Random Events at a Constant Average Rate

Many quantities count *random events* occurring through time or space:

- the number of calls arriving at a call centre in a minute;
- the number of misprints on a page of a newspaper;
- the number of radioactive decays detected in a second;
- the number of flaws in a metre of cloth.

In each case there is no fixed number of "trials" — a minute is not 20 coin flips — so the binomial does not apply directly. But the events happen *independently* of one another, at a *constant average rate*. The distribution governing such counts is the Poisson distribution.

The Poisson Distribution

Definition. A random variable X has the Poisson distribution with parameter $\lambda > 0$, written $X \sim \mathrm{Po}(\lambda)$, if

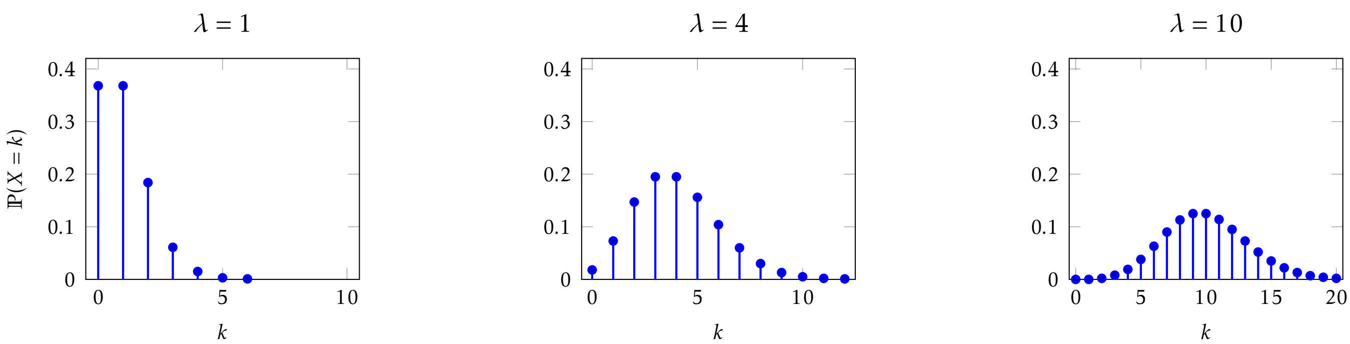
$$\mathbb{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k \in \{0, 1, 2, \ldots\}.$$

Here X counts the number of events in a given interval, and λ is the average number of events in an interval of that length.

A quick sanity check that the probabilities sum to 1 uses the Maclaurin series of the exponential:

$$\sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1.$$

The shape of the distribution depends on λ: for small λ it is sharply skewed with most mass near 0; as λ grows it becomes increasingly symmetric and bell-shaped about λ.



The modelling assumptions

Fact (Conditions for a Poisson model) — A count of events may be modelled by a Poisson distribution if:

1. events occur **independently** of one another;
2. events occur at a **constant average rate** (proportional to the length of the interval).

In an exam you must state these *in the context of the question* — e.g. "calls arrive independently of one another" and "calls arrive at a constant average rate".

Remark. The word **average** in assumption 2 is essential, and omitting it loses marks. "Calls arrive at a constant rate" would describe a perfectly regular, predictable process — one call exactly every 20 seconds — in which the number of calls in a minute is not random at all. The Poisson model says the *long-run average* rate is constant while the individual arrivals remain random.

Strictly we also require that events cannot occur simultaneously, but exam questions are not looking for this; in context it is usually subsumed by independence.

Example

A bridal shop sells, on average, 3 wedding dresses per week. The manager models the number of dresses sold in a week by a Poisson distribution. State, in context, two assumptions the manager is making.

- *Sales of wedding dresses occur independently of one another (one customer's purchase does not affect another's).*
- *Wedding dresses are sold at a constant* average *rate (the mean number of sales per week does not change, e.g. no seasonal wedding rush).*

Both assumptions are questionable here — weddings are seasonal, and friends may shop together — and exam questions often ask you to criticise the model in exactly this way.

Mean and Variance

Theorem

If $X \sim \text{Po}(\lambda)$ then

$$\mathbb{E}[X] = \lambda \qquad \text{and} \qquad \text{Var}[X] = \lambda.$$

The mean and variance of a Poisson random variable are *equal*.

The derivation is a nice piece of series manipulation, with two ideas worth remembering: cancel k into $k!$, and compute $\mathbb{E}[X(X-1)]$ rather than $\mathbb{E}[X^2]$.

For the mean, the $k=0$ term vanishes, and we cancel k into $k!$:

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \lambda e^{-\lambda}\sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda.$$

For the variance it is easiest to compute $\mathbb{E}[X(X-1)]$ first, cancelling $k(k-1)$ into $k!$:

$$\mathbb{E}[X(X-1)] = \sum_{k=2}^{\infty} k(k-1)\frac{e^{-\lambda}\lambda^k}{k!} = \lambda^2 e^{-\lambda}\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2 e^{-\lambda} \cdot e^{\lambda} = \lambda^2,$$

so $\mathbb{E}[X^2] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] = \lambda^2 + \lambda$, and

$$Var[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \lambda^2 + \lambda - \lambda^2 = \lambda.$$

Tip

"Mean = variance" is a quick diagnostic for real data: if the sample mean and sample variance of count data are close, a Poisson model is plausible; if the variance is much larger than the mean, it is not.

Example

A biologist counts the number of insect larvae in each of 100 equal-sized soil samples:

Number of larvae	0	1	2	3	4	5
Frequency	26	35	24	10	4	1

By calculating the sample mean and variance, comment on whether a Poisson model is plausible for these data.

Sample mean:

$$\bar{x} = \frac{0(26)+1(35)+2(24)+3(10)+4(4)+5(1)}{100} = \frac{134}{100} = 1.34.$$

For the variance, $\sum f x^2 = 0 + 35 + 96 + 90 + 64 + 25 = 310$, so

$$s^2 \approx \frac{310}{100} - 1.34^2 = 3.10 - 1.7956 = 1.30 \text{ (3 s.f.)}.$$

The sample mean (1.34) and variance (1.30) are very close, which is consistent with a Poisson model: larvae appear to be scattered randomly and independently through the soil at a constant average rate. (If larvae clustered together — eggs laid in batches — we would expect the variance to noticeably exceed the mean.)

Example (OCR Further Stats, June 2024 (parts))

Some bird-watchers study the song of chaffinches in a particular wood. They investigate whether the number, N, of separate bursts of song in a 5 minute period can be modelled by a Poisson distribution. They assume that a burst of song can be considered as a single event, and that bursts of song occur randomly.

(a) State two further assumptions needed for N to be well modelled by a Poisson distribution.

(b) The bird-watchers record the value of N in each of 60 periods of 5 minutes. The mean and variance of the results are 3.55 and 5.6475 respectively. Explain what this suggests about the validity of a Poisson distribution as a model in this context.

(c) It is known that chaffinches are more likely to sing in the presence of other chaffinches. Explain whether this fact affects the validity of a Poisson model for N.

(a) Bursts of song occur independently of one another; bursts of song occur at a constant average rate.

(b) For a Poisson model the mean and variance should be (approximately) equal. Here the variance (5.6475) is substantially greater than the mean (3.55), which casts doubt on the validity of a Poisson model.

(c) If chaffinches are more likely to sing when other chaffinches are singing, then bursts of song are not *independent of one another — one burst makes another more likely. This violates the independence assumption, so the Poisson model is not valid. (It also explains the overdispersion in (b): clustering inflates the variance.)*

Deriving the Poisson from the Binomial

Where does the formula $\frac{e^{-\lambda}\lambda^k}{k!}$ come from? The key idea: a Poisson process is a limit of binomials. Suppose events occur at an average rate of λ per unit of time. Chop the time interval into n tiny slices. If n is large enough, each slice contains at most one event, and the count of events becomes a binomial: n slices, each a "success" (contains an event) independently with probability $\frac{\lambda}{n}$, so that the mean number of events is $n \cdot \frac{\lambda}{n} = \lambda$, as it should be.

So consider $X_n \sim B\left(n, \frac{\lambda}{n}\right)$ and let $n \to \infty$.

$$
\begin{aligned}
\mathbb{P}(X_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{n(n-1)\cdots(n-k+1)}{k!} \cdot \frac{\lambda^k}{n^k} \cdot \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \cdot \underbrace{\frac{n}{n} \cdot \frac{n-1}{n} \cdots \frac{n-k+1}{n}}_{\to 1} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{\to e^{-\lambda}} \cdot \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-k}}_{\to 1} \\
&\to \frac{\lambda^k e^{-\lambda}}{k!} \quad \text{as } n \to \infty.
\end{aligned}
$$

The middle limit $\left(1 - \frac{\lambda}{n}\right)^n \to e^{-\lambda}$ *is the famous compound-interest limit, which you can verify with the Maclaurin series of* $\ln(1+x)$.

> **Remark.** Historically the Poisson, with $\lambda = np$, was used as a numerical *approximation* to $B(n,p)$ for large n and small p — exactly the limit above. Your calculator computes exact binomial probabilities, so the approximation is obsolete in practice; it is the limiting relationship itself that matters.

Calculating Poisson Probabilities

You are expected to have a calculator that gives Poisson probabilities. On the Casio fx-991EX: MENU → Distribution, then

- **Poisson PD** gives $\mathbb{P}(X = x)$;
- **Poisson CD** gives the cumulative probability $\mathbb{P}(X \le x)$.

You should still be fluent with the formula $\mathbb{P}(X = k) = \dfrac{e^{-\lambda}\lambda^k}{k!}$: questions with unknown λ or unknown k expect you to work with it algebraically.

Tip (The classic discrete pitfall)

The Poisson distribution is **discrete**, so strict and non-strict inequalities differ:

$$\mathbb{P}(X < k) = \mathbb{P}(X \le k-1), \qquad \mathbb{P}(X \ge k) = 1 - \mathbb{P}(X \le k-1).$$

Translate every probability into the form $\mathbb{P}(X \le \square)$ *before* reaching for Poisson CD.

Example

Calls arrive at a helpline at an average rate of 3 per minute. Find the probability that, in a randomly chosen minute,

(a) exactly 5 calls arrive;

(b) fewer than 2 calls arrive;

(c) at least 4 calls arrive.

Let X be the number of calls arriving in one minute; then $X \sim Po(3)$, assuming calls arrive independently at a constant average rate.

(a) $\mathbb{P}(X = 5) = \dfrac{e^{-3}3^5}{5!} = \dfrac{243e^{-3}}{120} = 0.101$ (3 *s.f.*)

(b) $\mathbb{P}(X < 2) = \mathbb{P}(X \le 1) = e^{-3}(1+3) = 4e^{-3} = 0.199$ (3 *s.f.*)

(c) $\mathbb{P}(X \ge 4) = 1 - \mathbb{P}(X \le 3) = 1 - e^{-3}\left(1 + 3 + \frac{9}{2} + \frac{27}{6}\right) = 1 - 13e^{-3} = 0.353$ (3 *s.f.*)

Scaling the rate with the interval

If events occur at average rate λ per unit interval, then the count in an interval of length t units is Po(λt): the parameter scales with the length of the interval (this was assumption 2 at work). Always *redefine* the variable when the interval changes.

Example

A weaver finds flaws in cloth at an average rate of 0.8 flaws per metre.

(a) Find the probability of exactly 2 flaws in a 2.5 metre length.

(b) Find the probability of no flaws in a 50 centimetre length.

(c) Find the length of cloth, to the nearest centimetre, for which the probability of it containing no flaws is 0.9.

(a) Let X be the number of flaws in 2.5 metres; $X \sim Po(0.8 \times 2.5) = Po(2)$.

$$\mathbb{P}(X = 2) = \frac{e^{-2}\,2^2}{2!} = 2e^{-2} = 0.271 \text{ (3 s.f.)}$$

(b) Let Y be the number of flaws in 0.5 metres; $Y \sim Po(0.4)$, so $\mathbb{P}(Y = 0) = e^{-0.4} = 0.670$ (3 s.f.).

(c) For length t metres, the number of flaws is $Po(0.8t)$ and we need $e^{-0.8t} = 0.9$. Taking logarithms, $-0.8t = \ln 0.9$, so

$$t = \frac{\ln 0.9}{-0.8} = 0.1317\ldots \approx 13\,\text{cm}.$$

Example

$X \sim \text{Po}(\lambda)$ satisfies $\mathbb{P}(X = 3) = \mathbb{P}(X = 2)$. Find λ, and hence find $\mathbb{P}(X \geq 2)$.

$$\frac{e^{-\lambda}\lambda^3}{3!} = \frac{e^{-\lambda}\lambda^2}{2!} \implies \frac{\lambda^3}{6} = \frac{\lambda^2}{2} \implies \lambda = 3 \quad (\lambda \neq 0).$$

Then $\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X \leq 1) = 1 - e^{-3}(1 + 3) = 1 - 4e^{-3} = 0.801$ (3 s.f.).

Example (In class)

The number of letters delivered to an office in a day is modelled by Po(6.5). Find the smallest number n such that the probability of receiving more than n letters in a day is less than 1%.

The Normal Approximation to the Poisson

Look back at the diagrams on page 2: as λ grows, the Poisson distribution becomes symmetric and bell-shaped. This is no accident — the Poisson is a limit of binomials, and large-λ Poissons behave like large-n binomials, which are approximately normal.

Fact — If $X \sim \text{Po}(\lambda)$ with λ sufficiently large (say $\lambda > 15$ or so), then approximately

$$X \approx N(\lambda, \lambda),$$

matching the mean λ and variance λ. Since we are approximating a discrete distribution by a continuous one, a continuity correction is needed: e.g.

$$\mathbb{P}(X \leq 90) \approx \mathbb{P}(W < 90.5), \qquad \mathbb{P}(X \geq 90) \approx \mathbb{P}(W > 89.5),$$

where $W \sim N(\lambda, \lambda)$.

Example

A website receives visits at an average rate of 100 per hour. Use a normal approximation to estimate the probability of at most 90 visits in a given hour.

Let X be the number of visits in an hour, $X \sim Po(100)$. Approximate $X \approx W \sim N(100, 100)$, so W has standard deviation 10. With the continuity correction,

$$\mathbb{P}(X \leq 90) \approx \mathbb{P}(W < 90.5) = \mathbb{P}\left(Z < \frac{90.5 - 100}{10}\right) = \mathbb{P}(Z < -0.95) = 0.171 \text{ (3 } s.f.).$$

(The exact Poisson value is 0.1714, so the approximation is excellent here.)

Sums of Independent Poisson Variables

Theorem

If $X \sim \text{Po}(\lambda)$ and $Y \sim \text{Po}(\mu)$ are *independent*, then

$$X + Y \sim \text{Po}(\lambda + \mu).$$

This is entirely believable: if calls arrive on line A at average rate λ per minute and independently on line B at rate μ per minute, then calls overall arrive randomly and independently at average rate $\lambda + \mu$ per minute. We can check the mean and variance are consistent using expectation algebra:

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] = \lambda + \mu, \qquad \text{Var}[X+Y] = \text{Var}[X] + \text{Var}[Y] = \lambda + \mu \quad \text{(independence)},$$

so $X + Y$ has equal mean and variance, exactly as a $\text{Po}(\lambda + \mu)$ variable should. This *suggests* the result but does not prove it (plenty of non-Poisson distributions could have mean equal to variance). The proof falls out in one line from probability generating functions, in the next set of notes.

Remark (Common error: $2X$ is not Poisson)**.** If $X \sim \text{Po}(\lambda)$, it is tempting to write $X + X \sim \text{Po}(2\lambda)$. This is **wrong**: $X + X = 2X$ is a single observation doubled, and X is not independent of itself. Indeed

$$\mathbb{E}[2X] = 2\lambda \quad \text{but} \quad \text{Var}[2X] = 4\lambda,$$

and since the mean and variance differ, $2X$ cannot have a Poisson distribution at all (it only takes even values, for that matter). The theorem needs two *independent* variables: if $X_1, X_2 \sim \text{Po}(\lambda)$ are independent then $X_1 + X_2 \sim \text{Po}(2\lambda)$.

Example

A newspaper has misprints in its news pages at an average rate of 2.5 per page, and independently in its sport pages at an average rate of 1.5 per page.

(a) Find the probability that one news page and one sport page contain exactly 3 misprints in total.

(b) Find the probability that four news pages contain at most 5 misprints in total.

(a) Let $N \sim Po(2.5)$ and $S \sim Po(1.5)$ be the numbers of misprints on the two pages. As N and S are independent, $T = N + S \sim Po(4)$, so

$$\mathbb{P}(T = 3) = \frac{e^{-4}4^3}{3!} = \frac{32}{3}e^{-4} = 0.195 \text{ (3 s.f.)}.$$

(b) The counts on the four news pages are independent $Po(2.5)$ variables, so their total is $U \sim Po(10)$. By calculator (Poisson CD), $\mathbb{P}(U \le 5) = 0.0671$ (3 s.f.).

Example (OCR Further Stats, June 2022)

The manager of a car breakdown service uses the distribution Po(2.7) to model the number of punctures, R, in a 24-hour period in a given rural area. The manager knows that, for this model to be valid, punctures must occur randomly and independently of one another.

(a) State a further assumption needed for the Poisson model to be valid.

(b) State the value of the standard deviation of R.

(c) Use the model to calculate the probability that, in a randomly chosen period of 168 hours, at least 22 punctures occur.

(d) The manager uses the distribution Po(0.8) to model the number of flat batteries in a 24-hour period in the same rural area, and he assumes that instances of flat batteries are independent of punctures. A day begins and ends at midnight, and a "bad" day is a day on which there are more than 6 instances, in total, of punctures and flat batteries. Assuming both the manager's models are correct, calculate the probability that a randomly chosen day is a "bad" day.

(e) It is found that 12 of the next 100 days are "bad" days. Comment on whether this casts doubt on the validity of the manager's models.

(a) Punctures occur at a constant average *rate.*

(b) $Var[R] = 2.7$*, so the standard deviation is* $\sqrt{2.7} = 1.64$ *(3 s.f.).*

(c) 168 hours is 7 days, so the count is $W \sim Po(7 \times 2.7) = Po(18.9)$*. Then*

$$\mathbb{P}(W \geq 22) = 1 - \mathbb{P}(W \leq 21) = 0.267 \text{ (3 s.f.)}.$$

(d) The total count in a day is the sum of independent $Po(2.7)$ *and* $Po(0.8)$ *variables, so* $T \sim Po(3.5)$*, and*

$$\mathbb{P}(T > 6) = 1 - \mathbb{P}(T \leq 6) = 0.0653 \text{ (3 s.f.)}.$$

(e) Under the models we would expect about $100 \times 0.0653 \approx 6.5$ *bad days in 100; observing 12 is nearly double the expected number, which casts some doubt on the validity of the models (e.g. the rates may be higher than modelled, or instances may not be independent).*

Example (Edexcel FS1, June 2022 (parts))

During the summer, mountain rescue team A receives calls for help randomly with a rate of 0.4 per day.

(a) Find the probability that during the summer, team A receives at least 19 calls for help in 28 randomly selected days.

(b) During the summer, mountain rescue team B receives calls for help randomly with a rate of 0.2 per day, independently of calls to team A. The random variable C is the total number of calls for help received by teams A and B during a period of n days in the summer. Given that $\mathbb{P}(C = 0) < 0.001$, calculate the minimum value of n.

(c) Write down an assumption that needs to be made for the model in (b) to be appropriate.

(a) Over 28 days the count is $X \sim Po(28 \times 0.4) = Po(11.2)$, so

$$\mathbb{P}(X \geq 19) = 1 - \mathbb{P}(X \leq 18) = 0.0208 \text{ (3 s.f.)}.$$

(b) Over n days, calls to A and B are independent $Po(0.4n)$ and $Po(0.2n)$, so $C \sim Po(0.6n)$ and $\mathbb{P}(C = 0) = e^{-0.6n}$. We need

$$e^{-0.6n} < 0.001 \iff 0.6n > \ln 1000 \iff n > 11.5127\ldots,$$

so the minimum value is $n = 12$. (Check: $e^{-6.6} = 0.00136 > 0.001$ but $e^{-7.2} = 0.00075 < 0.001$.)

(c) Calls to the two teams are independent of each other (and the rates remain constant across the period).

Example (In class)

Which other distributions add nicely? If $X \sim B(n,p)$ and $Y \sim B(m,p)$ are independent, what is the distribution of $X + Y$? What goes wrong if the success probabilities differ? What about geometric distributions?

Textbook Exercises: [CUP.S] Ch 3; [S2] Ch 3; [S3/4] S3 Ch 2 §2.5