Normal Models, the Sharpe Ratio, the CLT, Sampling Distributions

Reeves Johnson

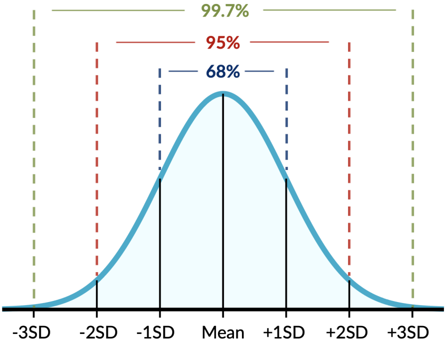
University of Virginia
McIntire School of Commerce

The Normal Distribution

The Normal Distribution and the 'Empirical Rule'

NB$_1$: Observing a value more than or less than 2 standard deviations from the mean is *rare*.

NB$_2$: The standard deviations shown are not exact. For example, 95% of the data are within 1.96 standard deviations, not 2.



University of Virginia
McIntire School of Commerce

The Normal Model and 'Rare' Events

- Consider the distribution of Sharpe ratios for the top 100 corporations from the S&P 500 (measured by market cap).
- Using the Normal model to approximate the distribution, how should we determine when a corporation has an extremely high or low Sharpe ratio? We need a rule . . .
- Signifcance levels (α) and p-values

Monte Carlo Simulation

A Brief Introduction to Monte Carlo Simulation

- Monte Carlo simulation: A computational technique using repeated random sampling to obtain numerical results
- Particularly useful for decision making in uncertain circumstances:
  - Risk management
  - Portfolio optimization
- Example: Appraising CapEx Projects
  - Inputs: Estimated cash flows, discount rates and time horizon
  - Process: Simulate various scenarios with different cash flows and economic conditions
  - Output: Range of possible NPVs

University of Virginia McIntire School of Commerce

Application: Portfolio Diversification

Using the same sample of the top 100 S&P corporations' Sharpe ratios, let's simulate *portfolios* to observe the effect diversification has on returns.

Here's the simulation recipe:

Step 1: Draw 10 random Sharpe ratios from our sample of 100 Sharpe ratios.

Step 2: Compute the mean Sharpe ratio for the 10 Sharpe ratios drawn.

Step 3: Repeat above steps 10,000 times to create a portfolio distribution of S&P top 100 corporations. (Sampling with replacement.)

Analysis of the Simulated Distribution

Compare the 2 distributions:

- Takeaways regarding the shapes, central tendencies and variabilities of the distributions?
- Why does the portfolio distribution tend towards a normal distribution?
- What if I increased the number of Sharpe ratios in each simulated portfolio to 30?
- This tendency towards normality is why many financial models assume normally distributed returns for diversified portfolios.

When Things Fall Apart

Yet, in periods of systemic financial distress (e.g., the Great Financial Crisis of 2008-09) such models were an accelerator of the distress. Why? There's a key assumption we made to arrive at a near-Normal distribution.

The Central Limit Theorem (CLT)

The Central Limit Theorem (CLT)

Mathematical reason for why the simulated portfolio distribution is near-Normal: **The Central Limit Theorem**.

Central Limit Theorem

Given certain conditions hold, the distribution of sample means approaches a Normal distribution as the sample size increases, regardless of the parent distribution.

It's the reason why:

- Normal models are ubiquitous in statistics.
- Statistical inference (our ability to generalize from samples) is so powerful.
- We have guidelines on minimal sample sizes.

Sampling Distributions

Keep these 3 Distributions Conceptually Distinct!

1. **Sample Distribution:** A distribution of sample data
2. **Population Distribution:** A distribution of the whole population
3. **Sampling Distribution:** A distribution of a *statistic* (e.g., sample means, $\bar{x}$)

The Classical Approach to Sampling Distributions

Central Limit Theorem (CLT): The sampling distribution of the sample mean approaches a normal distribution as the sample size, n, increases (regardless of the shape of the population distribution).

Build intuition on the CLT by simulating some sampling distributions of the sample mean, $\bar{x}$.

On the following site, click Central Limit Theorem:

Seeing Theory: The Central Limit Theorem

The CLT and Sampling Distributions

- The standard deviation of the sample means is known as the **standard error of the mean** and is given by:

$$\mathrm{se}_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

The CLT and Sampling Distributions

- The standard deviation of the sample means is known as the **standard error of the mean** and is given by:

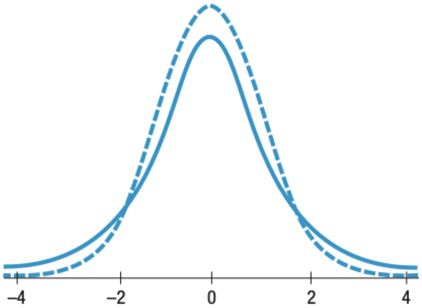
$$\text{se}_{\bar{x}} = \frac{\sigma}{\sqrt{n}}.$$

- You rarely know σ so you must *estimate* it.
- Use the *sample* standard deviation, s, to estimate the population standard deviation, σ.
- **Sample Standard Deviation:** Square root of the mean of the squared deviations:

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}.$$

A t-Distribution and the Standard Normal Distribution

- t-distributions replace Normal sampling distributions to account for the additional uncertainty of estimating σ.
- Compared to Normal distribtuions, t-dstributions have narrower peaks and heavier tails.



Dashed Curve: Standard Normal Distribution
Solid Curve: t-Distribtuion

So What, CLT? Confidence Intervals

Confidence Interval: An interval of *plausible* values a parameter of interest (e.g., μ) might assume.

To construct confidence intervals, you must know the standard error of the statistic/estimate, which you get from the CLT.

The confidence interval of $\bar{x}$ is generically written as:

$$\bar{x} \pm \text{Margin of Error}$$
$$\bar{x} \pm \text{Critial Value} \times \text{se}_{\bar{x}}$$

The **margin of error** depends on 2 pieces of information:

1. A *critical value*: Number of standard errors you must stretch out on either side of $\bar{x}$. The number of standard errors depends on your chosen *confidence level*.
2. The standard error of the sample mean, $\text{se}_{\bar{x}}$

Critical Values and Confidence Intervals

Critical Value: The *number of standard errors* you must stretch out on either side of $\bar{x}$ to construct a confidence interval.

Confidence Level: The long-run proportion of confidence intervals that contain the *true* parameter value; expressed as a percentage.

Critical Values for Standard Normal Distribution

CL	z^*
90%	1.65
95%	1.96
99%	2.58