Random Variables

Distributions, Moments, Transformations, and Limit Theorems

What Is a Random Variable?

Definition — Probability Space

A **probability space** is a triple $(\Omega, \mathcal{F}, P)$ where

- Ω is the sample space
- $\mathcal{F}$ is a σ-algebra on Ω
- $P : \mathcal{F} \to [0, 1]$ is a probability measure

Definition — Random Variable

A **random variable** X is a function that maps outcomes from a sample space Ω to real numbers:

$$X : \Omega \to \mathbb{R}$$

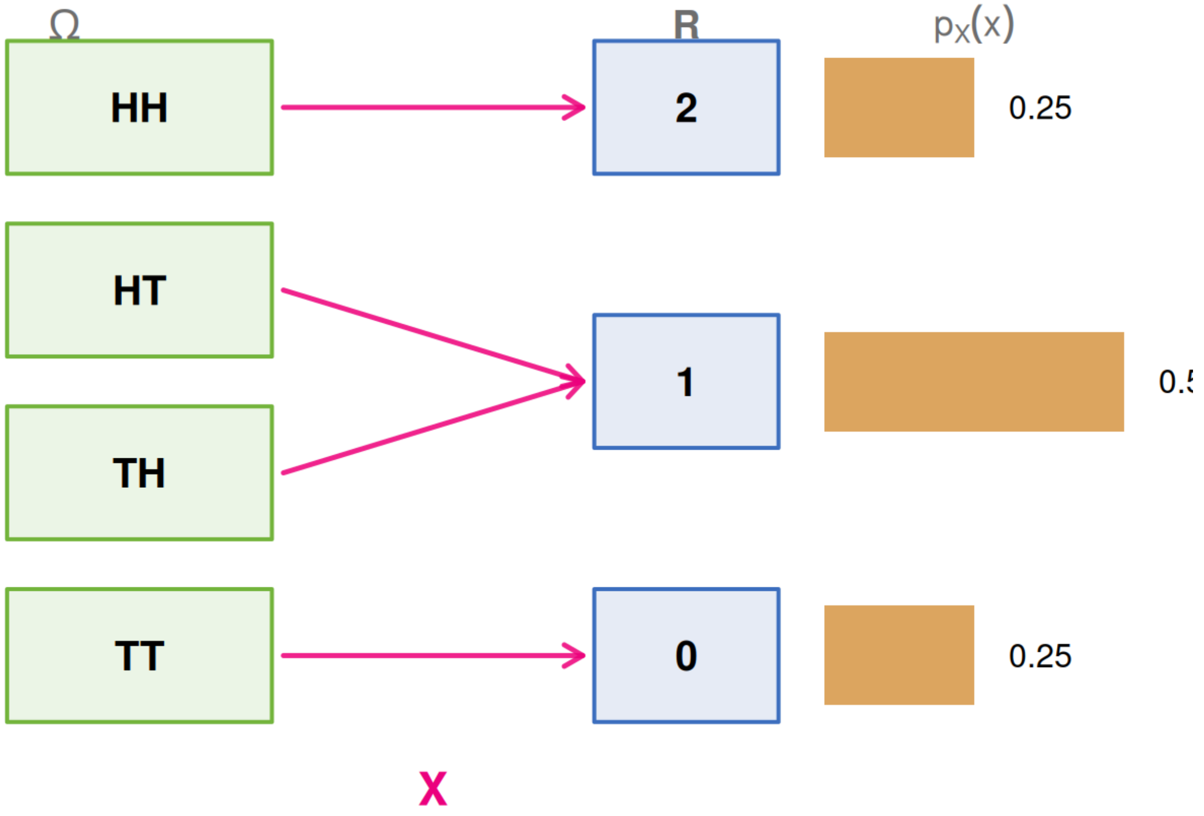
Key Insight

Random variables let us work with **numerical quantities** in probability theory — arithmetic, averages and calculus become available.

Random Variable Example

Example: tossing two coins.

- Sample space: $\Omega = \{HH, HT, TH, TT\}$
- Let X = number of heads
- $X(HH) = 2$
- $X(HT) = X(TH) = 1$
- $X(TT) = 0$



Cumulative Distribution Function (CDF)

Definition — CDF

The **cumulative distribution function** of a random variable X is

$$F_X(x) = P(X \leq x) \quad \text{for } x \in \mathbb{R}$$

Properties of the CDF:

(1) $0 \leq F_X(x) \leq 1$

(2) F_X is non-decreasing

(3) F_X is right-continuous: $\lim_{h \to 0^+} F_X(x+h) = F_X(x)$

Why the CDF matters

It is the one object that exists for **every** random variable — discrete, continuous, or a mixture of both. PMFs and PDFs are what the CDF looks like in two special cases.

Types of Random Variables

Discrete Random Variables

Definition — Discrete Random Variable

A random variable X is **discrete** if it takes values in a countable set $\{x_1, x_2, x_3, \ldots\}$.

Probability Mass Function (PMF):

$$p_X(x) = P(X = x)$$

Properties:

(1) $p_X(x) \geq 0$ for all x

(2) $\sum_x p_X(x) = 1$

(3) $P(X \in A) = \sum_{x \in A} p_X(x)$

Continuous Random Variables

Definition — Continuous Random Variable

A random variable X is **continuous** if there exists a non-negative function f_X such that

$$P(a \leq X \leq b) = \int_a^b f_X(x)\,dx$$

Probability Density Function (PDF):

$$f_X(x) = \frac{d}{dx} P(X \leq x)$$

Properties:

(1) $f_X(x) \geq 0$ for all x

(2) $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$

(3) $P(X = x) = 0$ for any single point x

Relationship Between CDF and PMF/PDF

Discrete case:

$$F_X(x) = \sum_{t \le x} p_X(t)$$

$$p_X(x) = F_X(x) - F_X(x^-)$$

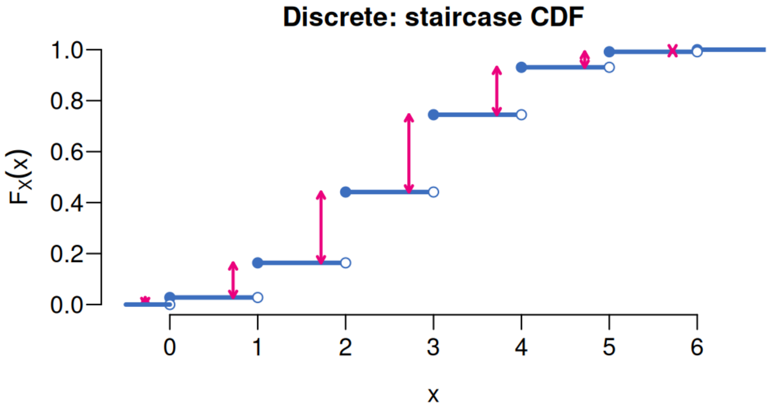
The CDF is a **staircase**; the jump heights *are* the probabilities.

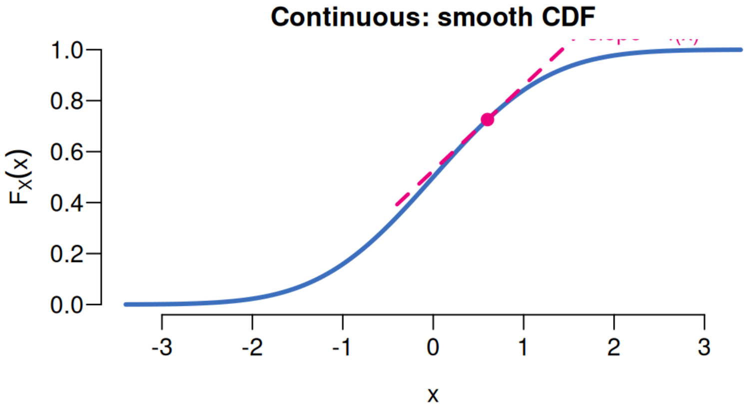
Continuous case:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

$$f_X(x) = \frac{d}{dx} F_X(x)$$

The CDF is **smooth**; the density is its slope.





Moments

Expected Value

Definition — Expected Value

The **expected value** (or mean) of a random variable X is

Discrete: $E[X] = \sum_x x \cdot p_X(x)$ **Continuous:** $E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)\, dx$

Properties:

- Linearity: $E[aX + bY] = aE[X] + bE[Y]$
- If $X \geq 0$, then $E[X] \geq 0$
- $E[c] = c$ for any constant c

Intuition

$E[X]$ is the **balance point** of the distribution: put the density on a seesaw, and this is where it tips level.

Variance and Standard Deviation

Definition — Variance

$$\mathrm{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2$$

Definition — Standard Deviation

$$\sigma_X = \sqrt{\mathrm{Var}(X)}$$

Properties:

- $\mathrm{Var}(X) \geq 0$
- $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$ — shifting does nothing, scaling squares
- If X and Y are independent: $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$

More Generally: Moments

Definition — Raw and Central Moments

The k-th raw moment is $\mu'_k = E[X^k]$.

The k-th central moment is $\mu_k = E[(X - E[X])^k]$.

Important special cases:

- $\mu'_1 = E[X]$ — the mean
- $\mu_2 = \mathrm{Var}(X)$ — the variance
- μ_3 measures **skewness**: $\gamma_1 = \dfrac{\mu_3}{\sigma^3}$
- μ_4 measures **kurtosis**: $\gamma_2 = \dfrac{\mu_4}{\sigma^4} - 3$ (excess kurtosis)

Moments, Live

Two knobs, four numbers. Watch which moments react to which knob.

Skew knob ε

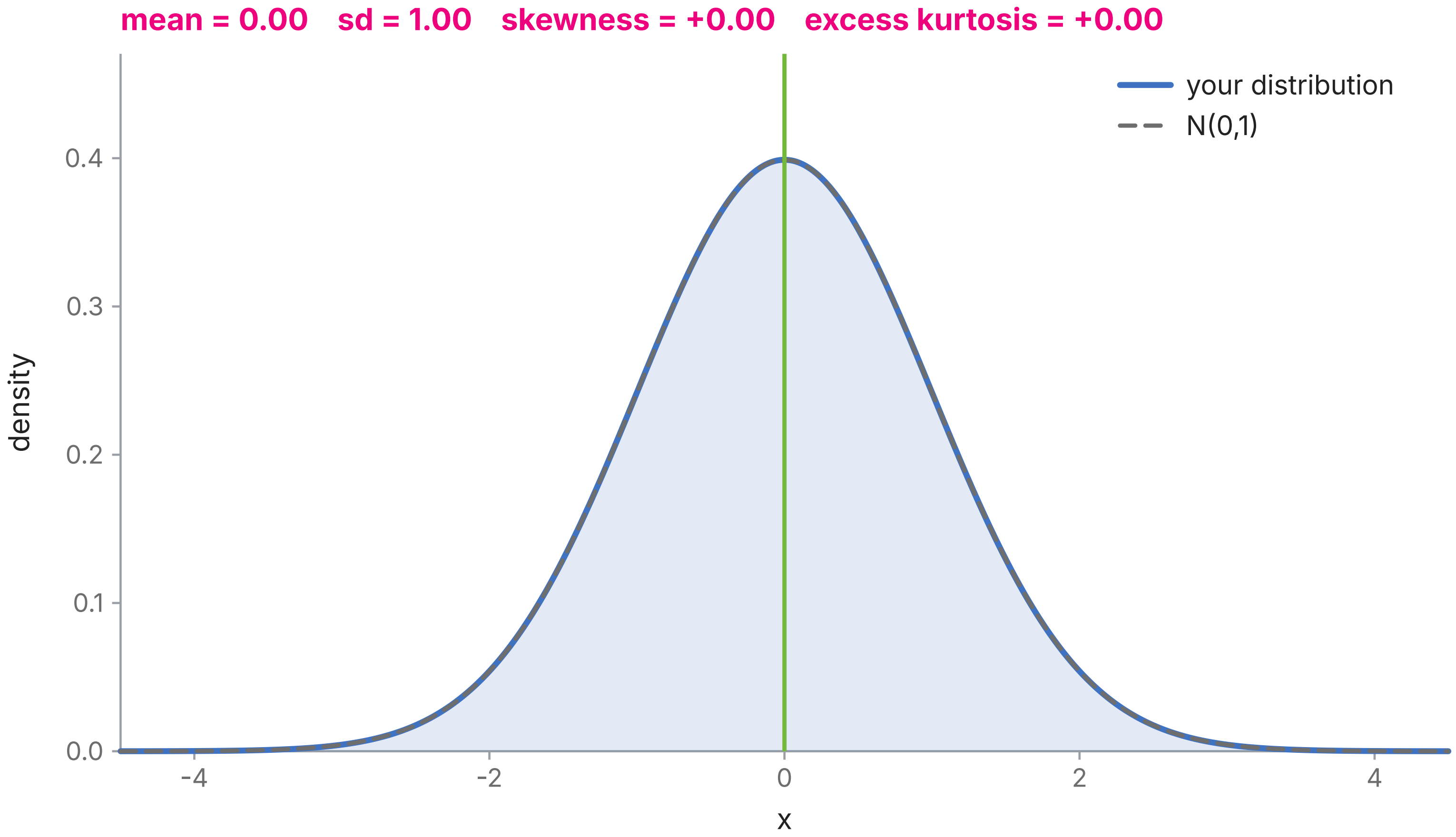
0

Tail knob δ

1

Show N(0,1) ☑

$X = \sinh\big((\sinh^{-1} Z + \varepsilon)/\delta\big)$ with $Z \sim N(0,1)$, standardised to mean 0, variance 1. At $\varepsilon = 0, \delta = 1$ it *is* the normal.



What the Knobs Do

- The **skew knob** moves mass to one side: γ_1 changes sign, while the mean and variance stay pinned at 0 and 1 by construction.
- The **tail knob** trades shoulder for tail: $\delta < 1$ gives heavy tails (large γ_2), $\delta > 1$ gives light ones.
- Mean and variance alone therefore say **nothing** about shape — two distributions can match on both and still look completely different.

Takeaway

Higher moments are how we describe the part of a distribution that μ and σ^2 cannot see.

Common Distributions

Bernoulli Distribution

Definition: models a single trial with two outcomes.

PMF:

$$p_X(x) = \begin{cases} p & \text{if } x = 1 \\ 1 - p & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

Parameters: $p \in [0, 1]$ (probability of success)

Moments: $E[X] = p$, $\mathrm{Var}(X) = p(1 - p)$

Notation: $X \sim \mathrm{Bernoulli}(p)$

Binomial Distribution

Definition: number of successes in n independent Bernoulli trials.

PMF:

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Parameters:

- $n \in \mathbb{N}$ (number of trials)
- $p \in [0, 1]$ (probability of success)

Moments:

- $E[X] = np$
- $\mathrm{Var}(X) = np(1-p)$

Notation: $X \sim \mathrm{Binomial}(n, p)$

Poisson Distribution

Definition: models rare events or counts in a fixed interval.

PMF:

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2, \ldots$$

Parameter: $\lambda > 0$ (rate)

Moments: $E[X] = \lambda$, $\mathrm{Var}(X) = \lambda$

Key Property

Poisson approximates the Binomial when n is large, p is small, and $np = \lambda$ is moderate.

Notation: $X \sim \mathrm{Poisson}(\lambda)$

Uniform Distribution

Definition: equal likelihood over an interval $[a, b]$.

PDF:

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \le x \le b \\ 0 & \text{otherwise} \end{cases}$$

CDF:

$$F_X(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \le x \le b \\ 1 & \text{if } x > b \end{cases}$$

Moments: $E[X] = \frac{a+b}{2}$, $\mathrm{Var}(X) = \frac{(b-a)^2}{12}$

Notation: $X \sim \mathrm{Uniform}(a, b)$

Exponential Distribution

Definition: models waiting times between events.

PDF:

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

CDF:

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \geq 0$$

Parameter: $\lambda > 0$ (rate)

Moments: $E[X] = \frac{1}{\lambda}$, $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

Memoryless Property

$$P(X > s + t \mid X > s) = P(X > t)$$

A used component is as good as new — the only continuous distribution with this property.

Notation: $X \sim \mathrm{Exp}(\lambda)$

Normal Distribution

Definition: the most important continuous distribution.

PDF:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Parameters:

- $\mu \in \mathbb{R}$ (mean)
- $\sigma^2 > 0$ (variance)

Notation: $X \sim N(\mu, \sigma^2)$

Moments: $E[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$

Standard Normal: $Z \sim N(0, 1)$, CDF written $\Phi(z)$

The Distribution Zoo, Live

Move the cut-off x_0 and watch the **shaded area** on top become the **height** of the CDF below.

Distribution

Binomial

n

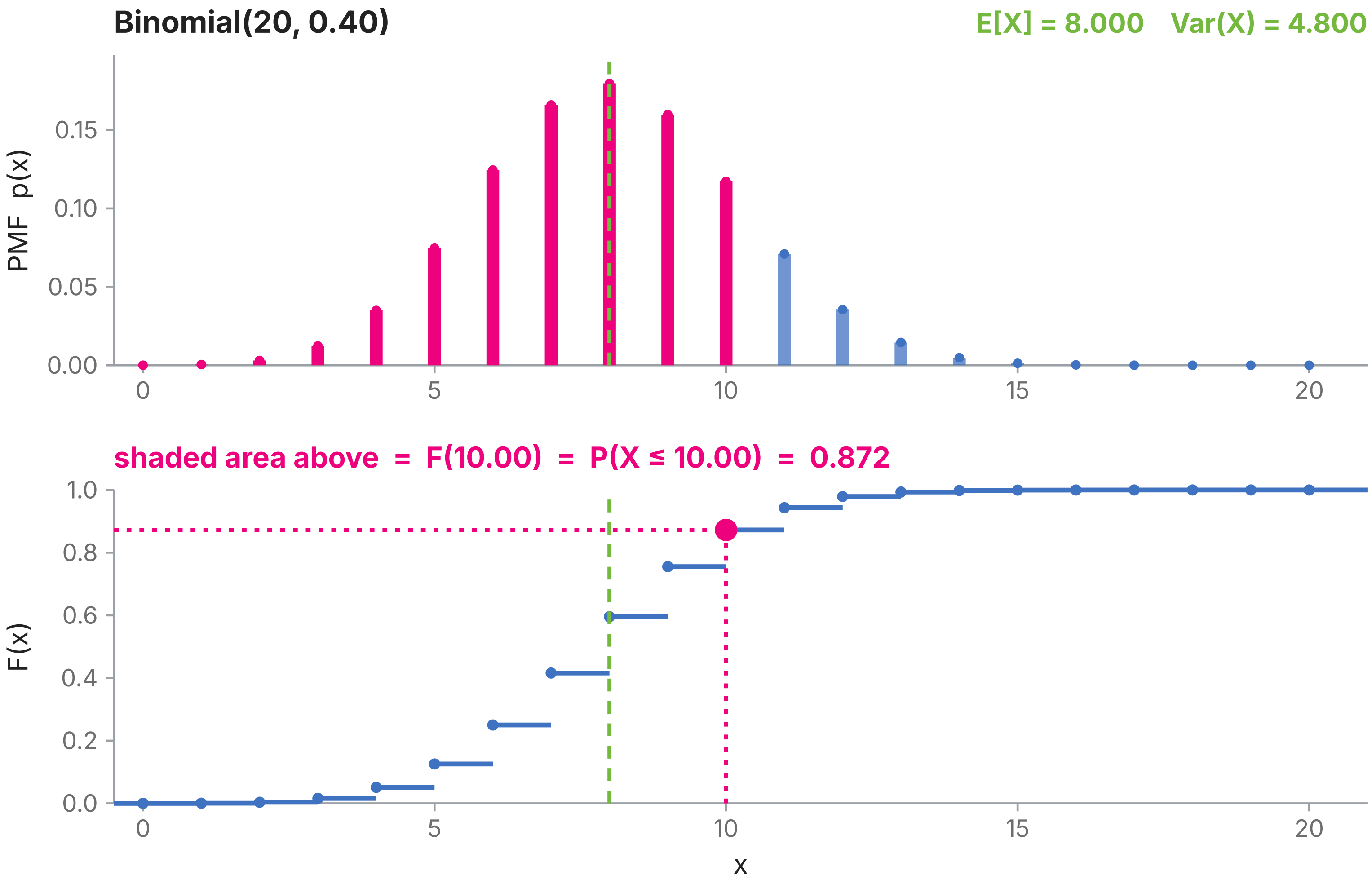
20

p

0.4

cut-off x_0

0.5



Binomial → Poisson, Live

Hold $\lambda = np$ fixed and let n grow: the binomial mass collapses onto the Poisson.

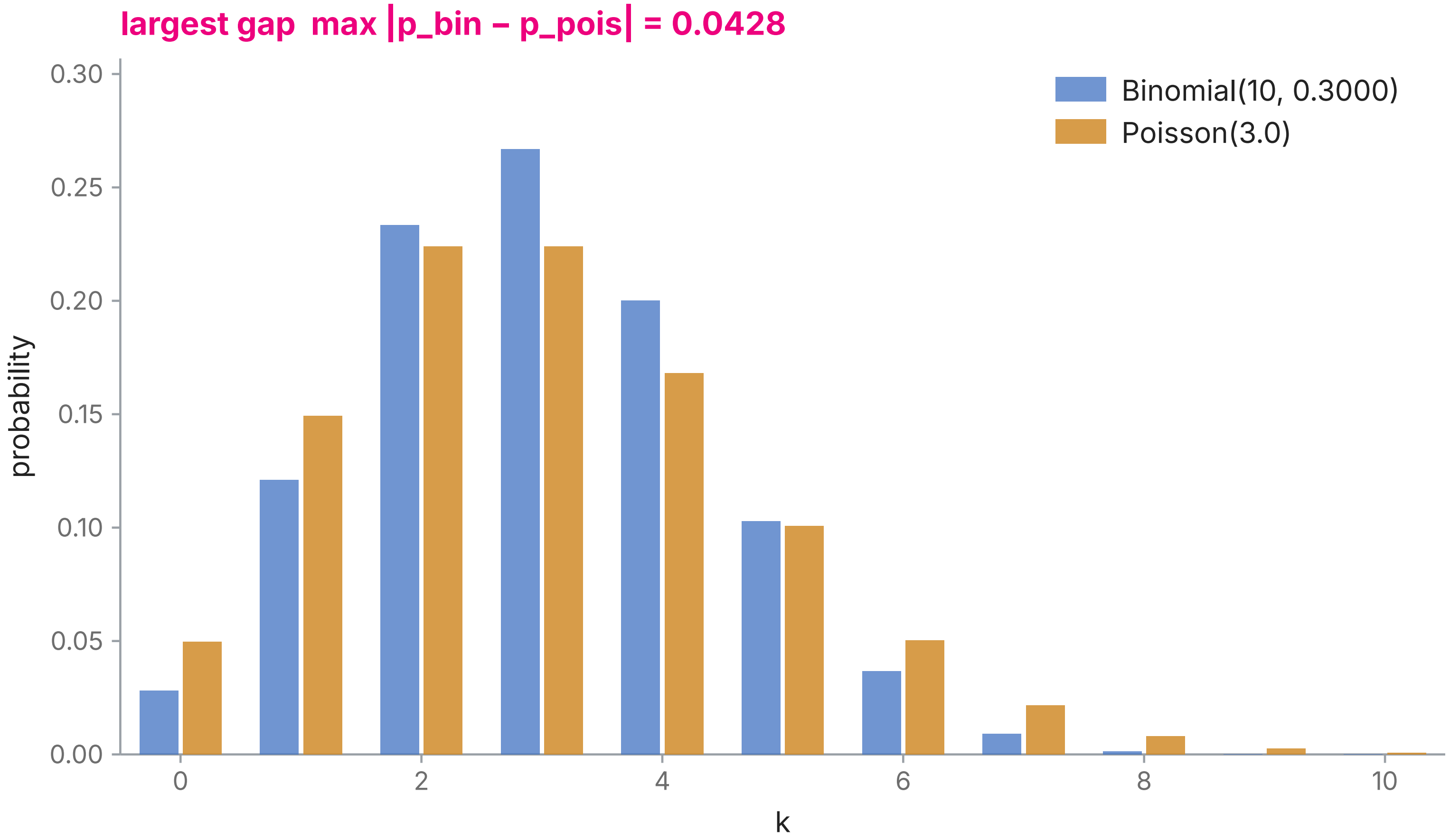
λ = np

3

n (p = λ/n)

10

Watch the largest gap in the readout — it shrinks roughly like λ^2/n.



Transformations of Random Variables

Functions of Random Variables

Question: if $Y = g(X)$, how do we find the distribution of Y?

Discrete case:

$$p_Y(y) = \sum_{x:\, g(x)=y} p_X(x)$$

Continuous case (monotonic g): if g is strictly monotonic with inverse g^{-1},

$$f_Y(y) = f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right|$$

Where the Jacobian comes from

Densities are probability *per unit length*. Stretching the axis by g dilutes the density; the derivative term is the bookkeeping that keeps the total area at 1.

Linear Transformations

If $Y = aX + b$:

Expected value:

$$E[Y] = aE[X] + b$$

Variance:

$$\mathrm{Var}(Y) = a^2\mathrm{Var}(X)$$

Special Case — Standardisation

$$Z = \frac{X - E[X]}{\sqrt{\mathrm{Var}(X)}}$$

Then $E[Z] = 0$ and $\mathrm{Var}(Z) = 1$.

Note what standardisation does **not** do: it moves and rescales, but it never changes the *shape*. A skewed X gives a skewed Z.

Example: Square of a Standard Normal

Let $X \sim N(0,1)$ and $Y = X^2$. Find the distribution of Y.

Solution. For $y > 0$:

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(X^2 \le y) \\ &= P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) \end{aligned}$$

Taking the derivative:

$$f_Y(y) = \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad y > 0$$

Recognise it

This is the **Chi-squared distribution with 1 degree of freedom** — and note g was *not* monotonic, which is why two branches had to be added up.

Transformations, Live

Left: the distribution of X. Right: what g does to it.

Distribution of X

N(0,1)

Transformation g

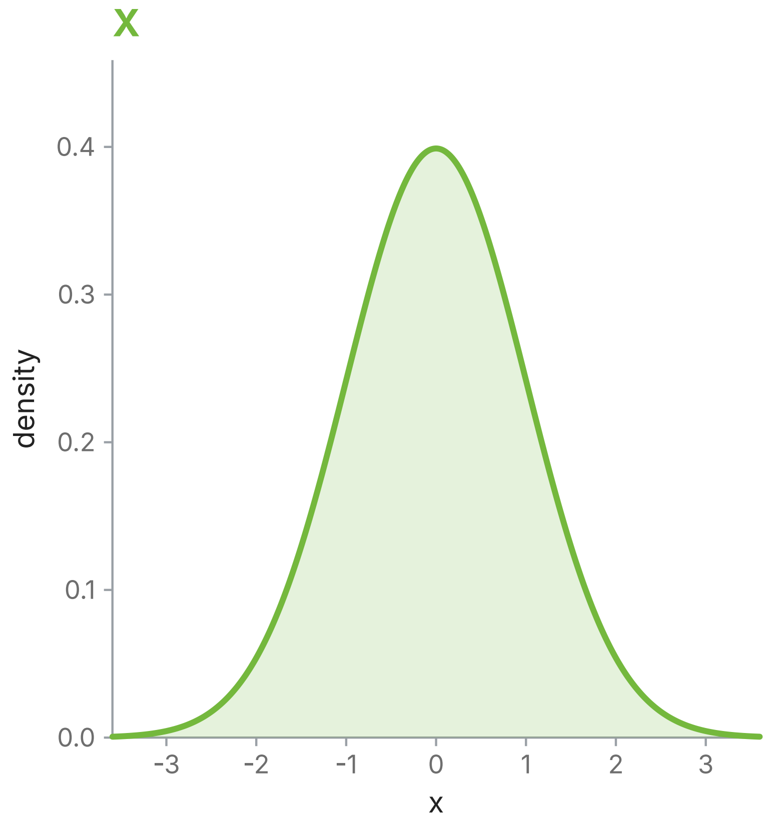
X²

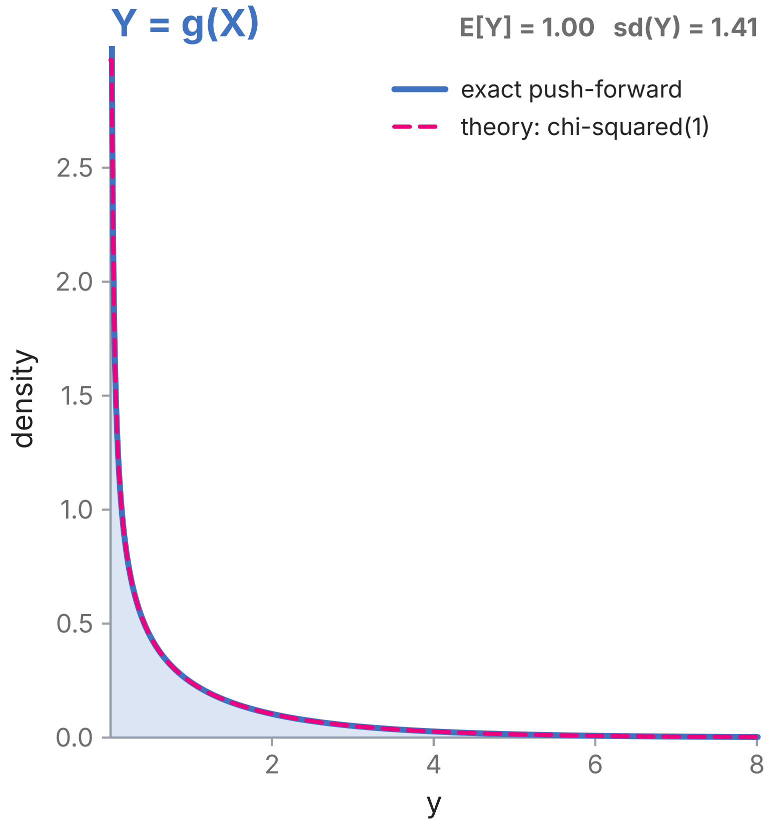
a (linear only)

2

b (linear only)

1





Limit Theorems

Laws of Large Numbers

Theorem — Law of Large Numbers

Let $X_1, X_2, \dots$ be i.i.d. with $E[|X_i|] < \infty$ and $E[X_i] = \mu$. Then

(1) for $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and any fixed $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - \mu\right| > \epsilon\right) = 0 \qquad \text{(WLLN)}$$

(2)

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1 \qquad \text{(SLLN)}$$

Key differences:

- **WLLN:** convergence in probability ($\bar{X}_n \xrightarrow{P} \mu$)
- **SLLN:** almost sure convergence ($\bar{X}_n \xrightarrow{a.s.} \mu$)

The Law of Large Numbers, Live

Each line is one sequence of draws, plotted as the running average $\bar{X}_n$.

Population

Exp(1)

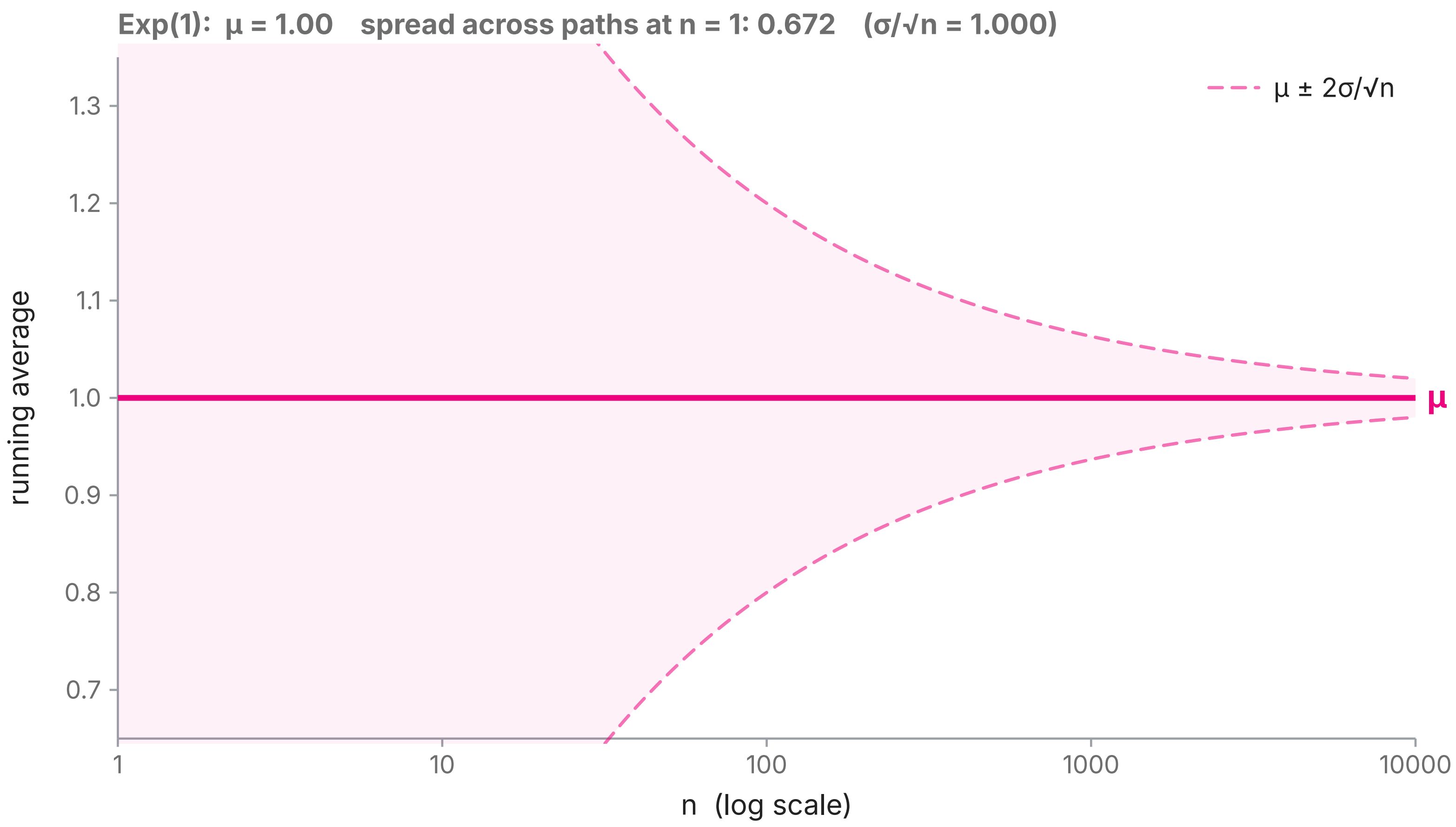
Paths

8

Draws per path

1

New simulation



What the Paths Show

- Every path wanders at first, then tightens around μ — the spread of $\bar{X}_n$ shrinks like $1/\sqrt{n}$.
- The **Cauchy** case is the useful counterexample: the sample mean jumps around forever, because $E|X| = \infty$ and the theorem's hypothesis fails.

Read the assumptions

"Averages converge" is not a law of nature. It is a theorem, and it has conditions.

Central Limit Theorem

Theorem — Central Limit Theorem

Let $X_1, X_2, \ldots$ be i.i.d. with $E[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2 < \infty$. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{d} N(0,1)$$

as $n \to \infty$.

Practical interpretation: for large n, $\bar{X}_n$ is approximately normal,

$$\bar{X}_n \approx N\left(\mu, \frac{\sigma^2}{n}\right)$$

Rule of thumb: $n \geq 30$ is often sufficient — but see the next two slides for how badly that can fail.

The Central Limit Theorem, Live

Add up n independent $\mathrm{Uniform}(0,1)$ draws. No simulation: each extra term is one more **convolution**, so this is the exact density.

Terms n

1

Centre: subtract a_n

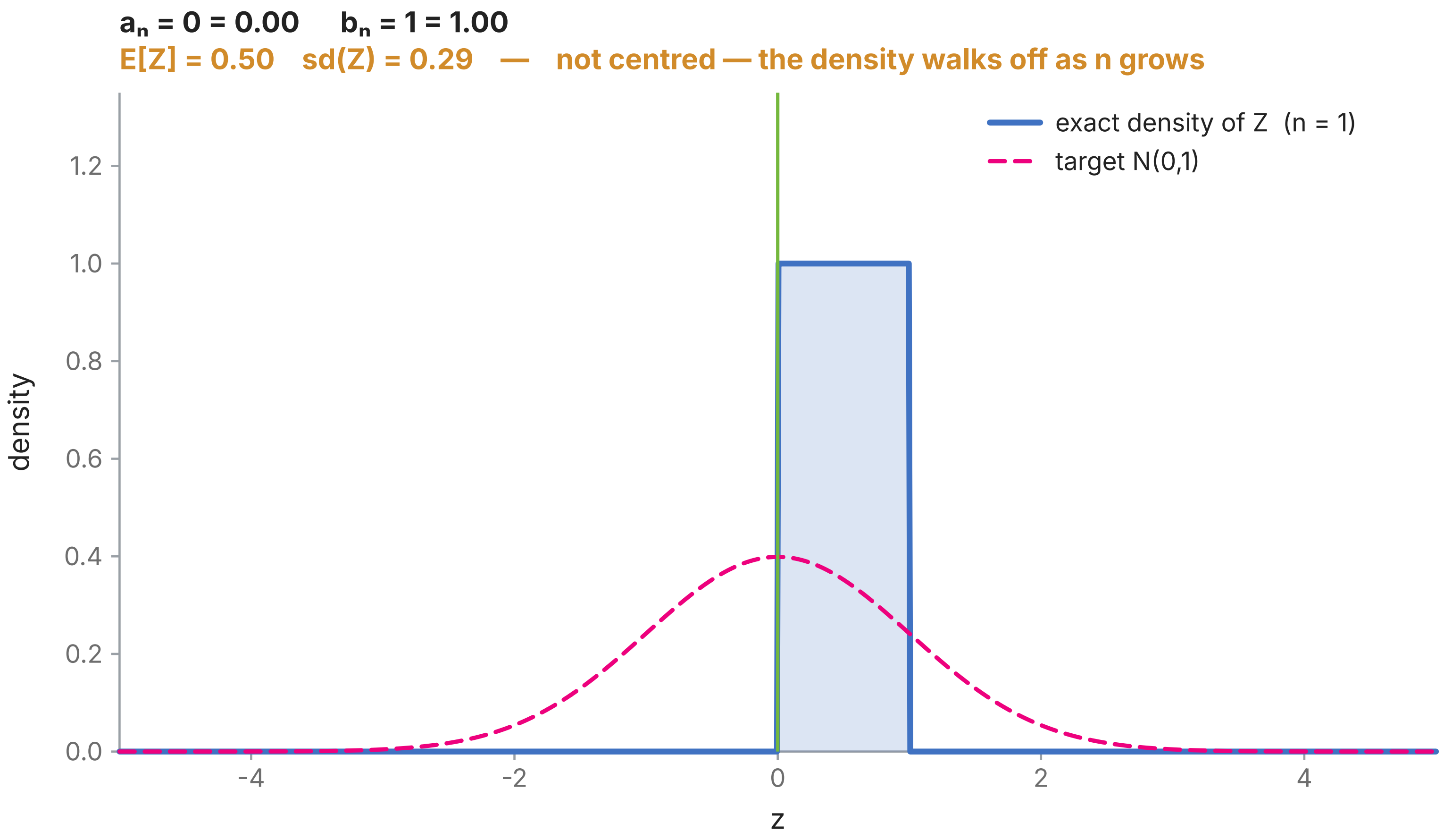
0 (none)

Scale: divide by b_n

1 (none)

$Z = \dfrac{S_n - a_n}{b_n}$, where $S_n = X_1 + \cdots + X_n$.

Only $a_n = n\mu$ together with $b_n = \sigma\sqrt{n}$ gives the standard normal. The frame never moves: a wrong a_n walks the density off the edge, too small a b_n flattens it, too large a b_n collapses it to a spike.



Any Population, Live

The population on the left can be as ugly as you like. Watch the right-hand panel anyway.

Population

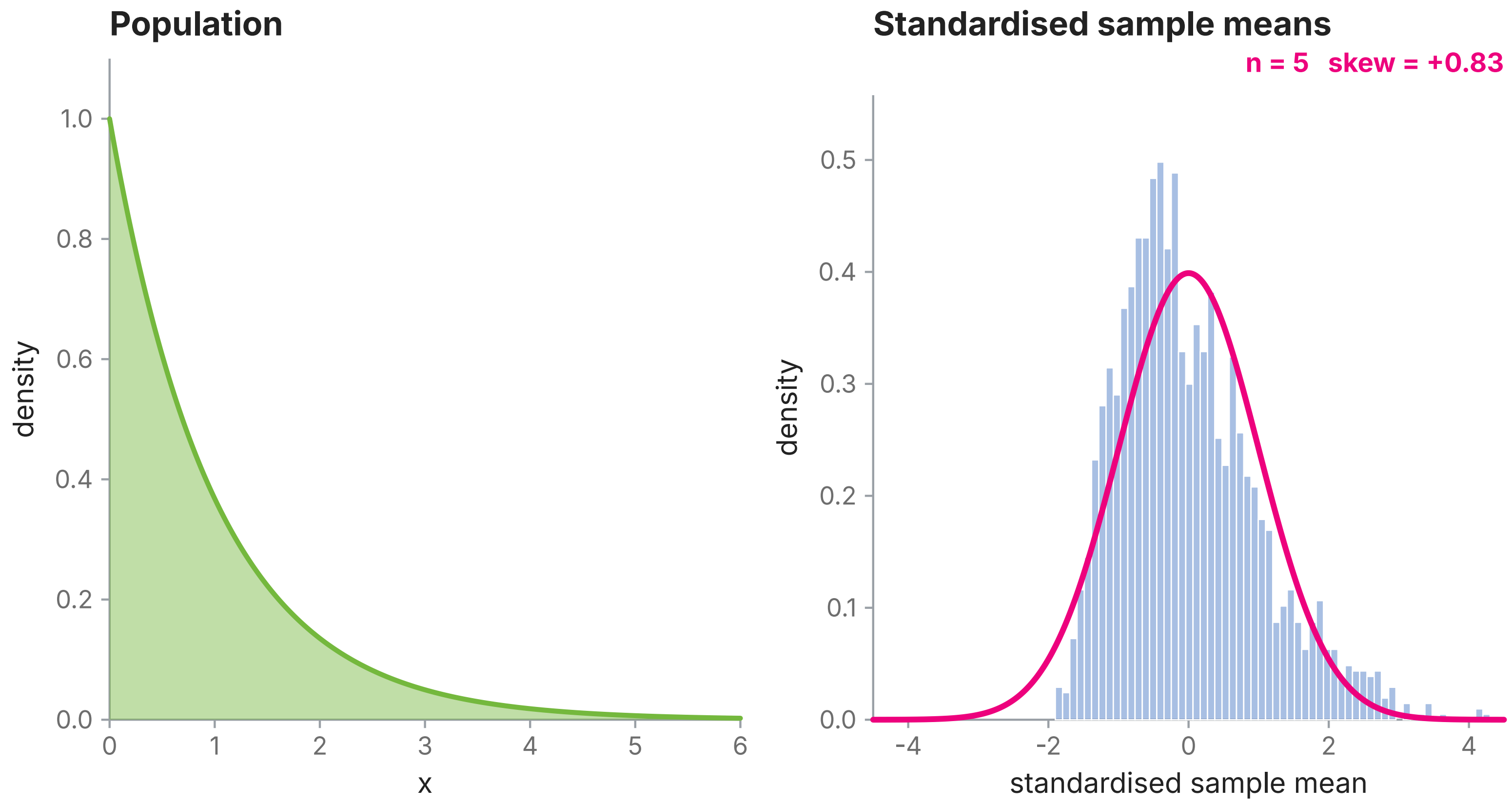
Exp(1) — skewed

Sample size n

5

Number of samples

2000



What the CLT Demo Teaches

- At $n = 1$ the right panel *is* the population — standardised, but the same shape.
- Skewness of the standardised mean falls like $\gamma_1/\sqrt{n}$: it fades, but slowly.
- **Bernoulli(0.08)** at $n = 30$ is still visibly lumpy and skewed — the “$n \geq 30$” rule of thumb quietly assumes the population is not too extreme.

Takeaway

The CLT is a statement about the **limit**. How large n must be before the approximation is usable depends on the population’s shape, not on a magic number.

Summary

Key Takeaways

(1) **Random variables** map outcomes to numbers, enabling mathematical analysis of uncertainty

(2) **Discrete vs continuous:** different tools (PMF vs PDF) but unified by the CDF

(3) **Expected value and variance** characterise the centre and spread — and higher moments the shape

(4) **Common distributions** model different real-world phenomena, and are related to one another (Binomial → Poisson, $N(0,1)^2 \to \chi^2_1$)

(5) **Transformations** let us derive new distributions from known ones

(6) **Limit theorems** describe the behaviour of averages: the LLN says *where* $\bar{X}_n$ goes, the CLT says *how it is spread out* on the way there

Next: Multivariate Random Variables

Two or more at a time

- Joint, marginal and conditional distributions
- Independence
- Covariance and correlation

And then

- The multivariate normal
- Conditional expectation
- Sums and sampling distributions