Introduction to Probability and Statistics

Lesson 01

Digital Finance

Outline

What is Statistics?

Statistics is the science of:

- Collecting, organizing, and summarizing data
- Drawing conclusions from data
- Making decisions under uncertainty

Two main branches:

- **Descriptive Statistics**: Summarizing and visualizing data
- **Inferential Statistics**: Drawing conclusions about populations from samples

Statistics helps us make sense of data and quantify uncertainty.

Why Study Statistics?

Everyday decisions under uncertainty:

- Will it rain tomorrow? (weather forecasts)
- Is this medical treatment effective? (clinical trials)
- Which route has less traffic? (navigation apps)

In business and finance:

- Which marketing campaign works better? (A/B testing)
- How risky is this investment? (risk assessment)
- What sales can we expect? (forecasting)

Statistics helps us make better decisions when outcomes are uncertain.

What is Probability?

Intuition: Probability measures *how likely* something is to happen.

- "Chance," "likelihood," "odds" – everyday words for probability
- Written as P(event), a number between 0 (impossible) and 1 (certain)
- Example: $P(\text{heads}) = 0.5$ means heads occurs 50% of the time

Three interpretations:

- **Frequentist**: Long-run frequency (flip a coin 1000 times $\rightarrow$ 500 heads)
- **Bayesian**: Degree of belief ("I'm 70% confident it will rain")
- **Classical**: Equally likely outcomes (dice: 1/6 chance for each face)

Probability gives us a precise language to quantify uncertainty.

Simple Examples First

Coin flip:

- Two outcomes: Heads or Tails
- P(heads) = 0.5 – each equally likely
- Flip 10 times: expect 5 heads, but could get 3 or 7

Rolling a die:

- Six outcomes: 1, 2, 3, 4, 5, 6
- P(six) $= 1/6 \approx 0.167$
- Roll 60 times: expect 10 sixes

Connection to finance:

- Stock goes up or down $\rightarrow$ like a coin flip!
- Many possible returns $\rightarrow$ like rolling a die with many sides

---

Simple random processes help us understand complex financial markets.

Probability vs Statistics

Probability (Forward reasoning):

- Given a known model, what outcomes are likely?
- "If I flip a fair coin 10 times, what is P(7 heads)?"
- Model $\rightarrow$ Data

Statistics (Backward reasoning):

- Given observed data, what model explains it?
- "I got 7 heads in 10 flips. Is the coin fair?"
- Data $\rightarrow$ Model

Probability and statistics are two sides of the same coin.

The Relationship

	Probability	**Statistics**
Direction	Model $\rightarrow$ Data	Data $\rightarrow$ Model
Question	What will happen?	What happened?
Known	Model parameters	Observed data
Unknown	Outcomes	Parameters

Coin flip example:

- Probability: "If P(heads) = 0.5, how likely is 7 heads in 10 flips?"
- Statistics: "I got 7 heads in 10 flips. Is P(heads) really 0.5?"

Probability assumes the model; statistics infers it from data.

A Brief History of Statistics

Key figures who shaped the field:

- **Blaise Pascal** (1654): Founded probability theory through gambling problems
- **Thomas Bayes** (1763): Developed Bayesian reasoning (updating beliefs with evidence)
- **Carl Friedrich Gauss** (1809): Discovered the normal distribution ("bell curve")
- **Ronald Fisher** (1920s): Created modern hypothesis testing and experimental design

Evolution of applications:

- Gambling → Astronomy → Medicine → Finance
- Today: Machine learning, AI, and data science all build on statistics

Statistics evolved from games of chance to the foundation of modern data science.

Uncertainty is Everywhere

Everyday uncertainty:

- Will my bus arrive on time?
- Will this job candidate perform well?
- How long will this battery last?

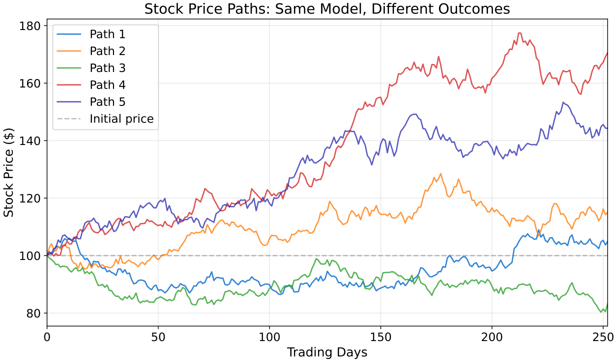
Why quantifying uncertainty matters:

- Uncertainty cannot be eliminated, only measured
- Better decisions require knowing "how uncertain"
- Probability gives us precise language for uncertainty

Statistics transforms "I don't know" into "Here's my best estimate with confidence bounds."

What is a random walk?

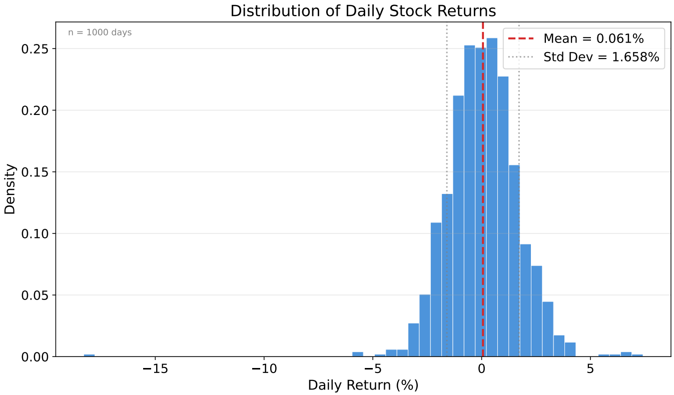
- Each step is unpredictable – like coin flips deciding up or down
- Same rules can produce vastly different paths
- **Volatility** = how much prices jump around (high volatility = big swings)



Five simulations from the same model – each path is equally likely.

Understanding Return Distributions

What is a return? The percentage change in price (e.g., +5% gain).
What is a distribution? How values spread out – which outcomes are common vs rare.



Most returns are small (center); extreme gains/losses (tails) are rare but matter.

Population vs Sample

Population:

- The entire group we want to study
- Often too large or impossible to measure completely
- Example: All stock returns ever

Sample:

- A subset of the population that we actually observe
- Must be representative to draw valid conclusions
- Example: Returns from the past 5 years

We use samples to learn about populations.

Parameters vs Statistics

Parameter (what we want to know):

- A fixed value describing the entire population
- Usually unknown – that's why we need statistics!
- Notation: Greek letters (μ = true mean, σ = true spread)

Statistic (what we can calculate):

- A value calculated from our sample data
- Notation: Latin letters ($\bar{x}$ = sample mean, s = sample spread)
- The "hat" symbol means "estimate of" ($\hat{\mu}$ estimates μ)

We use statistics (from samples) to estimate parameters (of populations).

Types of Data

Quantitative (Numerical):

- **Continuous**: Can take any value (stock prices, returns)
- **Discrete**: Can only take specific values (number of trades)

Qualitative (Categorical):

- **Nominal**: Categories without order (sector: tech, finance, health)
- **Ordinal**: Categories with order (credit rating: AAA, AA, A)

The type of data determines which statistical methods to use.

What We Will Cover

Part I: Foundations (Lessons 1-2)

- Introduction and descriptive statistics

Part II: Probability Theory (Lessons 3-8)

- Probability axioms, conditional probability
- Random variables and distributions
- Joint distributions, expectation

Part III: Statistical Inference (Lessons 9-12)

- Sampling and the Central Limit Theorem
- Estimation and hypothesis testing

Each concept builds on the previous ones.

Summary

Key takeaways from this lesson:

- Statistics helps us make decisions under uncertainty
- Probability reasons forward (model to data)
- Statistics reasons backward (data to model)
- We use samples to learn about populations

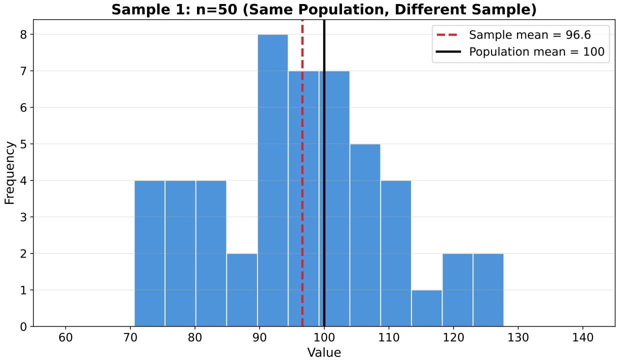
Next lesson: Descriptive Statistics – how to summarize and visualize data

---

Lesson 01 complete.

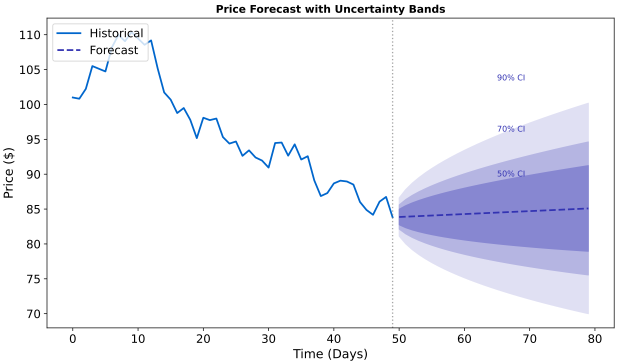
Sample Variability

Key insight: Different samples from the same population give different estimates.



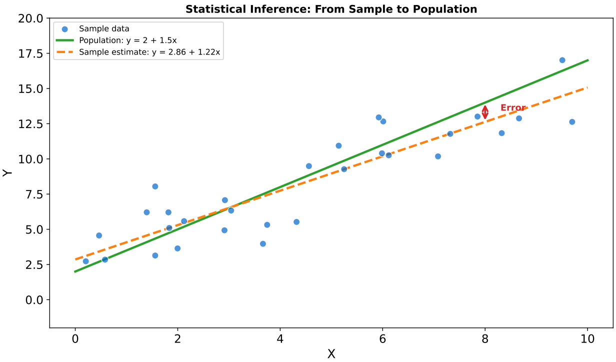
This is just ONE sample. Another sample would give a different mean!

Uncertainty in Forecasting



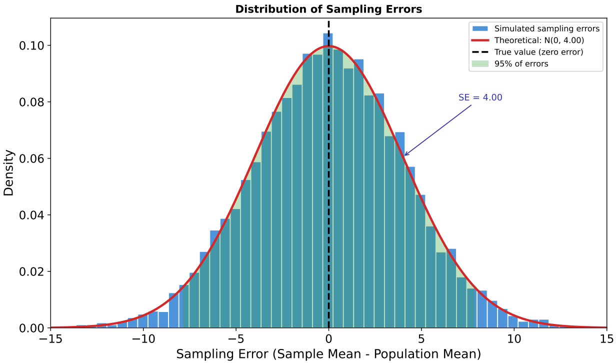
Uncertainty grows as we forecast further into the future.

Statistical Inference



We use sample data to estimate population relationships.

Sampling Error Distribution



Sampling errors follow a predictable distribution around zero.