

Probability & Statistics 2

A-Level Mathematics ■ Topic 6

A-Level Mathematics



What we will cover

- 1 The Poisson distribution
- 2 Combining random variables
- 3 Continuous random variables
- 4 Sampling & estimation
- 5 Hypothesis tests
- 6 Recap





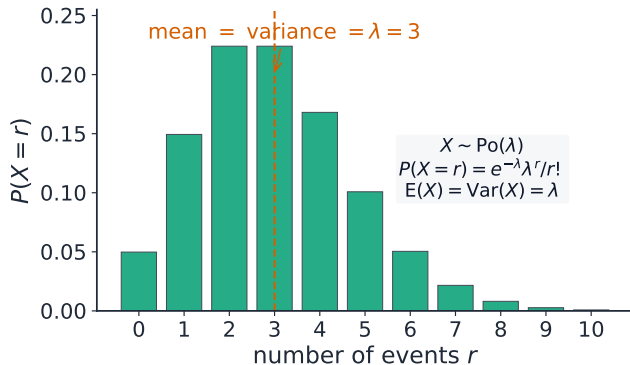
The Poisson distribution

Counts of rare, independent events in a fixed interval: $X \sim \text{Po}(\lambda)$,

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad E(X) = \text{Var}(X) = \lambda.$$

A binomial $X \sim B(n, p)$ with n large and p small is approximately Poisson with $\lambda = np$.

The Poisson distribution, in detail



The Poisson distribution $Po(3)$: for a Poisson variable the mean and variance both equ

Combining random variables

For a linear change and for **independent** variables:

$$E(aX + b) = aE(X) + b$$

$$\text{Var}(aX + bY) = a^2 \text{Var} X + b^2 \text{Var} Y$$

e.g. mean 5, variance 4: $E(3X - 1) = 14$, $\text{Var}(3X - 1) = 36$.



Continuous random variables

A **density function** $f(x) \geq 0$, $\int f = 1$.

$$P(a < X < b) = \int_a^b f, \quad E(X) = \int xf.$$

Median solves $F(m) = 0.5$.

The two functions of a continuous variable

Probability density function

The **probability density function** $f(x)$: probability is the *area* beneath it, and the total area is 1.

Cumulative distribution function

The **cumulative distribution function** $F(x) = P(X \leq x)$ – the integral of f , and differentiating F gives f back.

Expectation

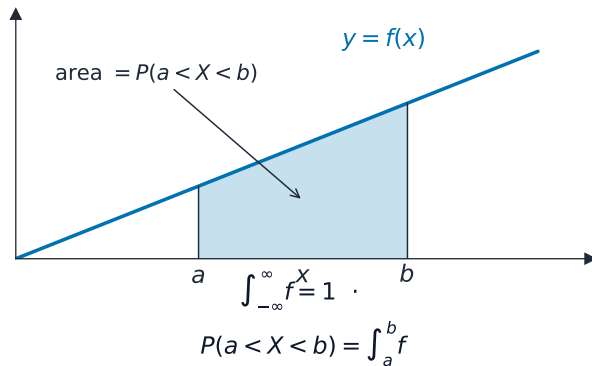
The **expectation** is $E(X) = \int xf(x) dx$ over the range where f is non-zero.

The normal distribution

The **normal distribution** is the case whose density is that symmetric bell – and the one every table is written for.

Medians and **percentiles** come from F , means and variances from f . Knowing which one a question needs saves most of the work.

Continuous random variables, in detail



A **density function** is never negative and encloses **unit area**: $f(x) \geq 0$ and $\int f = 1$. Probability is an **area**, so $P(X = a) = 0$ for any single value.

Sampling and estimation

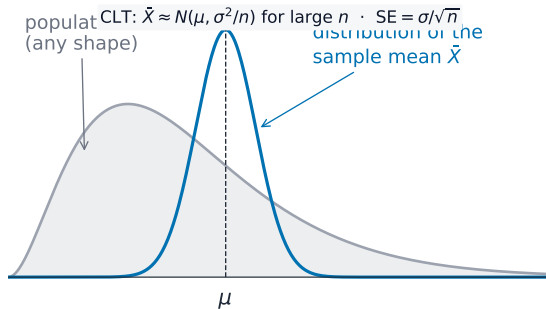
$E(\bar{X}) = \mu$, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$. A 95% **confidence interval** :

$$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}.$$

$$n = 64, \bar{x} = 50, \sigma = 8$$

$$50 \pm 1.96 \Rightarrow (48.0, 52.0)$$

Sampling and estimation, in detail



CLT : $\bar{X}$ near-normal

$$E(\bar{X}) = \mu, \text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

What makes a sample usable

a sample

a small group chosen from the whole **population**

a random sample

needs **randomness** : **every** member has an equal chance of being chosen

why it matters

only then do the sample statistics estimate the population's without systematic error

biased methods

sampling only volunteers, or the first 20 people to arrive – both over-represent a particular kind of person

the central limit theorem

for a large n the sample mean is **approximately normal**, whatever the population's distribution

the consequence

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ – which is what every mean-based test rests on

“Unsatisfactory” in a mark scheme means **biased**, and the mark is for naming *who* is over- or under-represented.



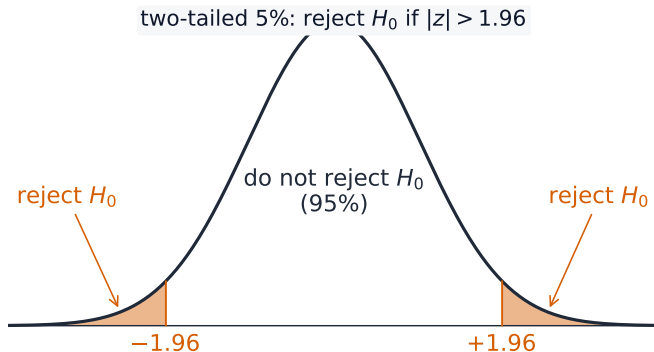
Hypothesis tests

Null H_0 vs **alternative** H_1 ; pick a **significance level** .

Test statistic in the **rejection region** $\Rightarrow$ reject H_0 .

Type I : reject a true H_0 ; **Type II** : accept a false H_0 .

Hypothesis tests, in detail



A two-tailed test at 5% rejects H_0 only if the test statistic falls in a shaded tail beyond ± 1.96 .

The parts of a hypothesis test

a hypothesis test

uses sample data to judge a claim

the null hypothesis

H_0 – the claim being tested, always an equality

the alternative

H_1 – what you conclude if H_0 is rejected; one-tailed or two-tailed

the significance level

often 5% – fixed **before** you look at the data

the test statistic

worked out from the data, and compared with the critical region

the conclusion

reject or do not reject H_0 – **in context**, and never “accept H_0 ”

Not rejecting is not proving. The test can only say the evidence was insufficient, which is why the wording of the conclusion is marked.

The parts of a test, and the two ways it fails

The hypotheses

H_0 states the value being tested; the **alternative hypothesis** H_1 states what you would accept instead, and its form decides one tail or two.

What is estimated

A sample gives **unbiased estimates** of the mean and variance – and, for a proportion, of the **population proportion**.

Type I error

A **type I error** : rejecting H_0 when it was true. Its probability *is* the significance level.

Type II error

A **type II error** : failing to reject H_0 when it was false. Lowering the significance level raises this one.

You never prove H_0 . The conclusion is either “sufficient evidence to reject” or “insufficient evidence” – in the context of the question.

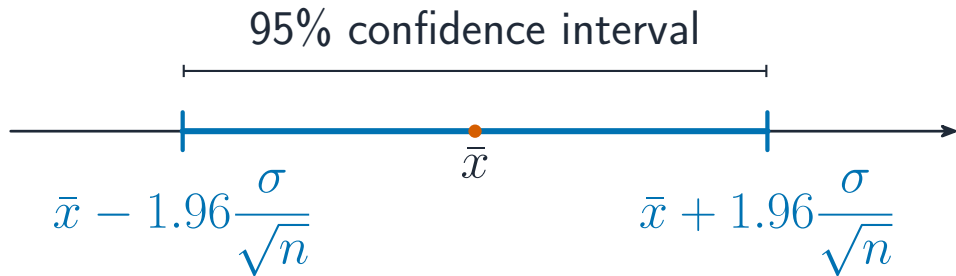
Worked example – a Poisson probability

$X \sim \text{Po}(3)$. Find $P(X = 2)$.

$$P(X = 2) = e^{-3} \frac{3^2}{2!} = e^{-3} \times 4.5 = 0.224$$

For a Poisson variable the mean **and** variance both equal λ . Use $P(X = r) = e^{-\lambda} \frac{\lambda^r}{r!}$.

Sampling and estimation, continued



A 95% confidence interval stretches 1.96 standard errors each side of the sample mean.

The Poisson distribution, continued



People arriving at random in a queue follow a Poisson distribution.

Exam tips

- Use the **Poisson** distribution for rare, random, independent events; its mean equals its variance ($= \lambda$).
- When combining independent random variables, **variances add** (they never subtract).
- For a **hypothesis test**, state H_0 and H_1 , the significance level, the test statistic, and a conclusion **in context**.
- For a **confidence interval**, use the correct z (or t) value and interpret it in words.



Topic 6 – key takeaways

1

Poisson

$\text{Po}(\lambda)$; mean = variance = λ

2

Combining

Var adds for independent variables

3

Sampling

CLT; $\bar{x} \pm 1.96\sigma/\sqrt{n}$

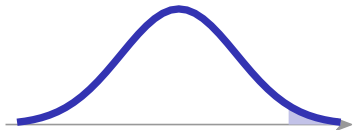
4

Testing

H_0 vs H_1 ; reject in the tail

Check yourself

- 1 Mean and variance of $Po(\lambda)$?
- 2 $\text{Var}(2X)$ if $\text{Var}(X) = 5$?
- 3 What does the area under $f(x)$ give?
- 4 Rejecting a true H_0 is which error?



Answers 1. both λ 2. $4 \times 5 = 20$ 3. a probability 4. Type I