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

Today: Continuous distributions — the **A-list** stars of statistics

- **Uniform** — equal probability, foundation for simulation
- **Normal** — the star of the show (CLT, regression, everything)
- **Exponential** — waiting times, survival analysis
- **Chi-square**, t — the B-list supporting cast for inference

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

The Pattern Continues: Parameters Define Everything

Same roadmap as discrete, now with PDFs instead of PMFs

Monday's pattern for discrete distributions:

$$\text{Parameters} \longrightarrow \text{PMF} \longrightarrow \mathbb{E}[X], \text{Var}[X]$$

Today's pattern for continuous distributions:

$$\text{Parameters} \longrightarrow \text{PDF} \longrightarrow \text{CDF} \longrightarrow \mathbb{E}[X], \text{Var}[X]$$

For each distribution today, we'll identify:

1. **Parameters**: What do you need to specify? (μ, σ^2, λ, a, b)
2. **PDF**: The density function $f(x)$
3. **CDF**: The cumulative distribution $F(x) = \mathbb{P}(X \leq x)$
4. **Moments**: Expected value and variance

Once you know the parameters, everything else follows.

Most of Your Statistical Life Will Be Normal

The A-list and B-list of continuous distributions

A-list actors — you'll model data with these:

- **Normal** — regression errors, polling, heights, test scores
- **Exponential** — waiting times, survival, duration models
- **Uniform** — simulation, randomization, probability foundations

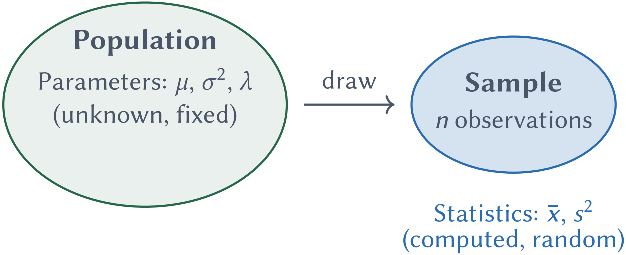
B-list actors — supporting roles for inference:

- **Chi-square** — variance estimation, goodness-of-fit
- ***t*-distribution** — hypothesis testing with estimated variance

The B-list are *derived* from Normal. You don't model data with them—you use them for inference.

Statistics Is Learning About Populations from Samples

The central distinction



The game: Use sample statistics to *estimate* population parameters.

Today's distributions describe *populations*. Estimation theory tells us how sample statistics behave.

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

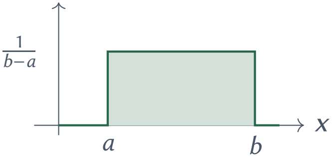
The Simplest Distribution: Equal Probability Everywhere

Political example: When does the voter arrive?

Example: A voter arrives at a polling station sometime between 8am and 8pm. If arrivals are "uniformly distributed," any moment is equally likely.

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$(0, 1)$

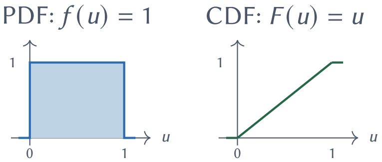
The building block for everything

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**



Support Tells You Where Outcomes Are Possible

You worked with this on Problem Set 2

Definition: The **support** of a random variable is where its PDF is positive:

$$\text{Supp}[X] = \{x : f(x) > 0\}$$

Three types of support we'll see today:

Distribution	Support	Type
Uniform(a, b)	$[a, b]$	Bounded (finite interval)
Normal(μ, σ^2)	$(-\infty, +\infty)$	Unbounded (whole line)
Exponential(λ)	$[0, +\infty)$	Half-line (non-negative)

PS2 Q2 asked you to find support. This concept matters for specifying models correctly.

Universality of the Uniform

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

The idea: Plug any continuous X into its own CDF, and you always get Uniform(0,1).

Why? The CDF maps outcomes to probabilities. Since probabilities live in $[0, 1]$ and the CDF "spreads" outcomes evenly across this interval, the result is uniform.

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

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

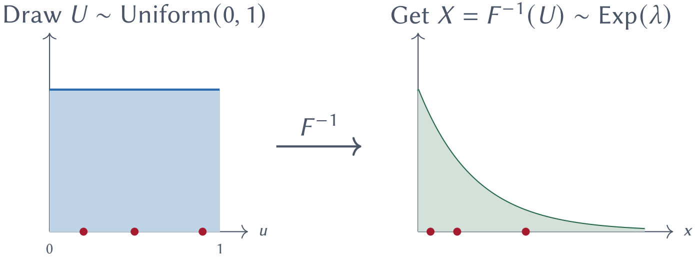
The converse (this is the useful part):
If $U \sim \text{Uniform}(0, 1)$ and F is any CDF, then:

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

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

De Moivre Discovered It; Gauss Got the Credit

Stigler's Law of Eponymy

Stigler's Law: "No scientific discovery is named after its original discoverer."

The Normal distribution is called "Gaussian"—but Gauss didn't discover it.

- **Abraham de Moivre (1733)**: French Huguenot exile in London, surviving by tutoring aristocrats in gambling mathematics. First derived the normal curve in *The Doctrine of Chances*.
- **Pierre-Simon Laplace (1774–1812)**: Developed the theory systematically. Proved early versions of the Central Limit Theorem.
- **Carl Friedrich Gauss (1809)**: Applied it to astronomical errors. Got the credit. But Gauss himself called it the "Laplacian curve."

De Moivre died impoverished. Gauss is called the "Prince of Mathematicians." Life isn't fair.

The Normal Distribution: The Star of the Show

Application first: Where do you see it?

Examples:

- Heights of adults, test scores, measurement errors
- **Polling errors** — why we talk about "margin of error"
- **Regression residuals** — the foundation of inference

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

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

Why everywhere? Three reasons:

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

Standardization Converts Any Normal to Z

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

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

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

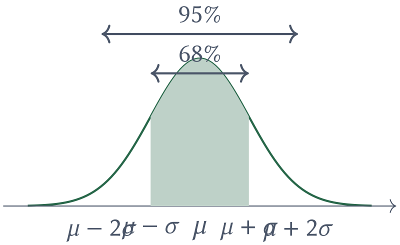
Key notation:

- $\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$ — 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.

Most Probability Concentrates Near the Mean

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 Has Unbounded Support — But Tails Vanish Fast

Theoretically infinite, practically finite

Support: $\text{Supp}[X] = (-\infty, +\infty)$ — any value is *theoretically* possible.

But probabilities decay exponentially in the tails:

- Outside 3 SDs: only 0.3% of probability
- Outside 4 SDs: only 0.006% of probability
- Outside 5 SDs: essentially zero (1 in 3.5 million)

Practical implication: For heights (mean 170cm, SD 10cm):

- Normal says negative heights are "possible" — but $P(X < 0) \approx 0$
- The model is an approximation; we accept tiny errors in exchange for tractability

Contrast with Exponential: support $[0, \infty)$ *enforces* non-negativity.

A Warning: The Normal Can Be Misused

"The Bell Curve" controversy

1994: Herrnstein & Murray publish *The Bell Curve*, claiming IQ differences between racial groups are genetic and immutable.

The statistical sin: They treated the Normal distribution as *destiny* rather than *description*.

James Heckman's critique (Nobel laureate, 1995):

- IQ is not fixed — it responds to environment and intervention
- The authors confused *description* with *explanation*
- Selection bias: who takes the tests, when, under what conditions?

Lesson: The Normal describes many phenomena. It doesn't explain them. Distributions are tools, not theories of causation.

Statistics without causal reasoning is dangerous.

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.

Galton and "Regression to the Mean"

Why tall parents have shorter children (on average)

Francis Galton (1886): Studied heights of fathers and sons.

Finding: Sons of very tall fathers were tall — but not *as* tall as their fathers. Sons of very short fathers were short — but not *as* short.

Galton called this "regression toward mediocrity":

- Extreme observations tend to be followed by less extreme ones
- This is a *statistical phenomenon*, not a biological force
- It happens whenever two variables are imperfectly correlated

Why "regression"? This is literally where the term comes from. Galton was "regressing" son's height on father's height.

The Normal distribution quantifies this: extreme Z-scores are rare by definition.

Part III

The Exponential Distribution

Waiting for an Event

How Long Until the Next Supreme Court Vacancy?

Application first: Waiting times in politics

Political science questions that involve waiting:

- How long until the next Supreme Court vacancy?
- How long will this ceasefire last?
- How long until a cabinet collapse?
- Time between terrorist attacks in a region?

Historical data: Supreme Court vacancies occur at rate $\lambda \approx 0.5$ per year.

$\Rightarrow$ Average wait: about 2 years between vacancies.

The **Exponential distribution** models these waiting times.

The Exponential Distribution

The math behind waiting times

Definition: $T \sim$ Exponential(λ) has PDF:

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

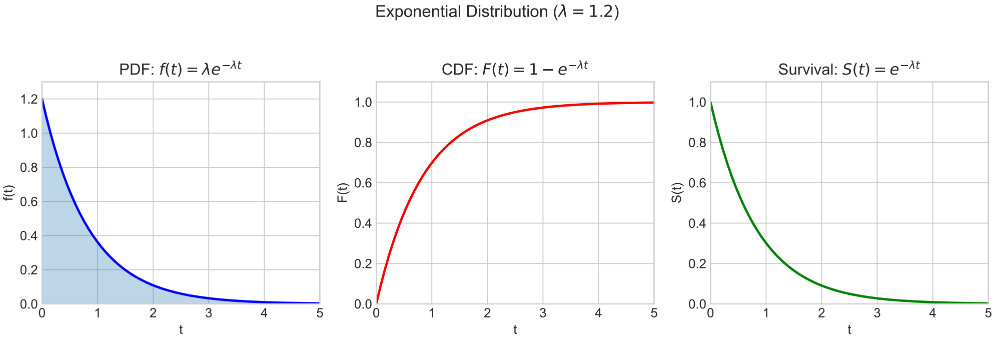
where $\lambda > 0$ is the **rate parameter**.

Support: $[0, +\infty)$ — waiting times are always non-negative.

PS2 Connection: Problem 7(c) asked you to find c for $f(y) = ce^{-2y}$. That's an Exponential($\lambda = 2$)! The answer was $c = 2$.

The Survival Function Decays Exponentially

PDF, CDF, and Survival — three views of the same distribution



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

The probability of "surviving" (not yet experiencing the event) past time t decays exponentially.

Average Wait Is the Inverse of the Rate

Key properties of the Exponential

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 Exponential Distribution Has No Memory

How long you've waited doesn't affect how much longer you'll wait

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.

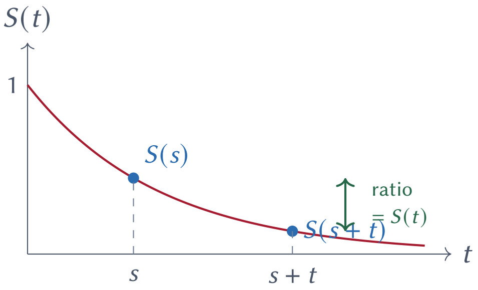
Proof:

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

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.

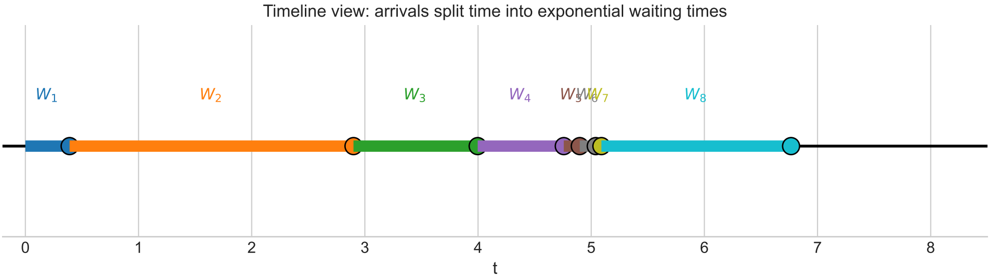
Part IV

The Poisson–Exponential Connection

Two Sides of One Process

Poisson Counts Events; Exponential Measures Waiting Times

Same process, different questions



- **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 rate λ. 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; $\mathbb{P}(T > 1) = e^{-3} \approx 0.05$.

Part V

Chi-Square and *t* Distributions

The B-List: Supporting Actors for Inference

Chi-Square and t Are Inference Tools, Not Data Models

You don't model data with these—you use them for hypothesis testing

A-list vs B-list:

- **A-list** (Normal, Exponential, Uniform): You model *data* with these
- **B-list** (Chi-square, t): You use these for *inference about parameters*

Why do they exist?

- **Chi-square**: When you estimate variance from data, your estimate follows a χ^2
- **t-distribution**: When you test hypotheses using an estimated (not known) variance

The punchline: In a few weeks, when you run a regression and ask "is this coefficient statistically significant?"—the t-distribution will give you the answer.

We're planting seeds. You'll see these again in the regression unit.

Chi-Square Is a Sum of Squared Normals

Derived from Normal—support is $[0, \infty)$

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 Is Normal with Heavier Tails

What happens when you don't know the true variance

The problem: In real life, you don't know σ. You estimate it from data.
The consequence: Your estimate $\hat{\sigma}$ is uncertain. This makes extreme values more likely than the Normal predicts.

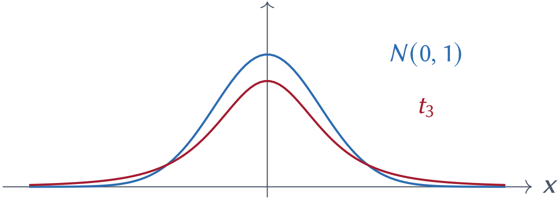
The solution: Use the t-distribution, which has heavier tails to account for this.

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

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

Key insight: Small k = more uncertainty about σ = heavier tails.
As $k \to \infty$, the t becomes Normal (you've estimated σ precisely).
This is why we use "t-tests" — they account for estimating variance from data.

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 A-List and B-List: A Summary

Which distributions model data? Which are for inference?

A-list actors — you model *data* with these:

- **Uniform**: Simulation, randomization, probability foundations
- **Normal**: CLT, regression errors, test scores, polling
- **Exponential**: Waiting times, survival analysis, duration

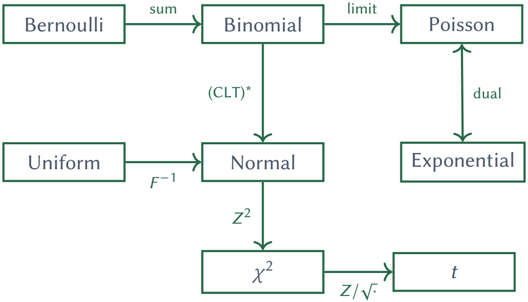
B-list actors — you use these for *inference*:

- **Chi-square**: Variance estimation, goodness-of-fit tests
- t: Hypothesis testing with estimated variance

The relationship: Normal $\xrightarrow{Z^2}$ Chi-square $\xrightarrow{Z/\sqrt{\cdot}}$ t

Most of your statistical life will be Normal. But when you estimate variance from data, the B-list appears.

How Distributions Connect: The Big Picture



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

*CLT = Central Limit Theorem (Week 5). Sample means of *any* distribution approach Normal.

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 3: Coming soon

PS2 due tonight (Feb 10). Hope you enjoyed working with PDFs, support, and MGFs!