Continuous Distributions

Uniform, Normal, Exponential, and More

Scott Cunningham

Gov 2001 · Harvard University

Spring 2026

Where We Are

From discrete counts to continuous measurements

Monday: Discrete distributions

- Bernoulli, Binomial, Poisson
- Counting successes and rare events
- Sum properties: Binomials add (same p), Poissons add

Today: Continuous distributions

- **Uniform** — equal probability, and why it's fundamental
- **Normal** — the star of the show (CLT, regression)
- **Exponential** — waiting times and memorylessness
- **Connections**: Poisson $\leftrightarrow$ Exponential, Chi-square, t

Reading: Aronow & Miller §1.4–1.5, Blackwell 2.4–2.5

Reminder: Why We Care About Distributions

The link to inference

The roadmap:

1. **Population**: Described by a distribution with unknown parameters
2. **Sample**: Data we observe (drawn from the population)
3. **Estimation**: Use data to learn about parameters (μ, σ^2, λ)
4. **Uncertainty**: Quantified via sampling distributions (which we derive from population distributions)

Today's distributions matter because:

- The **Normal** is the sampling distribution of the mean (CLT)
- The **Chi-square** appears when estimating variance
- The ***t*-distribution** is what we use for hypothesis tests with estimated variance

Everything connects. Today we're building the vocabulary you'll use for inference.

Part I

The Uniform Distribution

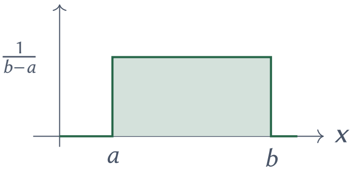
Simple but Fundamental

The Uniform Distribution: Quick Review

You've seen this — let's go deeper

Definition: $X \sim \text{Uniform}(a, b)$ has PDF:

$$f(x) = \frac{1}{b-a} \quad \text{for } x \in [a, b]$$



Key formulas: $\mathbb{E}[X] = \frac{a+b}{2}$ (midpoint), $\quad \text{Var}[X] = \frac{(b-a)^2}{12}$

The Standard Uniform: $U \sim$ Uniform$(0, 1)$

The building block for everything

The **standard uniform** $U \sim \text{Uniform}(0, 1)$ is special:

- PDF: $f(u) = 1$ for $u \in [0, 1]$
- CDF: $F(u) = u$ for $u \in [0, 1]$
- $\mathbb{E}[U] = 0.5, \quad \text{Var}[U] = 1/12$

Why is it fundamental?

- Every random number generator starts with Uniform(0,1)
- Randomization in experiments: "treat if $U < 0.5$"
- And something deeper: the **universality of uniform**

Universality of the Uniform

How to generate *any* distribution from Uniform(0,1)

Theorem (Probability Integral Transform):
Let X be a continuous random variable with CDF F. Then:

$$F(X) \sim \text{Uniform}(0, 1)$$

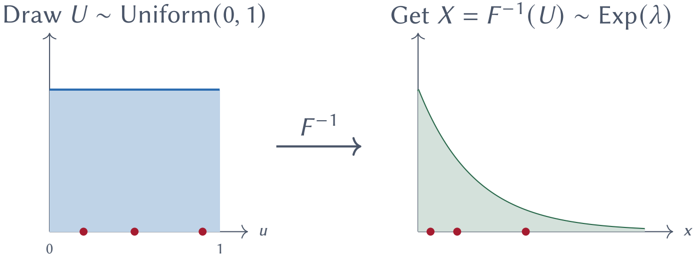
The converse (this is the useful part):
If $U \sim \text{Uniform}(0, 1)$ and F is any CDF, then:

$$X = F^{-1}(U) \text{ has CDF } F$$

Translation: To simulate from any distribution, just apply its inverse CDF to uniform draws.

Universality: Visualized

Generating Exponential from Uniform



For Exponential: $F(x) = 1 - e^{-\lambda x}$, so $F^{-1}(u) = -\frac{1}{\lambda}\ln(1-u)$

This is how statistical software simulates from *any* distribution.

The Normal Distribution

The Star of the Show

The Normal Distribution

Why it dominates statistics

Definition: $X \sim \text{Normal}(\mu, \sigma^2)$ has PDF:

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

for $x \in \mathbb{R}$, where μ is the mean and σ^2 is the variance.

Why is it everywhere?

1. **Central Limit Theorem**: Sample means are approximately normal
2. **Closure**: Sums of normals are normal
3. **Tractability**: Easy to compute probabilities, confidence intervals

The Standard Normal: $Z \sim N(0, 1)$

Definition: $Z \sim N(0, 1)$ is the **standard normal**.

Standardization: If $X \sim N(\mu, \sigma^2)$, then:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

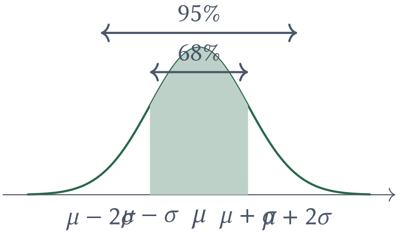
Key notation:

- $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ — the standard normal PDF
- $\Phi(z) = \mathbb{P}(Z \leq z)$ — the standard normal CDF

Tables, software, and formulas are all in terms of Φ. Standardize first.

The 68–95–99.7 Rule

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



- **68%** of values within 1 SD of mean
- **95%** within 2 SDs (more precisely: 1.96)
- **99.7%** within 3 SDs

Normal Closure Properties

Sums and linear combinations stay normal

Property 1 (Scaling and shifting):
If $X \sim N(\mu, \sigma^2)$, then for constants a, b:

$$aX + b \sim N(a\mu + b,\ a^2\sigma^2)$$

Property 2 (Sum of independent normals):
If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$ are **independent**, then:

$$X + Y \sim N(\mu_X + \mu_Y,\ \sigma_X^2 + \sigma_Y^2)$$

These properties make the normal uniquely tractable. No other distribution has both.

Part III

The Exponential Distribution

Waiting for an Event

The Exponential Distribution

Time until the next event

Setup: How long until the next event, if events occur at constant rate λ?

Definition: $T \sim \text{Exponential}(\lambda)$ has PDF:

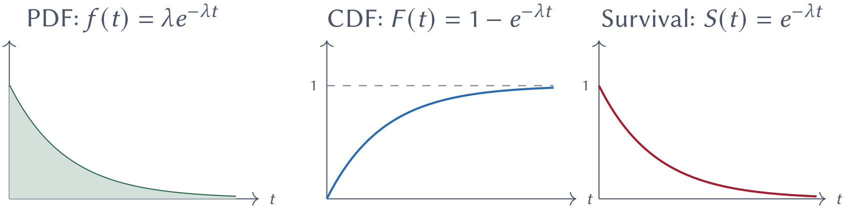
$$f(t) = \lambda e^{-\lambda t} \quad \text{for } t \geq 0$$

where $\lambda > 0$ is the **rate parameter**.

Political science examples:

- Time until a cabinet collapse
- Duration of ceasefires
- Time between Supreme Court vacancies

Exponential: PDF, CDF, and Survival Function



Survival function: $S(t) = \mathbb{P}(T > t) = 1 - F(t) = e^{-\lambda t}$

The probability of “surviving” past time t decays exponentially.

Exponential: Key Properties

For $T \sim \text{Exponential}(\lambda)$:

Expected value: $\mathbb{E}[T] = \frac{1}{\lambda}$
Variance: $\text{Var}[T] = \frac{1}{\lambda^2}$

Interpretation: If events occur at rate λ per unit time, the average wait is $1/\lambda$.

Example: Supreme Court vacancies at rate $\lambda = 0.5$ per year $\rightarrow$ average wait = 2 years.

The Memoryless Property

The exponential "forgets" how long it's waited

Property: For $T \sim \text{Exponential}(\lambda)$:

$$\mathbb{P}(T > s + t \mid T > s) = \mathbb{P}(T > t)$$

In words: Given that you've already waited s units, the probability of waiting *another* t units is the same as if you'd just started waiting.

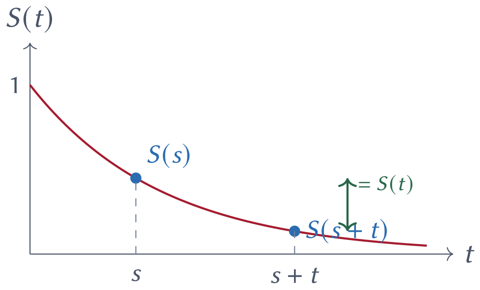
Proof:

$$\mathbb{P}(T > s + t \mid T > s) = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(T > t)$$

Only the exponential (continuous) and geometric (discrete) have this property.

Memorylessness Visualized

The survival curve “restarts” at any point



The *ratio* of survival probabilities depends only on the additional wait t, not on s.

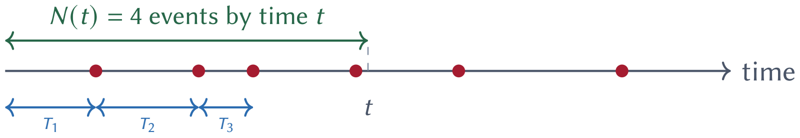
The Poisson–Exponential Connection

Two Sides of One Process

Poisson and Exponential: The Dual View

Counts vs. waiting times

If events occur at rate λ per unit time:



- **Poisson question**: How many events in time t? $N(t) \sim \text{Poisson}(\lambda t)$
- **Exponential question**: How long until next event? $T_i \sim \text{Exp}(\lambda)$

Same process, different questions.

The Key Identity

Connecting Poisson and Exponential

Let T_1 be the time until the first event. Then:

$$\mathbb{P}(T_1 > t) = \mathbb{P}(\text{no events by time } t) = \mathbb{P}(N(t) = 0)$$

Using Poisson:

$$\mathbb{P}(N(t) = 0) = \frac{(\lambda t)^0 e^{-\lambda t}}{0!} = e^{-\lambda t}$$

This is exactly the survival function of Exponential(λ)!

The Poisson count and exponential waiting time are two views of the same process.

Political Science Example

Supreme Court vacancies

Suppose vacancies occur at rate $\lambda = 0.5$ per year.

Poisson question: What's $\mathbb{P}$(at least 2 vacancies in a 4-year term)?

- $N(4) \sim \text{Poisson}(0.5 \times 4) = \text{Poisson}(2)$
- $\mathbb{P}(N \geq 2) = 1 - \mathbb{P}(N = 0) - \mathbb{P}(N = 1) = 1 - e^{-2} - 2e^{-2} \approx 0.59$

Exponential question: What's the average wait for the next vacancy?

- $T \sim \text{Exp}(0.5)$
- $\mathbb{E}[T] = 1/0.5 = 2$ years

Same λ, different questions, complementary answers.

Your Turn: Continuous Practice

Work through these with a partner

1. **Normal**: Adult heights are $N(170, 100)$ cm (mean 170, variance 100).
   - What's the standard deviation?
   - What range contains about 95% of heights?
2. **Exponential**: Congressional hearings occur at rate $\lambda = 3$ per month.
   - What's the expected wait for the next hearing?
   - What's $\mathbb{P}(\text{wait} > 1 \text{ month})$?

Answers: (1) SD = 10 cm; 150–190 cm. (2) $\mathbb{E}[T] = 1/3$ month; $\mathbb{P}(T > 1) = e^{-3} \approx 0.05$.

Chi-Square and *t* Distributions

Preview: You'll Need These for Regression

Why Are We Showing You These?

Planting seeds for inference

You might wonder: why Chi-square and t distributions now?

The short answer: In a few weeks, when you run a regression and want to know if a coefficient is "statistically significant," you'll need these distributions.

- **Chi-square**: Appears when we estimate variance from data
- ***t*-distribution**: What we use for hypothesis tests when variance is estimated (which is always in practice)

We're not going deep today—just building familiarity. When you see these again in the regression unit, you'll have context.

The Chi-Square Distribution

Sum of squared standard normals

Definition: If $Z_1, \ldots, Z_k \overset{\text{iid}}{\sim} N(0, 1)$, then:

$$X = Z_1^2 + Z_2^2 + \cdots + Z_k^2 \sim \chi_k^2$$

where k is the **degrees of freedom**.

Key facts:

- $\mathbb{E}[X] = k$
- $\text{Var}[X] = 2k$
- Support: $[0, \infty)$ — always non-negative (it's a sum of squares)

You'll see this when we estimate variance, test hypotheses about multiple coefficients, and compute R^2.

The t Distribution

Normal divided by Chi-square

Definition: If $Z \sim N(0, 1)$ and $V \sim \chi^2_k$ are independent, then:

$$T = \frac{Z}{\sqrt{V/k}} \sim t_k$$

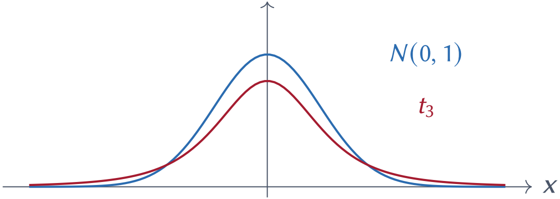
where k is the degrees of freedom.

Key facts:

- Symmetric around 0, like the normal
- **Heavier tails** than normal — more probability in extremes
- As $k \to \infty$, $t_k \to N(0, 1)$

The t distribution is why we use "t-tests" and "t-statistics" in regression.

Normal vs. t: Heavier Tails



The t distribution has more probability in the tails.

With small samples, extreme values are more likely — the t accounts for this.

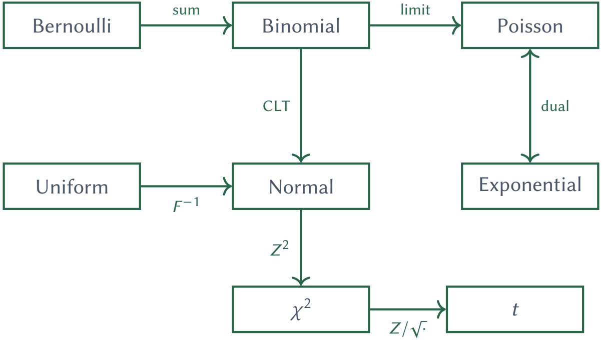
Summary: Continuous Distributions

Distribution	$\mathbb{E}[X]$	$\text{Var}[X]$	**Use case**
Uniform(a, b)	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	Equal probability, simulation
Normal(μ, σ^2)	μ	σ^2	CLT, regression errors
Exponential(λ)	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	Waiting times, memoryless
χ^2_k	k	$2k$	Variance estimation, tests
t_k	0	$\frac{k}{k-2}$	Small-sample inference

Key connections:

- Uniform(0,1) $\longrightarrow$ any distribution via inverse CDF
- Poisson $\leftrightarrow$ Exponential: counts vs. waiting times
- Normal $\longrightarrow$ Chi-square (sum of squares) $\longrightarrow$ t (ratio)

The Big Picture: How Distributions Connect



Understanding these connections helps you see why certain distributions appear in certain contexts.

Looking Ahead

Next week: Joint distributions and the CEF

- Joint, marginal, and conditional distributions
- Covariance and correlation
- The Conditional Expectation Function (CEF)

Reading:

- Aronow & Miller, §1.3 and §2.2
- Blackwell, Chapter 2.4–2.5

Problem Set 2: Due February 21