large drum containing 200 tickets numbered 1 to 200.
(a) State the distribution used to model the number on the selected ticket.
(b) Let X denote the number on the ticket. Find $P(X < 120 \mid X \geq 100)$.

Q3. A factory produces lightbulbs. Historical data suggests that 5% of the lightbulbs produced are defective. A quality control officer selects a random sample of 20 bulbs for inspection.
(a) Suggest a distribution to model the number of defective bulbs in the sample.
(b) State one assumption necessary for this distribution to be valid.
(c) Find the probability that the sample contains exactly 1 defective bulb.
(d) Find the probability that the sample contains at least 2 defective bulbs.

Q4. James plays a carnival game where he throws a ball through a hoop. He has 10 attempts. He knows from experience that his probability of success on any single throw is 0.3.
(a) Write down the distribution for X, the number of successful throws.
(b) Comment on the validity of the assumption that successive trials are independent in this context.
(c) Find the probability that James succeeds at least 3 times, given that he succeeds at most 6 times.

Q5. An IT support desk receives emails requesting password resets. These emails arrive at a mean rate of 3.5 per hour.
(a) State the distribution used to model the number of emails, E, arriving in a one-hour period.
(b) Find the probability that the desk receives fewer than 3 emails in a randomly selected hour.
(c) Find the probability that the desk receives between 2 and 5 emails (inclusive) in a randomly selected hour.

Q6. The number of cars passing a "Welcome to Wales" sign in a 10-minute interval is modelled by a Poisson distribution with mean 30.
(a) Explain why a Poisson distribution might not be a perfect model for traffic flow (Think: independence).
(b) Let X denote the number of cars passing in a **5-minute** interval. Write down the distribution of X.
(c) Determine the probability that more than 20 cars pass in 5 minutes, given that no more than 25 cars passed in that interval.

Q7. The continuous random variable X has a probability density function $f(x)$ defined as a horizontal line between $x = 0$ and $x = 5$, and 0 elsewhere.
(a) Sketch the graph of $f(x)$.
(b) Using the fact that the total area must be 1, find the height of the line.
(c) Calculate $P(1 < X < 3)$ by calculating the relevant area under the graph.

Q8. Suppose Y is a continuous random variable.
(a) Write down the value of $P(Y = 10)$.
(b) Hence, explain why $P(Y \leq 10) = P(Y < 10)$.
(c) A student calculates a probability for a discrete distribution $B(10,0.5)$ and writes "As n is small, $P(X \leq 5) \approx P(X < 5)$". Explain the mistake in the student's reasoning.

---

4. Teacher Solutions for Exercise 1

Q1.

(a) Discrete Uniform $P(X = x) = \frac{1}{12}$.

(b) $P(X > 8) = P(X \in \{9,10,11,12\}) = \frac{4}{12} = \frac{1}{3}$.

(c) $P(3 < X \leq 6) = P(X \in \{4,5,6\}) = \frac{3}{12} = \frac{1}{4}$.

Q2.

(a) Discrete Uniform $X \sim U[1,200]$ (or, e.g. $P(X = x) = \frac{1}{200}$) with probability $\frac{1}{200}$.

(b)

$$P(X < 120 \mid X \geq 100) = \frac{P(X < 120 \cap X \geq 100)}{P(X \geq 100)}$$

Outcomes: $\{100, 101, 102, \dots, 119\}$ (20 outcomes).

Outcomes ≥ 100: $\{100...200\}$ (101 outcomes).
Probability =

$$\frac{20/200}{101/200} = \frac{20}{101}$$

Q3.

(a)

$$X \sim B(20,0.05)$$

(b) The defective status of one bulb must be independent of another (random sampling from a large batch).

(c)

$$P(X = 1) = \binom{20}{1}(0.05)^1(0.95)^{19} = 0.3774$$

(d)

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - 0.7358 = 0.2642$$

Q4.

(a)

$$X \sim B(10,0.3)$$

(b) Unlikely to be independent. James might improve with practice (learning effect) or get worse due to fatigue/frustration.

(c)

$$P(X \geq 3 \mid X \leq 6) = \frac{P(3 \leq X \leq 6)}{P(X \leq 6)}$$

$$P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 2) = 0.9894 - 0.3828 = 0.6066$$

$$P(X \leq 6) \approx 0.9894$$

Result

$$\approx 0.6131$$

Q5.

(a) $E \sim Po(3.5)$

(b) $P(E < 3) = P(E \leq 2) = 0.3208$

(c) $P(2 \leq E \leq 5) = P(E \leq 5) - P(E \leq 1) = 0.8576 - 0.1359 = 0.7217$

Q6.

(a) Independence: Cars often travel in groups/convoys. Constant rate: Traffic is heavier at rush hour than mid-day.

(b) Mean for 10 mins = 30. Mean for 5 mins = 15.

$$X \sim Po(15)$$

(c)

$$P(X > 20 \mid X \leq 25) = \frac{P(20 < X \leq 25)}{P(X \leq 25)}$$

Numerator:

$$P(X \leq 25) - P(X \leq 20) = 0.9938 - 0.9170 = 0.0768$$

Denominator:

$$0.9938$$

Result:

$$0.0768/0.9938 = 0.0773$$

Q7.

(a) Rectangle from $x = 0$ to $x = 5$.

(b) Area = width × height.

$$1 = 5 \times h \Longrightarrow h = 0.2$$

(c) Area from 1 to 3. Width = 2. Height = 0.2. Probability = $2 \times 0.2 = 0.4$.

Q8.

(a) 0

(b) Since $P(Y = 10) = 0$, including the value 10 in the inequality adds zero probability.

(c) For discrete distributions, $P(X = 5)$ is not zero. In $B(10,0.5)$, $P(X = 5) \approx 0.246$.

Therefore, $P(X \leq 5)$ is significantly larger than $P(X < 5)$.