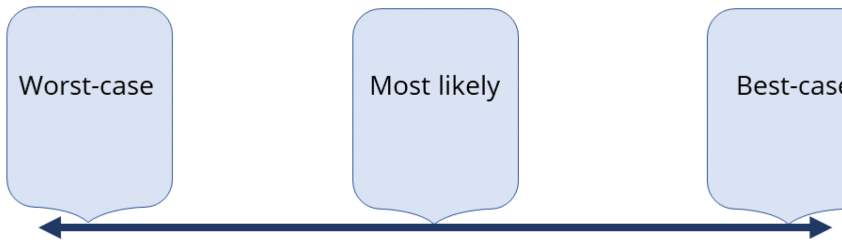
M8L2. Monte Carlo Simulation

Slide #1

Monte Carlo Simulation.

Slide #2

Review



2

In our scenario analysis, we evaluated three distinct possibilities: a most-likely scenario complemented by two less likely extremes- the best-case and worst-case scenarios.

By examining a spectrum of high probability scenarios within these outer limits, we shall gain a deeper understanding of the uncertainties and their impacts on the business.

Expanding the number of scenarios assessed allows us to gather more insights, thereby enhancing the informed decision-making process by considering a broad range of potential futures.

Slide #3

Monte Carlo Simulation

A method of estimating the value of an unknown quantity using the principles of inferential statistics.

Working Principle:

- Construct a digital model with random variables
- Run the model repeatedly with random inputs

Result:

A statistical distribution of outcomes and their probabilities.

3

The Monte Carlo simulation is a statistical technique that uses repeated random sampling to estimate the probabilities for a variety of outcomes.

This method constructs a digital model of a specific business case incorporating variables that are uncertain.

The process then runs the model repeatedly, each time with random inputs aligned with the statistical distributions of these uncertain variables.

This iterative process generates a range of outcomes and the probabilities they occur, establishing a statistical distribution of results.

Slide #4

The Monte Carlo Method

4

The innovative Monte Carlo method, a statistical approach for simulating uncertainties in business scenarios, was invented by the mathematician Stanislaw Ulam.

While contributing to nuclear weapons research at the prestigious Los Alamos National Lab, Ulam developed this method.

While contributing to nuclear weapons, research at the prestigious Los Alamos National Lab, Ulam developed this method.

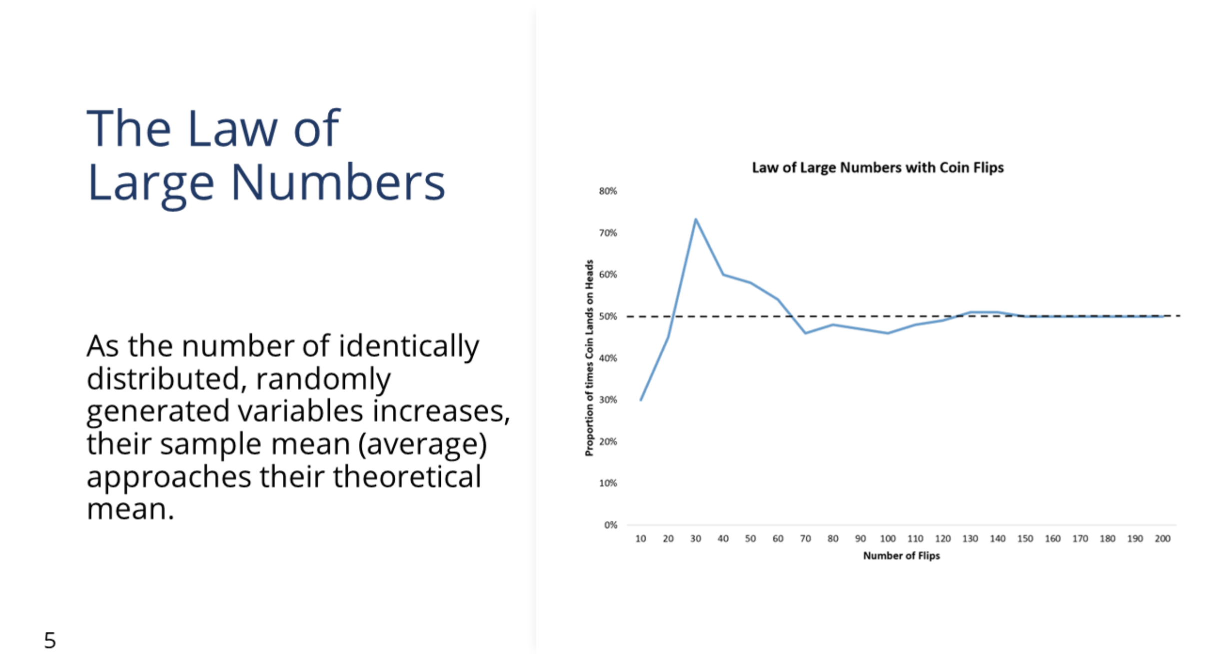
He collaborated with John von Neumann to leverage the ENIAC computer, marking the dawn of the first fully automated statistical analysis.

Due to the sensitive nature of their work, a discrete code name was essential.

Thus, Monte Carlo was chosen, inspired by the renowned Monte Carlo Casino in Monaco, a place familiar to Ulam through a relative's adventures in gambling.

This name fittingly captures the essence of the method, making informed decisions amidst the unpredictable twists and turns of business.

Slide #5



To understand the method, we begin with a foundational overview of the statistical principles it's built upon.

The law of large number theorem. The law of large numbers, introduced by Helvetia, is a fundamental concept in inferential statistics.

It suggests that as we increase the number of randomly generated, identically distributed variables, their average or sample mean gets closer to the theoretical average.

In simpler terms, the more data we collect, the more our observed results align with the true average. Consider the act of flipping a coin multiple times.

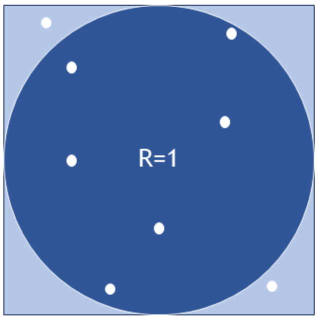
Over time, we expect to see a roughly equal distribution of heads and tails, about 50 percent each time. For decision makers, this principle underscores the value of extensive information gathering and analysis.

It ensures that our insights become clearer and more reliable, allowing us to make decisions rooted in consistent trends, rather than being swayed by outliers or random variations.

Slide #6

Large Number Theorem - Application

Estimate π using large number theorem



Area (square) = $2 \times 2 = 4$
Area (Circle) = π

Probability of the dart falling in the circle vs. in the square (inclusive):

$$\frac{Area\ (Circle)}{Area\ (square)} = \frac{\pi}{4}$$

$\pi = 4 \times$ Probability (in circle vs. in square)

Randomly generate N pairs of variables X and Y [-1,1]
Total Number of (X,Y) $X^2 + Y^2 \leq 1$ represents the area of "circle"
Total number of (X,Y) represents the area of "square"

6

Utilizing the law of large numbers offers a clever method to estimate pi, showcasing how data analytics simulations can produce precise outcomes.

Picture a square with a circle neatly fitted inside.

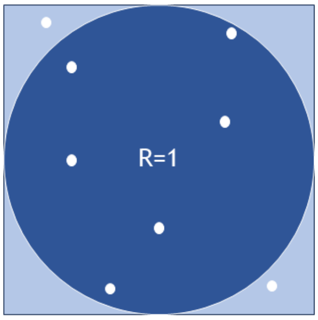
When we randomly toss darts within this boundary, the likelihood of a dart landing in the circle versus anywhere in the square guides us toward pi.

By generating a large number of random points and observing those that fall within the circle, we compute pi as four times the ratio of points inside the circle to those in the square.

This method simplifies a sophisticated mathematical concept into an accessible, data driven approach, illustrating the power of Monte Carlo simulation modeling in uncovering precise insights.

Slide #7

Large Number Theorem - Application




```
Sub PI()
Sim = 5
Dim X As Double
Dim Y As Double
Dim PICal As Double
Dim Numb As Double
Dim CirCount As Double


For I = 1 To Sim
Num = Cells(I + 2, 4)
Numb = Cells(I + 2, 4)

For Trial = 1 To 10
CirCount = 0
    For J = 1 To Num
        X = 1 - 2 * Rnd()

        Y = 1 - 2 * Rnd()
        R = X * X + Y * Y

        If R <= 1 Then
        CirCount = CirCount + 1
        End If
    Next

    PICal = (CirCount / Numb)
    Cells(I + 2, 4 + Trial) = PICal * 4

Next
Next
```

7

We've designed a pi value simulation model using Microsoft's macro function.

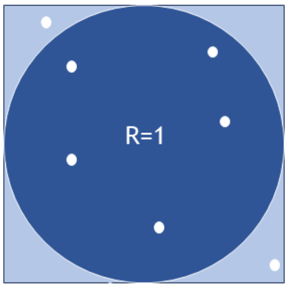
This model includes five different sets of simulations with varying dart counts, 100, 1000, 10,000, 100,000 and 1,000,000.

Highlighting how sample size affects simulation accuracy.

Conducting 10 trials for each set further underscores our method's reliability and consistency.

Slide #8

Large Number Theorem - Application



The larger the number of samples (N), the closer the calculated π to its theoretical value (3.1415926...)

Trial

Dart Number	1	2	3	4	5	6	7	8	9	10
100	3.39393939393939	2.94949494949495	3.03030303030303	3.27272727272727	3.39393939393939	3.19191919191919	3.19191919191919	3.19191919191919	3.19191919191919	3.11111111111111
1,000	3.12312312312312	3.15115115115115	3.13113113113113	3.12312312312312	3.08708708708709	3.12312312312312	3.11111111111111	3.16716716716717	3.15515515515516	3.14714714714715
10,000	3.15471547154715	3.10831083108311	3.14671467146715	3.14911491149115	3.15311531153115	3.13631363136314	3.17551755175518	3.15991599159916	3.16991699169917	3.11071107110711
100,000	3.13871387138714	3.14343434343434	3.13367336733673	3.14039403940394	3.14007400740074	3.13315331533153	3.14943494349435	3.14631463146315	3.13935393539354	3.14467446744674
1,000,000	3.14143814381438	3.14332233223322	3.14206620662066	3.14260226022602	3.14224222422242	3.13963796379638	3.14139813981398	3.13856985698570	3.14187818781878	3.14382238223822

The simulation outcomes reveal a crucial insight.

As the sample size n increases, the estimated pi value approaches its true figure 3.1415926.

This trend affirms the method's reliability and repeatability.

For decision makers, it means confidence in leveraging such simulations for real world decision making, providing a sturdy, data driven foundation for business decisions.

Slide #9

Question

- You made 10 trials, and under each, you throw 99 darts (samples).
- The estimated Pi values are shown in the below table.
- The estimated Pi is between 2.94949494949 and 3.393933939.

	Trial									
Dart Number	1	2	3	4	5	6	7	8	9	10
99	3.39393939393939	2.94949494949495	3.03030303030303	3.27272727272727	3.39393939393939	3.19191919191919	3.19191919191919	3.19191919191919	3.19191919191919	3.11111111111111

If you do NOT know the theoretical value of π, to what degree do you believe the estimated number, Pi?

or

What's your confidence level that the theoretical value of π falls within the range of Pi value estimated from 99 samples?

9

We have demonstrated the large number theorem, the principle underlying the Monte Carlo simulation.

Suppose you conduct 10 trials, each with 99 darts.

The estimated pi ranges from 2.949 to 3.393

This experiment raises a crucial question for you.

Without knowing the theoretical value of pi, to what degree do you believe the estimated number?

Or, how confident are you that the true value of pi falls within these estimates?

Slide #10

Confidence Level and Interval

A confidence interval is a range of values that is expected to contain an unknown parameter of a population.(e.g. Pi is between 3.1919191919 and 3.3332

It is based on a confidence level, which is a percentage that indicates how often the interval will capture the parameter in repeated samples.

A confidence interval is a way of expressing the uncertainty of an estimate.

e.g. $\pi \; falls \; between \; 3.1919191919 \pm 0.141253717$ @ a 68% confidence level

Dart Number	Average	STDEV
100	3.19191919191919	0.141253717

10

To determine our confidence in the estimated range of pi, we need to introduce the concept of a confidence interval from inferential statistics.

Instead of estimating an unknown parameter by a single value, a confidence interval is a range of estimated values that are expected to contain the true value of the parameter.

For instance, we might say pi lies between 3.1919 and 3.3333332 with a certain level of confidence, say 68%.

This means there is a 68% probability that the range would contain the true value pi.

The statement is based on a confidence level, which is a percentage that indicates how often the interval will capture the parameter in repeated events.

A confidence interval measures the level of certainty in our estimate.

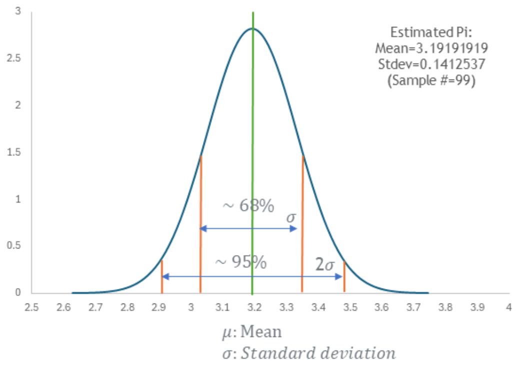
Slide #11

Empirical Rule (three-sigma rule)

~ 68% of data within one standard deviation of mean

~ 95% of data within 2 standard deviations of mean

~ 99.7% of data within 3 standard deviations of mean



11

The empirical rule provides a simple method for calculating confidence intervals for parameters in datasets that follow a normal, bell-shaped distribution around the mean.

This rule delineates that data clusters within distinct ranges falling under three key confidence intervals.

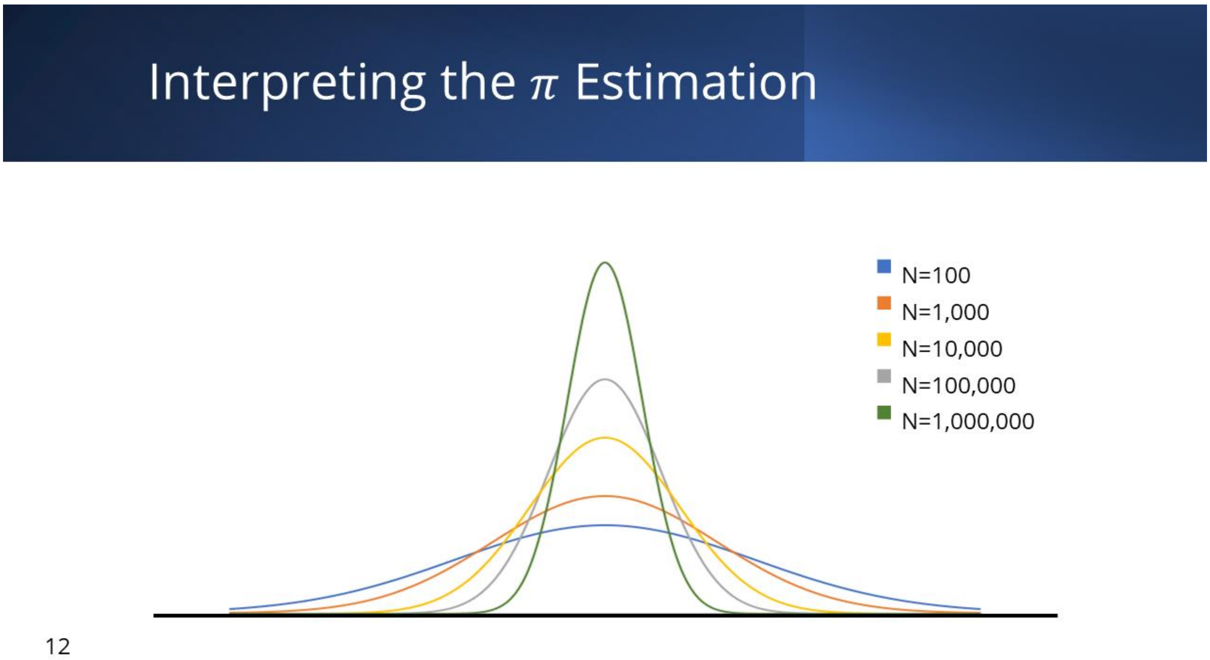
Approximately 68% of data points lie within one standard deviation, sigma of the mean, mu. This 68% confidence interval spans from mu minus sigma to mu plus sigma, representing the central range of the data.

About 95% of data points fall within two standard deviations from the mean, mu, broadly covering most outcomes. The 95% interval forms the basis of the commonly used confidence interval in scientific and statistical analysis.

Nearly 99.7% of data points are found within three standard deviations from the mean providing extensive coverage of the data set.

This range is essential for identifying outliers, the exceptional data points that lie beyond this boundary.

Slide #12



Assuming the data set of estimated pi values follows a normal distribution, we calculate the mean and standard deviation for each of the five simulations based on 10 trials results.

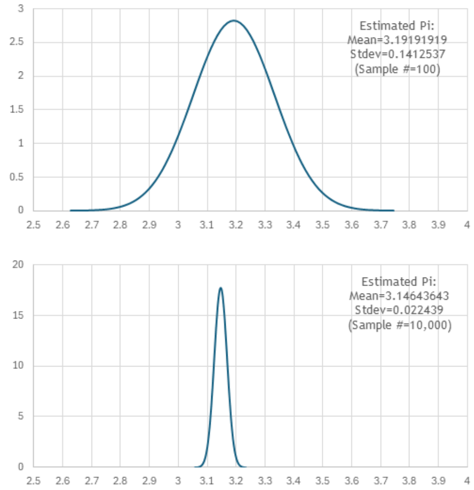
By plotting a bell curve for each data set, it becomes evident 10 that the range of pi values tightens as the sample size expands.

This trend is consistent with the Law of Large Numbers Theorem, which suggests that as the sample size grows, the sample mean converges towards the theoretical mean.

Slide #13

Confidence Intervals and # of Samples

Dart Number	Average	STDEV
100	3.19191919191919	0.141253717
1,000	3.13193193193193	0.023710556
10,000	3.14643464346435	0.022438762
100,000	3.14092209220922	0.005197982
1,000,000	3.14169776977698	0.001585344



13

For each simulation, we determine the estimated pi interval at a 68% confidence level.

The data reveals that the interval contracts with an increase in sample size.

For example, when estimating pi with 100 darts, pi falls between 3.05 and 3.33.

This range sharpens to between 3.140 and 3.143 as the dart count escalates to 1 million.

In this scenario, we can assert that pi lies between 3.140 and 3.143 with 68% confidence, closely mirroring the theoretical value of pi.

The empirical rule's significance lies in its ability to provide a straightforward method for estimating the probability, or confidence, of a variable falling within a certain range without complex calculations.

Slide #14

Confidence Intervals and # of Samples

Dart Number	Mean-Stdev	Mean+Stdev
100	3.05067	3.33317
1,000	3.10822	3.15564
10,000	3.12400	3.16887
100,000	3.13572	3.14612
1,000,000	3.14011	3.14328

The number of samples increases, the confidence interval becomes narrower at a confidence level (e.g., 68%)

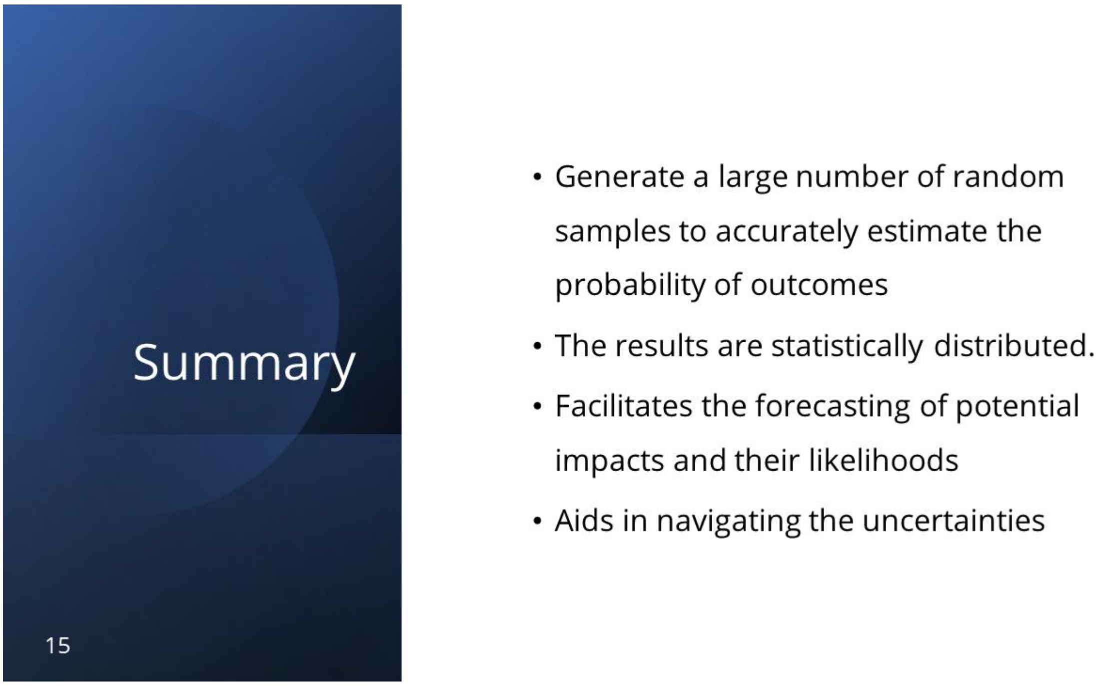
14

It simplifies the analysis of confidence intervals for the average of a normally distributed parameter, making it a practical tool for analysts who need to quickly assess the variability and reliability of their data.

This rule is most reliable with large sample sizes and when the underlying distribution of the data is known to be approximately normal.

An assumption often made in real-world simulation application.

Slide #15



Monte Carlo simulation involves generating a large number of random samples for uncertain input variables to accurately estimate the probability of outcomes in complex business models.

The results of the simulation are statistically distributed, enabling their presentation as an interval around the mean value at a specified confidence level.

As a practical tool for businesses, Monte Carlo simulation facilitates the forecasting of potential impacts and their likelihoods by simulating a wide range of uncertain business scenarios, from the best to the worst condition.

This method aids in navigating the uncertainties of business decision making with enhanced precision and communicable confidence.