# Lesson 14 Summary: Distributions

## Data Science with Python – Key Concepts

Data Science Program

# Distributions Overview

## Probability Distributions

| Normal | Binomial | t-dist | Log-normal |
|---|---|---|---|
| N(mu, sigma) | B(n, p) | t(df) | LN(mu, s) |
| Bell curve | Discrete | Fat tails | Prices |

**Key Concepts:**

PDF | CDF | Quantiles | Q-Q plots | Goodness of fit

*Financial returns have fat tails - not normal!*

**Understanding distributions is key to statistical analysis**

# Normal Distribution

**The bell curve – most important distribution:**

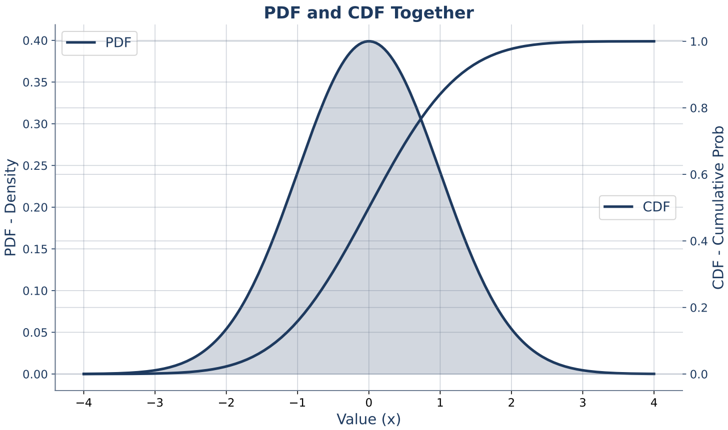
- Symmetric around mean
- Defined by $\mu$ (mean) and $\sigma$ (std)
- 68-95-99.7 rule for standard deviations

**Python:**

```
from scipy import stats
stats.norm.pdf(x, loc=mu, scale=sigma)
```

**Many natural phenomena follow normal distribution**

# PDF and CDF




**PDF: probability density; CDF: cumulative probability**

# Binomial Distribution

**Discrete outcomes with fixed probability:**

- n trials, each with probability p of success
- Number of successes in n trials
- Example: Coin flips, trading win rates

**Python:**

```
stats.binom.pmf(k, n=10, p=0.5) # P(X=k)
```

**Converges to normal as n increases**

# t-Distribution

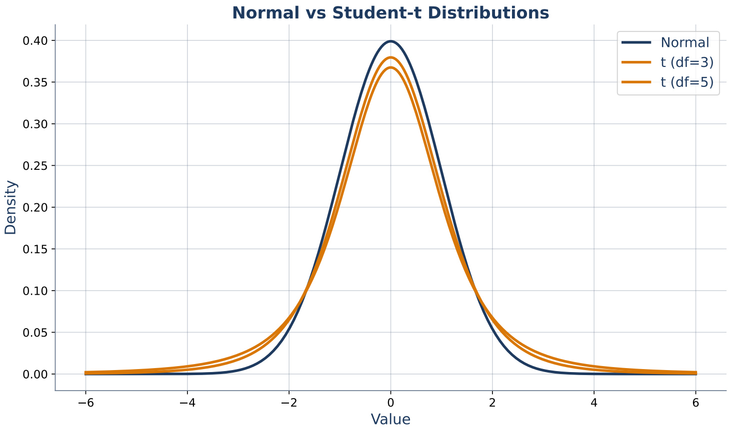
**Like normal but with heavier tails:**

- Used when sample size is small
- Degrees of freedom (df) parameter
- Converges to normal as df increases

**Finance:** Better model for stock returns than normal

**t-distribution captures fat tails in financial data**

# Fat Tails




**Financial returns have fatter tails than normal predicts**

# Q-Q Plots

**Compare empirical vs theoretical distribution:**

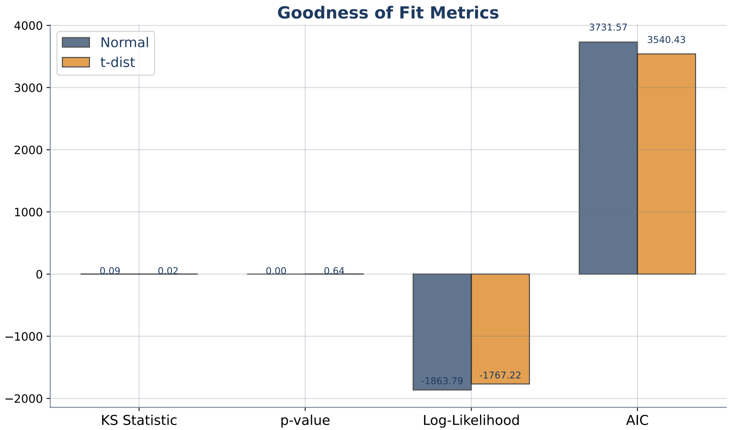
- Points on diagonal = good fit
- Deviations at tails = fat tails
- S-curve = skewness

**Python:**

```
from scipy import stats
stats.probplot(data, dist="norm", plot=plt)
```

**Q-Q plots reveal departures from normality**

# Goodness of Fit




**Statistical tests verify distribution assumptions**

# Quick Reference

**Essential Distribution Operations:**

| Distribution | scipy.stats |
|---|---|
| Normal | `stats.norm(loc, scale)` |
| t-distribution | `stats.t(df, loc, scale)` |
| Binomial | `stats.binom(n, p)` |
| Lognormal | `stats.lognorm(s, loc, scale)` |
| **Method** | |
| PDF | `.pdf(x)` |
| CDF | `.cdf(x)` |
| Quantile | `.ppf(q)` |
| Random | `.rvs(size=n)` |

**scipy.stats provides comprehensive distribution support**