Continuous Probability Distributions

Advanced Module A4

Digital Finance

Outline

Learning Objectives

By the end of this module, you will be able to:

1. Distinguish between PDF and CDF for continuous random variables
2. Apply the normal distribution, z-scores, and empirical rule
3. Use exponential, gamma, and Weibull distributions for waiting times
4. Recognize when to use t, chi-squared, and F distributions in inference
5. Identify heavy-tailed distributions (Pareto, Cauchy) and their implications
6. Interpret Q-Q plots to assess distributional fit

This module covers 14 continuous distributions with finance applications.

What is a Continuous Random Variable?

Definition: A random variable X is **continuous** if it can take any value in an interval.
Examples:

- Stock prices, returns, portfolio values
- Time until default, waiting time between trades
- Temperature, height, weight

Key insight: For continuous variables, $P(X = x) = 0$ for any specific value x.

- Why? Infinitely many possible values means each has zero probability
- Instead, we ask: $P(a < X < b)$ — probability of an interval

---

We measure probability as area under a curve, not height at a point.

Probability Density Function (PDF)

The PDF $f(x)$ describes the "density" of probability at each point.
Key properties:

- $f(x) \geq 0$ for all x (non-negative)
- $\int_{-\infty}^{\infty} f(x)dx = 1$ (total area = 1)
- $P(a < X < b) = \int_a^b f(x)dx$ (probability = area)

Common confusion:

- $f(x)$ is NOT probability — it's *density*
- $f(x)$ can exceed 1 (think of a narrow, tall curve)
- Only the *area* under the curve represents probability

Think: probability = area under the PDF curve between two points.

Cumulative Distribution Function (CDF)

The CDF $F(x) = P(X \leq x)$ gives cumulative probability.
Key properties:

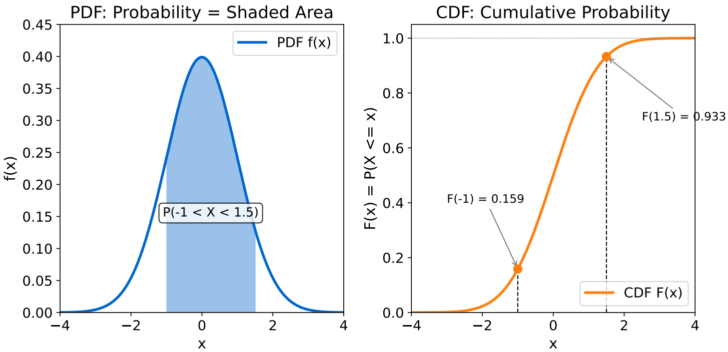
- $F(x)$ is non-decreasing: if $a < b$, then $F(a) \leq F(b)$
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
- For continuous X: $F(x) = \int_{-\infty}^{x} f(t)dt$

Practical calculations:

- $P(X > a) = 1 - F(a)$
- $P(a < X < b) = F(b) - F(a)$

CDF accumulates probability from left to right; use tables or software for values.

PDF and CDF Relationship



Left: shaded area = probability. Right: CDF value = total area up to that point.

Continuous Uniform Distribution

All values in interval $[a, b]$ equally likely:

$$X \sim \text{Uniform}(a, b)$$

PDF:

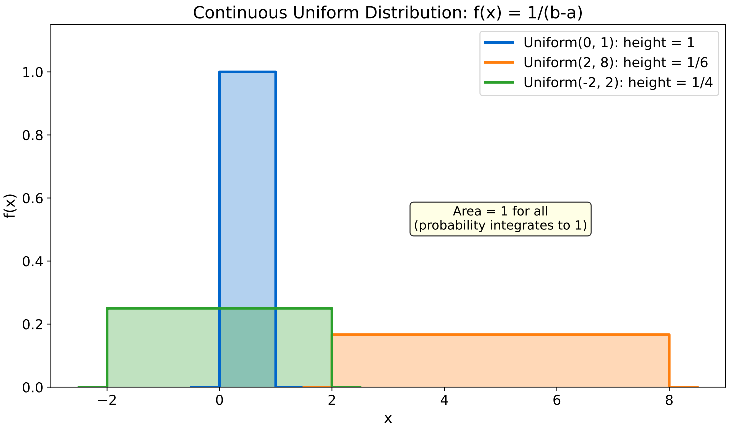
$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Properties:

- $E[X] = \frac{a+b}{2}$ (midpoint)
- $\text{Var}(X) = \frac{(b-a)^2}{12}$

Used in Monte Carlo simulation and when "all outcomes equally likely."

Uniform Distribution: Visualization



Height = 1/(b-a) ensures total area equals 1.

Normal (Gaussian) Distribution

The most important distribution in statistics!

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

Characteristics:

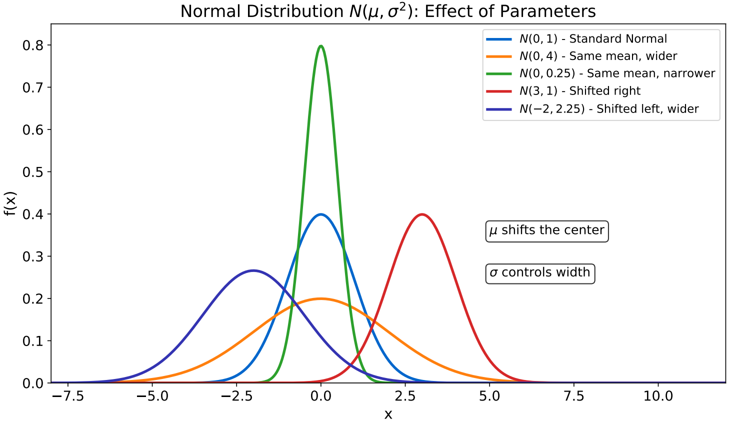
- Bell-shaped, symmetric around μ
- Tails extend to $\pm\infty$ but decay rapidly
- Completely characterized by mean μ and variance σ^2

PDF:

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

Central Limit Theorem: sums of many variables tend to normal.

Normal Distribution: Effect of Parameters



μ shifts location; σ controls spread (width of the bell).

Standard Normal Distribution

Special case: $Z \sim N(0,1)$

Standardization: Convert any $X \sim N(\mu, \sigma^2)$ to standard normal:

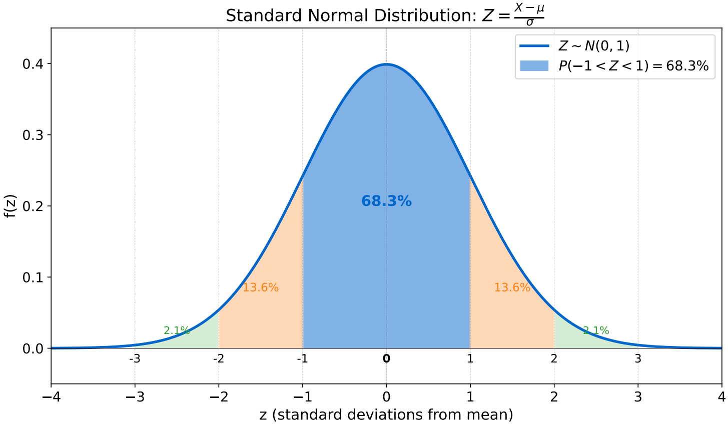
$$Z = \frac{X - \mu}{\sigma}$$

Z-score interpretation:

- $Z = 0$: at the mean
- $Z = 1$: one standard deviation above mean
- $Z = -2$: two standard deviations below mean

Z-tables give $P(Z \leq z)$; software: pnorm(), norm.cdf(), NORM.S.DIST().

Standard Normal: Z-Scores and Areas



Most probability concentrated within 2-3 standard deviations of the mean.

The Empirical Rule (68-95-99.7)

For any normal distribution:

- **68%** of data within 1 standard deviation of mean
- **95%** of data within 2 standard deviations
- **99.7%** of data within 3 standard deviations

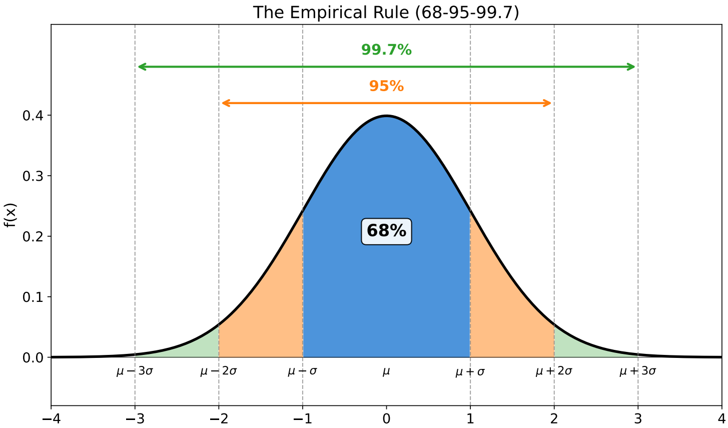
Finance application:

If daily returns $\sim N(0, 1.5\%)$:

- 68% of days: returns between -1.5% and $+1.5\%$
- 95% of days: returns between -3% and $+3\%$
- Moves beyond $\pm 4.5\%$ are rare (0.3% of days)

This rule helps quickly assess outliers and unusual observations.

Empirical Rule: Visualization



Memorize: 68-95-99.7 for quick probability assessments.

Log-Normal Distribution

Definition: X is log-normal if $\ln(X)$ is normally distributed:

$$X \sim \text{LogNormal}(\mu, \sigma^2) \Leftrightarrow \ln(X) \sim N(\mu, \sigma^2)$$

Key properties:

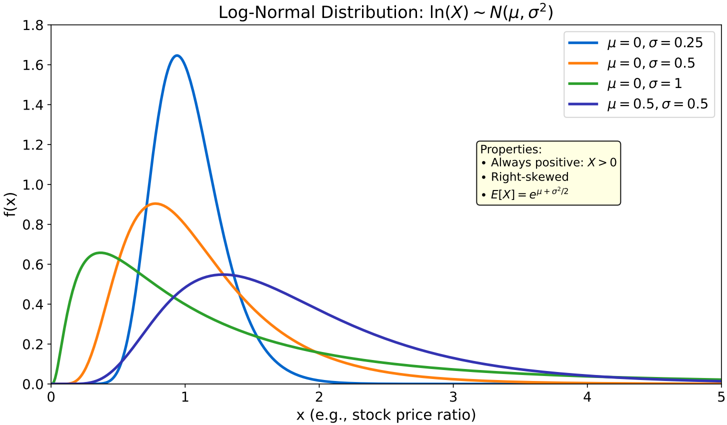
- Always positive: $X > 0$
- Right-skewed (long right tail)
- $E[X] = e^{\mu+\sigma^2/2}$ (note: NOT e^{μ}!)

Why finance uses log-normal:

- Stock prices cannot be negative
- Log-returns are approximately normal
- Black-Scholes assumes log-normal prices

If returns are normal, price levels are log-normal.

Log-Normal Distribution: Shapes



Higher σ creates more right-skew and heavier right tail.

Log-Normal in Finance: Stock Prices

Geometric Brownian Motion model:

$$S(T) = S(0) \cdot \exp\left(\left(\mu - \frac{\sigma^2}{2}\right) T + \sigma\sqrt{T} \cdot Z\right)$$

where $Z \sim N(0, 1)$.
Example: $S(0) = \$100$, $\mu = 8\%$, $\sigma = 20\%$, $T = 1$ year

- Expected price: $E[S(1)] = 100 \cdot e^{0.08+0.5 \cdot 0.04} = \$110.52$
- Median price: $100 \cdot e^{0.08} = \$108.33$
- $P(S > \$120)$: requires solving for $Z > 0.51 \Rightarrow 30.5\%$

Log-normal enforces positive prices; mean ¿ median due to right skew.

Exponential Distribution

Time until event occurs (continuous analog of geometric):

$$X \sim \text{Exponential}(\lambda)$$

PDF:

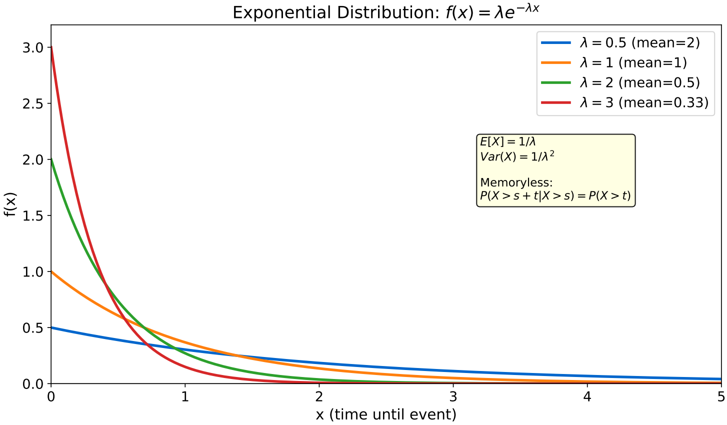
$$f(x) = \lambda e^{-\lambda x}, \quad x \geq 0$$

Properties:

- $E[X] = 1/\lambda$ (average waiting time)
- $\text{Var}(X) = 1/\lambda^2$
- CDF: $F(x) = 1 - e^{-\lambda x}$

λ = rate parameter (events per unit time).

Exponential Distribution: Visualization



Higher λ means more frequent events, shorter expected wait.

Memoryless Property

Unique to exponential (and geometric):

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

Interpretation:

- Past waiting time doesn't affect future probability
- "Starting fresh" at any point
- Constant hazard rate over time

Example: If average time between trades is 10 seconds, and you've already waited 20 seconds, the expected additional wait is still 10 seconds.

Memoryless = no "aging" or "wearing out" effect.

Exponential in Finance

Applications:

- Time between high-frequency trades
- Time until next limit order arrives
- Default time (simplified credit risk models)
- Interarrival times in Poisson processes

Example: If trades arrive at rate $\lambda = 120/\text{minute}$:

- Mean time between trades: $1/120$ minute $= 0.5$ seconds
- $P(\text{wait} > 1 \text{ second}) = e^{-120\cdot(1/60)} = e^{-2} \approx 13.5\%$

Exponential works when events occur randomly at constant rate.

Gamma Distribution

Sum of α independent exponential waiting times:

$$X \sim \text{Gamma}(\alpha, \beta)$$

PDF:

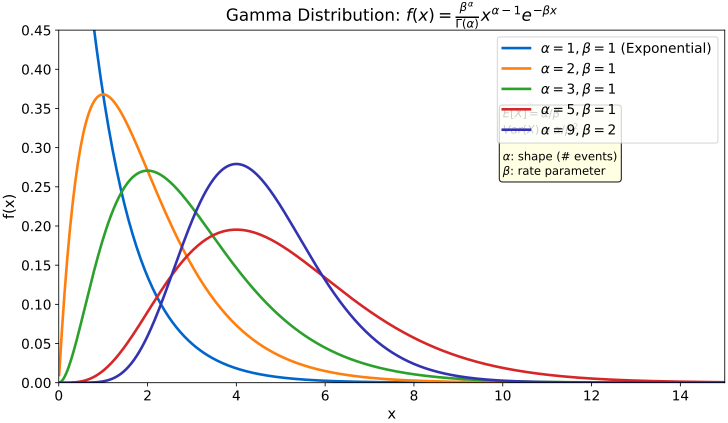
$$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0$$

Properties:

- $E[X] = \alpha/\beta$
- $\text{Var}(X) = \alpha/\beta^2$
- Special case: $\text{Gamma}(1, \beta) = \text{Exponential}(\beta)$

α = shape (number of events), β = rate.

Gamma Distribution: Shapes



As α increases, distribution becomes more symmetric and bell-shaped.

Weibull Distribution

Flexible model for failure times:

$$X \sim \text{Weibull}(k, \lambda)$$

PDF:

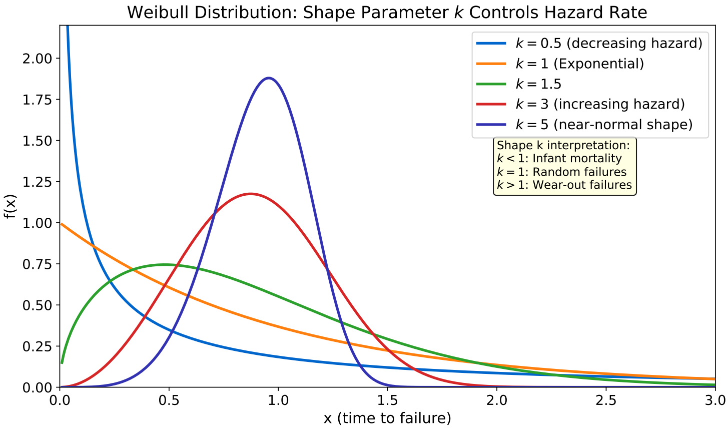
$$f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}, \quad x \geq 0$$

Shape parameter k interpretation:

- $k < 1$: Decreasing failure rate (infant mortality)
- $k = 1$: Constant failure rate (Exponential)
- $k > 1$: Increasing failure rate (wear-out)

Weibull generalizes exponential to non-constant hazard rates.

Weibull Distribution: Shape Variations



Shape k determines whether hazard increases or decreases over time.

Weibull in Finance and Risk

Applications:

- Operational risk: equipment failure times
- Credit risk: time-to-default modeling
- Insurance: claim arrival patterns
- Project management: task completion times

Why Weibull over Exponential?

- More realistic: failure rates often change over time
- $k > 1$: aging systems fail more as time passes
- $k < 1$: early failures decrease (burn-in period)

Weibull provides flexibility that exponential lacks.

Beta Distribution

Flexible distribution on the interval $[0, 1]$:

$$X \sim \text{Beta}(\alpha, \beta)$$

PDF:

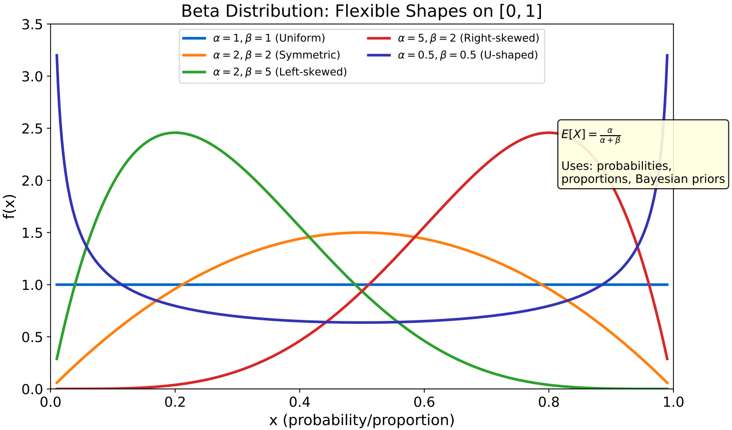
$$f(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}, \quad 0 \leq x \leq 1$$

Properties:

- $E[X] = \frac{\alpha}{\alpha+\beta}$
- $\text{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
- Special case: $\text{Beta}(1, 1) = \text{Uniform}(0, 1)$

Perfect for modeling probabilities, proportions, and percentages.

Beta Distribution: Shape Variations



Parameters α, β control skewness and concentration.

Beta Distribution Applications

Finance applications:

- Default probability estimation
- Recovery rate modeling (bounded between 0 and 1)
- Market share analysis
- Bayesian prior for unknown probabilities

Bayesian interpretation:

- $\alpha - 1$: “prior successes”
- $\beta - 1$: “prior failures”
- Posterior after s successes, f failures: Beta$(\alpha + s, \beta + f)$

Beta is the conjugate prior for binomial likelihood.

Student's t-Distribution

Used when estimating mean with unknown variance:

$$T \sim t_\nu \quad (\nu \text{ degrees of freedom})$$

Key characteristics:

- Symmetric around 0 (like standard normal)
- Heavier tails than normal (more probability in extremes)
- As $\nu \to \infty$: $t_\nu \to N(0,1)$

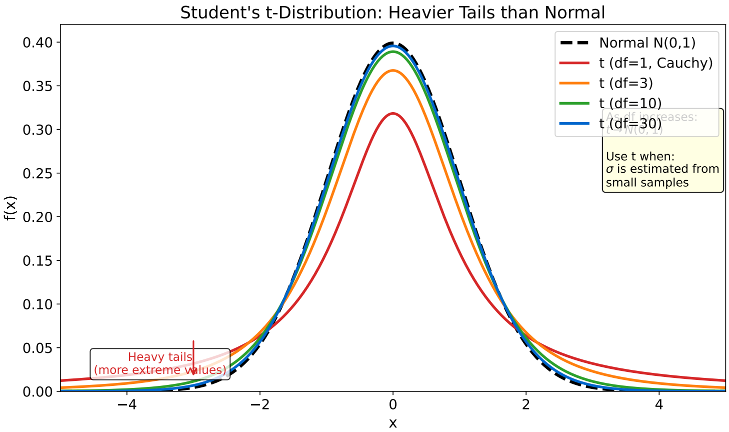
Origin:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

when sampling from a normal population.

Use t instead of z when sample size is small and σ is estimated.

t-Distribution vs Normal



Lower df = heavier tails = wider confidence intervals.

t-Distribution: Degrees of Freedom

How df affects the distribution:

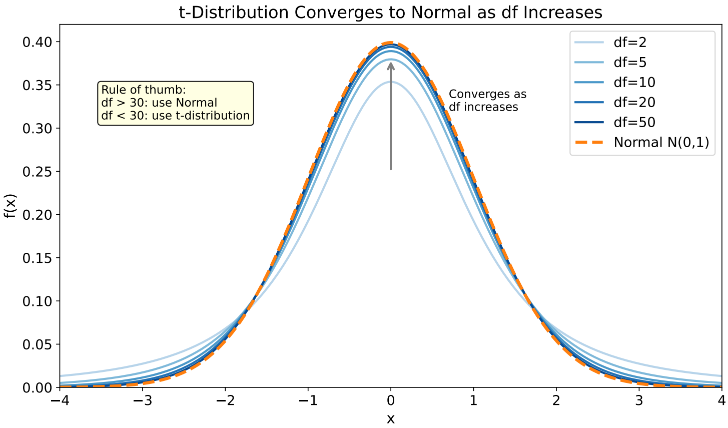
- $\mathbf{df = 1}$: Cauchy distribution (no mean!)
- $\mathbf{df = 2}$: Variance exists but infinite kurtosis
- $\mathbf{df = 30}$: Nearly indistinguishable from normal

Practical rule of thumb:

- $n < 30$: use t-distribution
- $n \geq 30$: normal approximation is reasonable

Critical values: $t_{0.025,10} = 2.23$ **vs** $z_{0.025} = 1.96$.

t-Distribution: Convergence to Normal



By df = 30, the t-distribution is practically normal.

Chi-Squared Distribution

Sum of squared standard normals:

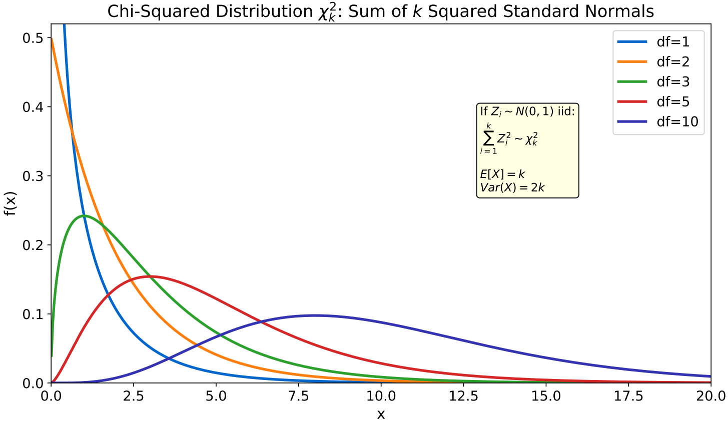
$$\chi_k^2 = \sum_{i=1}^{k} Z_i^2 \quad \text{where } Z_i \sim N(0,1) \text{ iid}$$

Properties:

- Support: $[0, \infty)$
- $E[\chi_k^2] = k$
- $\text{Var}(\chi_k^2) = 2k$
- Related to Gamma: $\chi_k^2 = \text{Gamma}(k/2, 1/2)$

Used in variance testing, goodness-of-fit, and contingency tables.

Chi-Squared Distribution: Shapes



As df increases, chi-squared becomes more symmetric.

Chi-Squared in Hypothesis Testing

Common applications:

- **Variance testing:** $(n-1)S^2/\sigma_0^2 \sim \chi^2_{n-1}$
- **Goodness-of-fit:** $\sum \frac{(O_i - E_i)^2}{E_i} \sim \chi^2_k$
- **Independence tests:** contingency table analysis

Finance example:

Testing if portfolio variance equals benchmark:

- H_0: $\sigma^2 = \sigma_0^2$ vs H_1: $\sigma^2 \neq \sigma_0^2$
- Test statistic: $(n-1)s^2/\sigma_0^2$
- Compare to χ^2_{n-1} critical values

Chi-squared tests are always one-sided (squared values are positive).

F-Distribution

Ratio of two chi-squared variables:

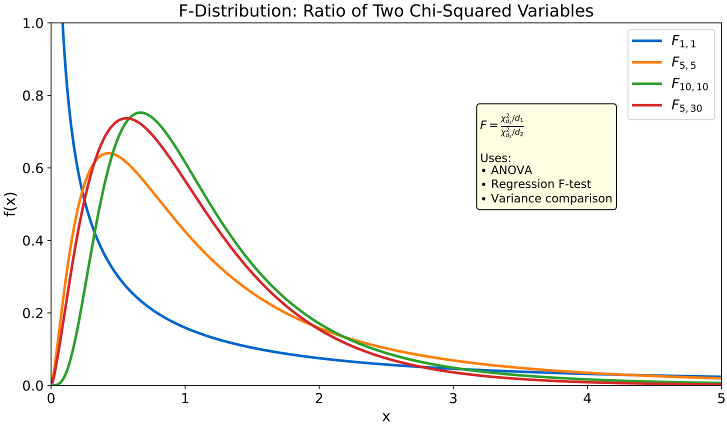
$$F = \frac{\chi^2_{d_1}/d_1}{\chi^2_{d_2}/d_2} \sim F_{d_1,d_2}$$

Properties:

- Support: $[0, \infty)$
- Two degrees of freedom: numerator (d_1) and denominator (d_2)
- $E[F] = \frac{d_2}{d_2 - 2}$ for $d_2 > 2$

Used for comparing variances and in ANOVA/regression.

F-Distribution: Shapes



Shape depends on both degrees of freedom parameters.

F-Distribution Applications

Primary uses:

- **ANOVA:** comparing means across groups
- **Regression F-test:** overall model significance
- **Variance comparison:** two-sample variance test

Regression F-test:

$$F = \frac{\text{Explained variance}/p}{\text{Residual variance}/(n-p-1)} \sim F_{p,n-p-1}$$

Interpretation:

- Large F: model explains significantly more than noise
- Compare to $F_{p,n-p-1}$ critical value

F-test asks: "Is the model better than just using the mean?"

Pareto Distribution

Power-law distribution for extreme values:

$$X \sim \text{Pareto}(\alpha, x_m)$$

PDF:

$$f(x) = \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}, \quad x \geq x_m$$

Key property — Power law tails:

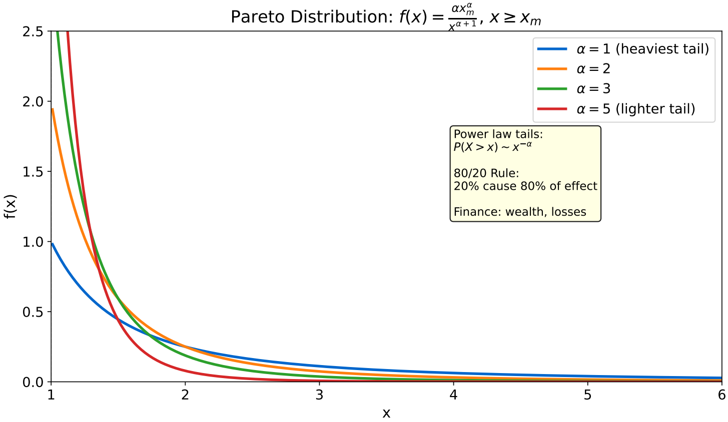
$$P(X > x) \sim x^{-\alpha}$$

Properties:

- $E[X] = \frac{\alpha x_m}{\alpha - 1}$ (exists only if $\alpha > 1$)
- $\text{Var}(X)$ exists only if $\alpha > 2$

Pareto: “80% of effects come from 20% of causes.”

Pareto Distribution: Visualization



Lower α = heavier tail = more extreme values.

Pareto in Finance

Applications:

- **Wealth distribution:** few have most wealth
- **Extreme losses:** large losses are more common than normal predicts
- **Insurance claims:** few large claims dominate
- **City sizes, firm sizes:** power-law patterns

80/20 Rule (Pareto Principle):

- 20% of clients generate 80% of revenue
- 20% of stocks drive 80% of market moves
- 20% of risks cause 80% of losses

Normal distributions severely underestimate extreme events.

Cauchy Distribution

The "pathological" distribution:

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

PDF:

$$f(x) = \frac{1}{\pi(1 + x^2)}$$

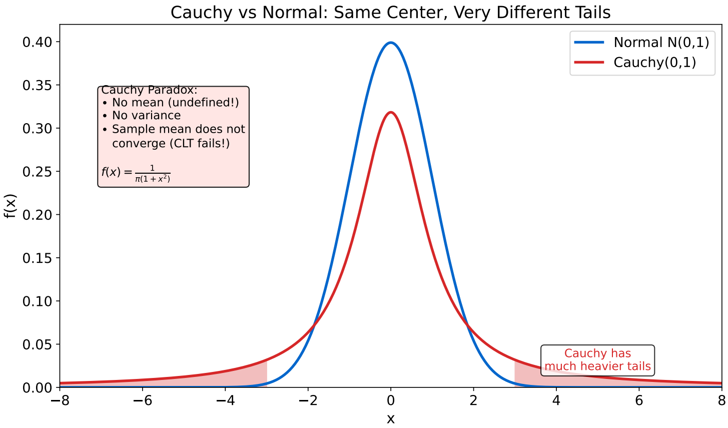
Pathological properties:

- **No mean** — integral does not converge
- **No variance** — even worse!
- **Sample mean does not converge** — CLT fails!

Origin: Ratio of two independent standard normals: $Z_1/Z_2 \sim \text{Cauchy}$

Cauchy shows that not all distributions have finite moments.

Cauchy vs Normal: Heavy Tails



Cauchy looks similar at center but has dramatically heavier tails.

Cauchy: Why It Matters

Lessons from Cauchy:

- Not all distributions have means or variances
- Averaging does NOT always help (sample mean is as variable as one observation!)
- CLT has conditions that can fail

Finance relevance:

- Price-to-earnings ratios (ratio of random variables)
- Some hedge fund return distributions
- Warns against assuming normality for everything

When you see "infinite variance," think: sample average is unreliable.

Laplace Distribution

Double exponential (symmetric exponential tails):

$$X \sim \text{Laplace}(\mu, b)$$

PDF:

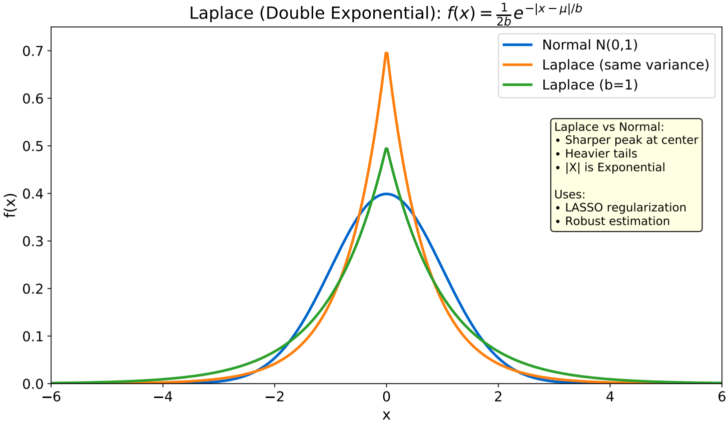
$$f(x) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$$

Properties:

- Symmetric around μ
- Sharper peak than normal, heavier tails
- $E[X] = \mu$, $\text{Var}(X) = 2b^2$

Also called "double exponential" — exponential in both directions.

Laplace Distribution: Visualization



Laplace vs Normal:
- Sharper peak at center
- Heavier tails
- |X| is Exponential

Uses:
- LASSO regularization
- Robust estimation

Sharper peak at center, heavier tails than normal.

Laplace in Statistics and Finance

Applications:

- **LASSO regularization:** Laplace prior induces sparsity
- **Robust estimation:** median minimizes Laplace likelihood
- **Currency returns:** better fit than normal for some pairs

Why Laplace for LASSO?

- Laplace prior on coefficients $\Rightarrow$ L1 penalty
- Encourages coefficients to be exactly zero
- Creates sparse, interpretable models

Connection: $|X| \sim$ **Exponential when** $X \sim$ **Laplace**$(0, b)$**.**

Q-Q Plots: Concept

Quantile-Quantile plot: Compare sample quantiles to theoretical quantiles.
How to read:

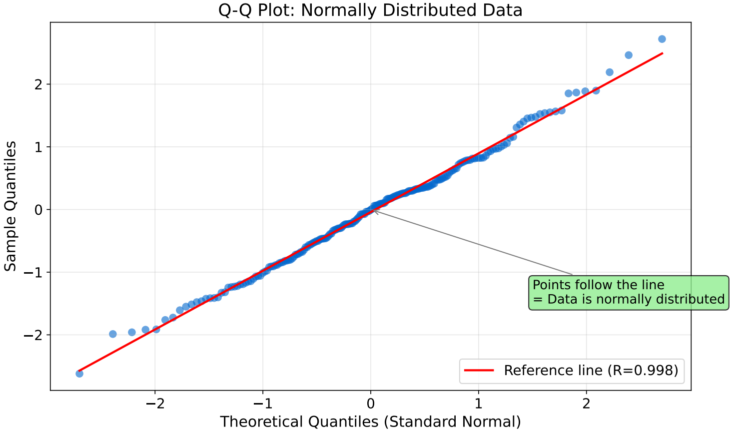
- Points on diagonal line: data follows the theoretical distribution
- Systematic deviations: data differs from theoretical
- S-curve at tails: heavy-tailed data
- Curved pattern: skewed data

Construction:

1. Sort data: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$
2. Compute theoretical quantiles for same probabilities
3. Plot sample quantiles vs theoretical quantiles

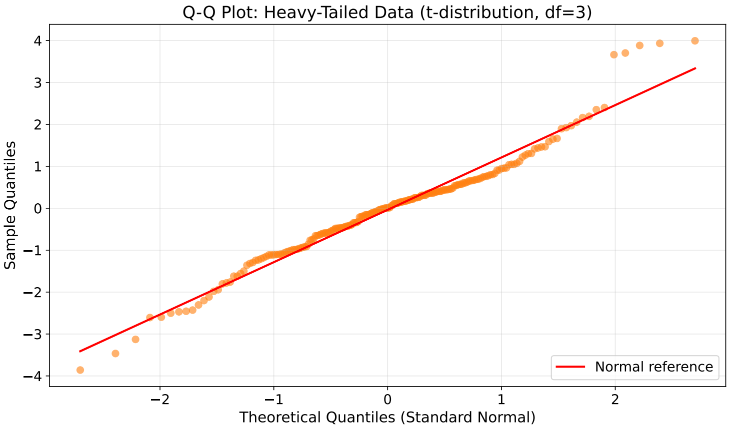
Q-Q plots are the standard diagnostic for normality.

Q-Q Plot: Normal Data (Good Fit)



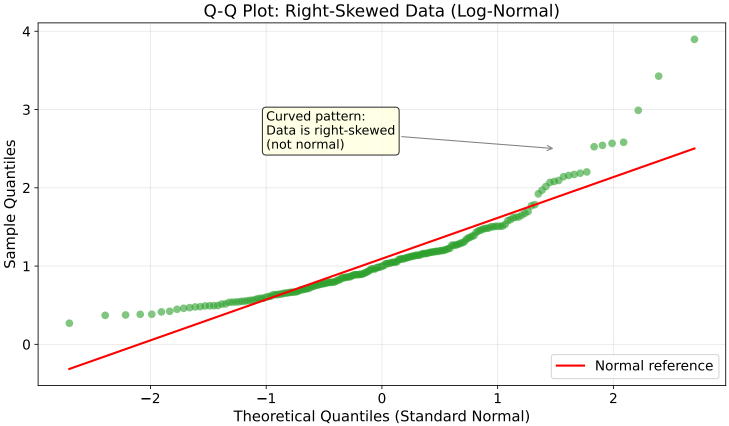
Points follow the line = data is well-approximated by normal.

Q-Q Plot: Heavy-Tailed Data



S-curve at tails = more extreme values than normal predicts.

Q-Q Plot: Skewed Data



Curved pattern = data is skewed, not symmetric like normal.

Choosing the Right Distribution

Decision guide:

Data characteristic	**Consider**
Symmetric, bounded tails	Normal
Positive values only, right-skewed	Log-normal, Gamma, Weibull
Waiting times, constant hazard	Exponential
Waiting times, changing hazard	Weibull, Gamma
Proportions, probabilities	Beta
Heavy tails, extreme values	Pareto, t-distribution
Small sample, unknown σ	t-distribution
Variance testing	Chi-squared
Comparing variances/groups	F-distribution

Always check fit with Q-Q plots and goodness-of-fit tests.

Distribution Summary Table

Distribution	**Support**	**Mean**	**Variance**	**Use**
Uniform(a, b)	$[a, b]$	$(a+b)/2$	$(b-a)^2/12$	Simulation
Normal(μ, σ^2)	$\mathbb{R}$	μ	σ^2	General
Log-Normal	$(0, \infty)$	$e^{\mu+\sigma^2/2}$	complex	Prices
Exponential(λ)	$[0, \infty)$	$1/\lambda$	$1/\lambda^2$	Waiting
Gamma(α, β)	$[0, \infty)$	α/β	α/β^2	Aggregated
Weibull(k, λ)	$[0, \infty)$	complex	complex	Reliability
Beta(α, β)	$[0, 1]$	$\alpha/(\alpha+\beta)$	complex	Proportions

---

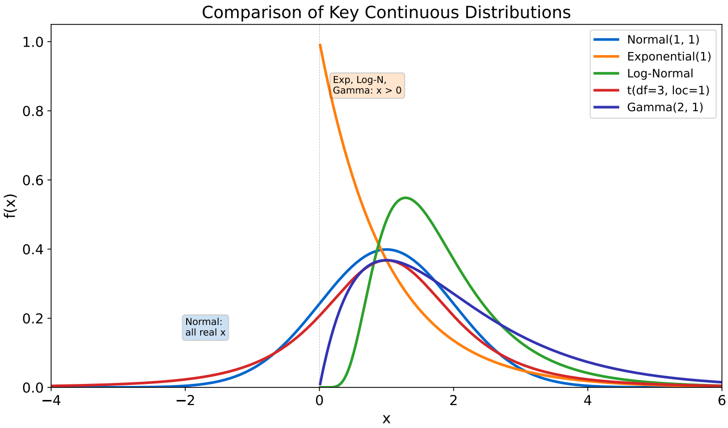
Part 1 of 2 — location-scale and waiting time distributions.

Distribution Summary Table (continued)

Distribution	**Support**	**Mean**	**Variance**	**Use**
Student-t$_\nu$	$\mathbb{R}$	0	$\nu/(\nu-2)$	Small samples
χ^2_k	$[0,\infty)$	k	$2k$	Variance tests
F_{d_1,d_2}	$[0,\infty)$	$d_2/(d_2-2)$	complex	ANOVA
Pareto(α, x_m)	$[x_m,\infty)$	$\alpha x_m/(\alpha-1)$	varies	Extremes
Cauchy	$\mathbb{R}$	undefined	undefined	Pathological
Laplace(μ, b)	$\mathbb{R}$	μ	$2b^2$	Robust/LASSO

Part 2 of 2 — inference and heavy-tailed distributions.

Distribution Comparison



Different distributions for different data characteristics.

Finance Application: Value at Risk (VaR)

VaR: Maximum expected loss at given confidence level.

$$\text{VaR}_\alpha = \mu - z_\alpha \cdot \sigma$$

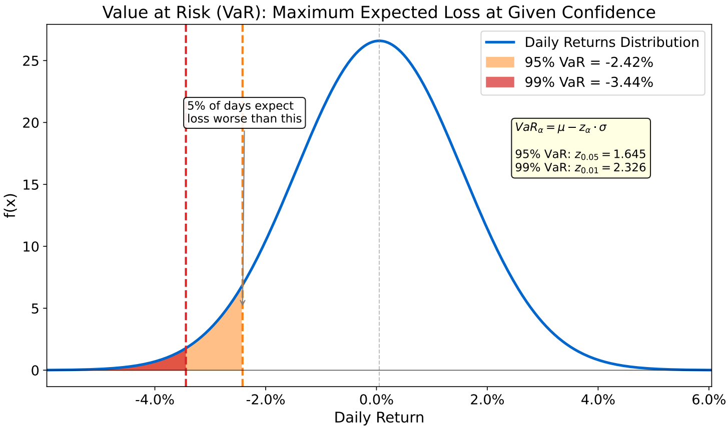
Example: Daily returns $\sim N(0.05\%, 1.5\%)$, portfolio = $1M

- 95% VaR: $0.05\% - 1.645 \times 1.5\% = -2.42\% \Rightarrow \$24{,}175$
- 99% VaR: $0.05\% - 2.326 \times 1.5\% = -3.44\% \Rightarrow \$34{,}390$

Interpretation: "5% of days, we expect to lose more than $24,175."

VaR assumes normality; fat tails require t or other distributions.

VaR Visualization



VaR is the quantile of the loss distribution.

Finance Application: Option Pricing

Black-Scholes assumes log-normal prices:

$$S_T = S_0 \exp\left(\left(r - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T} \cdot Z\right)$$

Call option price:

$$C = S_0\Phi(d_1) - Ke^{-rT}\Phi(d_2)$$

where Φ is the standard normal CDF. **Key insight:**

- Normal distribution appears via Φ
- Log-normal price distribution underlies the model

Understanding normal/log-normal is essential for derivatives pricing.

Common Misconceptions

Misconception 1: “PDF height = probability”

- Reality: Probability is AREA, not height

Misconception 2: “Normal distribution works for everything”

- Reality: Financial data often has fat tails

Misconception 3: “Sample mean always converges”

- Reality: Cauchy shows CLT has conditions

Misconception 4: “$N(\mu, \sigma)$”

- Reality: Second parameter is VARIANCE σ^2, not SD

Understanding these details prevents costly errors.

Key Takeaways

1. **PDF $\neq$ probability** — probability is area under the curve
2. **Normal distribution** is central but not universal
3. **Log-normal** is essential for positive quantities (prices)
4. **Exponential** is memoryless; Weibull/Gamma generalize it
5. **Beta** is perfect for proportions and probabilities
6. **t-distribution** handles small samples and uncertainty
7. **Heavy tails** (Pareto, Cauchy) warn against normality assumptions
8. **Q-Q plots** are the standard diagnostic for distribution fit

---

Match distribution to data characteristics, not convenience.

References and Further Reading

Textbooks:

- Casella & Berger: *Statistical Inference*
- Ross: *Introduction to Probability Models*
- McNeil, Frey & Embrechts: *Quantitative Risk Management*

Finance applications:

- Hull: *Options, Futures, and Other Derivatives*
- Jorion: *Value at Risk*

Software:

- Python: scipy.stats (all distributions)
- R: built-in d/p/q/r functions

Next: Joint Distributions (L07) or Hypothesis Testing (A0).