Mathematics Used In My Career

Dr. Carlo Lisi

SENIOR MANGER BUSINESS INSIGHTS AND ANALYTICS

Introduction

- Name: Dr. Carlo Lisi
- Current Position: Senior Manager, Analytics and Insights
- Company: TD Bank
- PhD, 1995, Mathematics (C* Algebras and Operator Theory)
- University of Toronto, Supervisor: Professor Man Duen Choi

In my current role I focus on:

- Performing data analytics for internal audit to support control assessments
- Using Advanced analytics (i.e predictive modelling) to test controls, accuracy & completeness of data and fraud assessment
- Successful completion of the annual audit plan
- Continuous Audit initiative
- Risk Intelligence for the Global Audit Team
- Robotics Process Automation

Introduction

The main software that are used:

- ACL (Audit Command Language – audit specific)
- SAS
- SQL
- R
- Python
- VBA
- Matlab

Computer Programming Skills Required

- Experience with the tools above or equivalent data analysis tools

Introduction

Programming skills

- From my experience in the current market environment I would recommend some basic programming skills
- Microsoft Excel (VBA)
- SQL
- SAS
- R, Python, Julia, Scala, Java Script Apache Spark – other machine learning tools
- Hadoop -- Impala, Pig

Data Visualization

- Tableau, Spotfire, Qlikview or equivalent

Introduction

Workplace learning

I learned all the listed specialized software and programming languages after being employed

- Unix Shell Scripting
- ACL
- SAS
- SQL
- Python and R [machine learning packages]
- Apache Spark
- Tableau
- Monarch

Introduction

- Mathematics used in Finance
- Pricing Derivatives and Exotic Options
- Extreme Value Theory
- Machine Learning

Introduction

Recommendation for your future professional development

- Network with individuals currently engaged in industry and ask for their recommendation
- Continue learning programming languages that are used in data analytics & machine learning

MATHEMATICS USED IN MY CAREER

MARKET RISK

Market Risk

TYPES OF MARKET RISK

- Equity risk: the risk that stock prices will change.
- Interest rate risk: the risk that interest rates will change.
- Currency risk: the risk that foreign exchange rates will change.
- Commodity risk: the risk that commodity prices (e.g. corn, copper, crude oil) will change.

Market Risk

- My first job at a Financial Institution (Royal Bank of Canada) was Senior Analyst, FX products on the market risk side
- Responsible for monitoring Market Risk for FX and commodities in the trading books
- Produced VaR (Value at Risk), and the Greeks, i.e. Delta, Gamma, Vega, Rho, and Theta reports
- **Math Used:**
- **Deterministic calculus** – single variable and multi-variable was used to check the calculations done by the Vendor Software (yes, first derivative, second derivatives are used!!)
- **Stochastic Calculus** – for pricing more exotic products (i.e. partial barrier options)
- Black-Scholes Option Pricing Formula
- Monte-carlo simulation

Market Risk

- Stochastic calculus

Itô's formula

If X is a stochastic process, satisfying $dX_t = \sigma_t\, dW_t + \mu_t\, dt$, and f is a deterministic twice continuously differentiable function, then $Y_t := f(X_t)$ is also a stochastic process and is given by

$$dY_t = \left(\sigma_t f'(X_t)\right) dW_t + \left(\mu_t f'(X_t) + \tfrac{1}{2}\sigma_t^2 f''(X_t)\right) dt.$$

Returning to our W_t^2, we can apply Itô with $X = W$ and $f(x) = x^2$ and we have

$$d(W_t^2) = 2W_t\, dW_t + dt, \quad \text{or} \quad W_t^2 = 2\int_0^t W_s\, dW_s + t,$$

which at least has the right expectation.

More generally, if X is still just the Brownian motion W, then $f(X)$ has differential

$$df(W_t) = f'(W_t)\, dW_t + \tfrac{1}{2} f''(W_t)\, dt,$$

Market Risk

Black-Scholes Formula

- current underlying price
- options strike price
- time until expiration, expressed as a percent of a year
- implied volatility
- risk-free interest rates

$$C = SN(d_1) - N(d_2)Ke^{-rt}$$

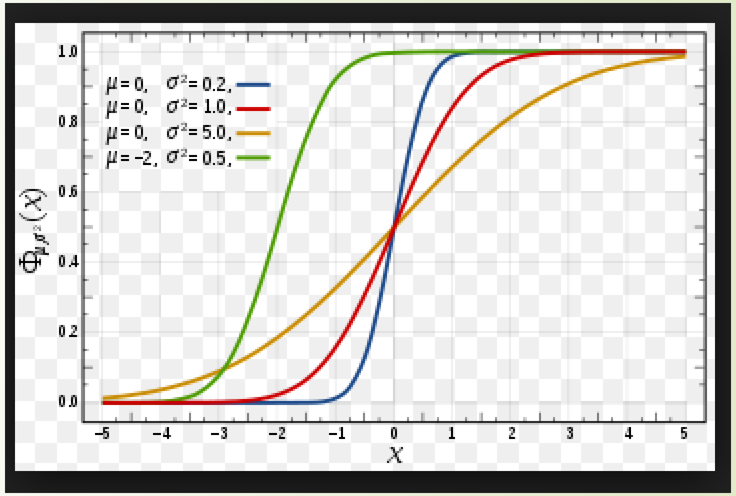
$$d_1 = \frac{\ln\left(S/K\right) + \left(r + s^2/2\right)t}{s \cdot \sqrt{t}}$$

$$d_2 = d_1 - s \cdot \sqrt{t}$$

C = Call premium
S = Current stock price
t = Time until option exercise
K = Option striking price
r = Risk-free interest rate
N = Cumulative standard normal distribution
e = Exponential term

s = St. Deviation
ln = Natural Log

Normal Distribution



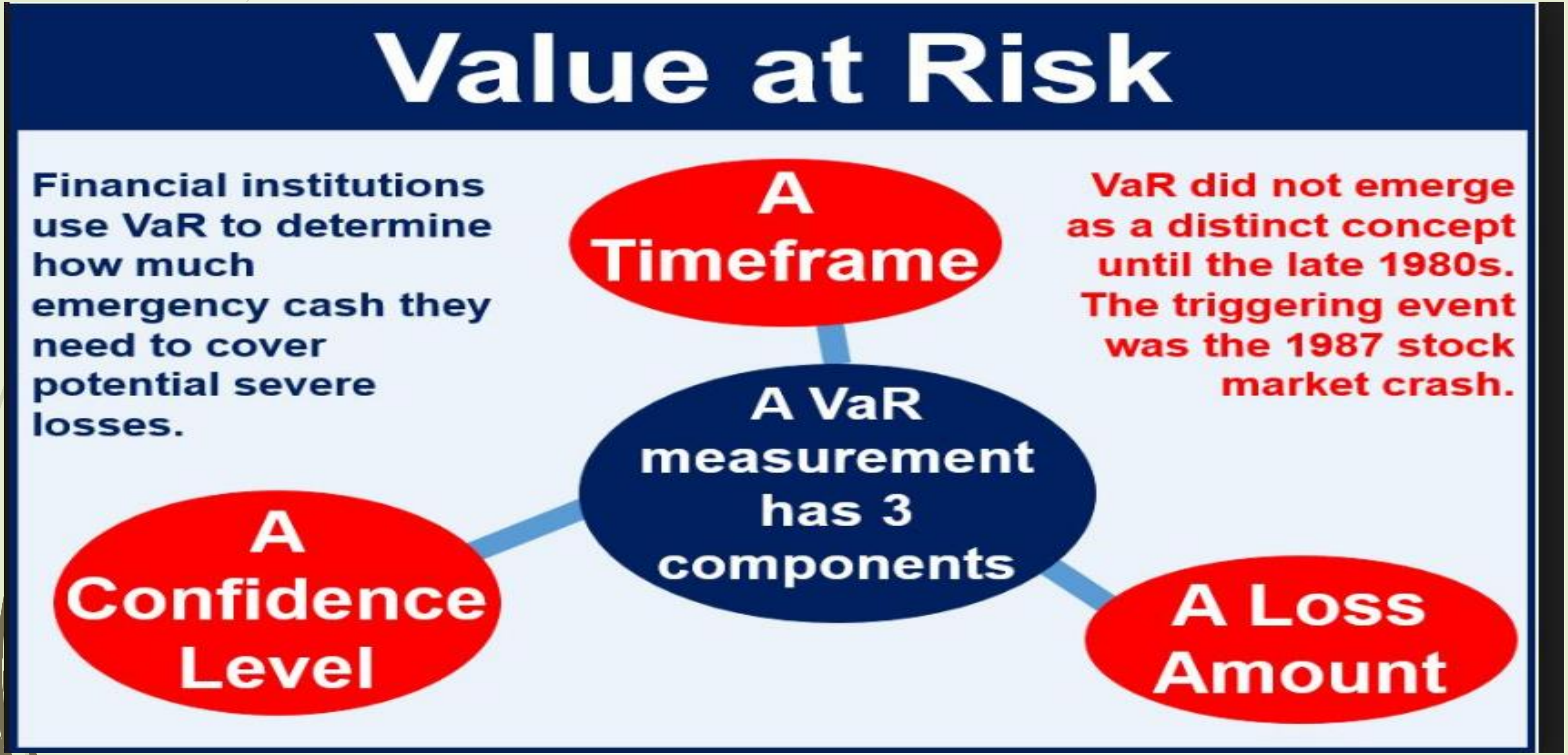
Greeks

- **DELTA**
- Change in price of the option with respect to the spot price
- To derive this, you need to remember the chain rule, and partial derivatives
- Learned financial mathematics from John Hull's book, "Options, Futures and other Derivatives"
- Gamma, Vega, Rho, Theta are all derived in John Hull's
- books

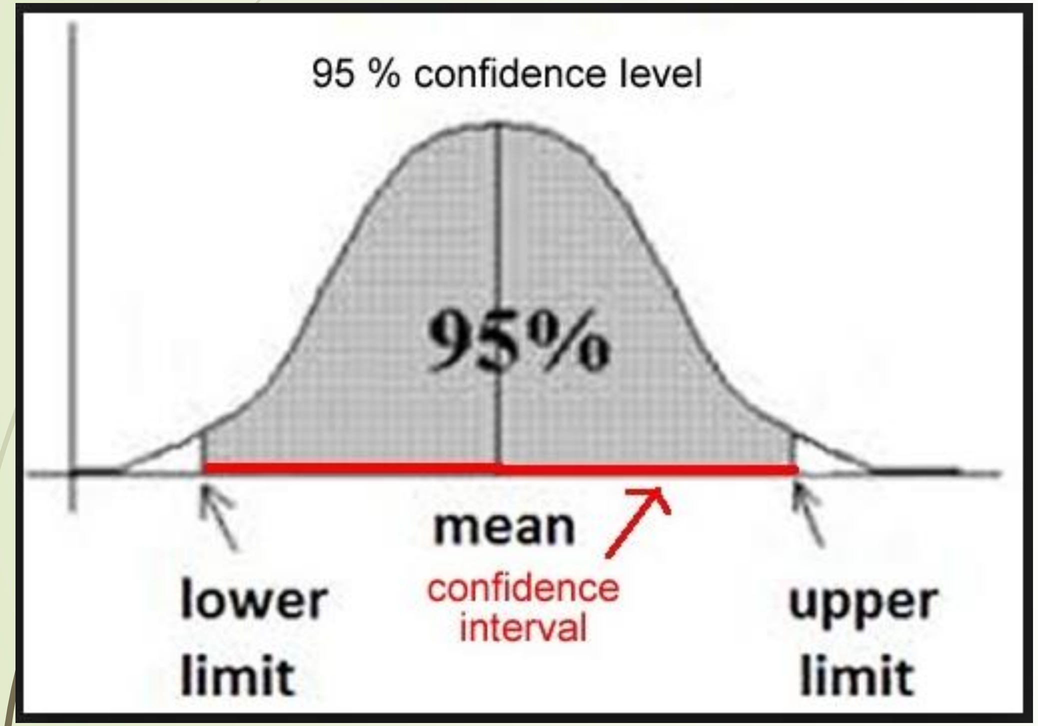
Greeks

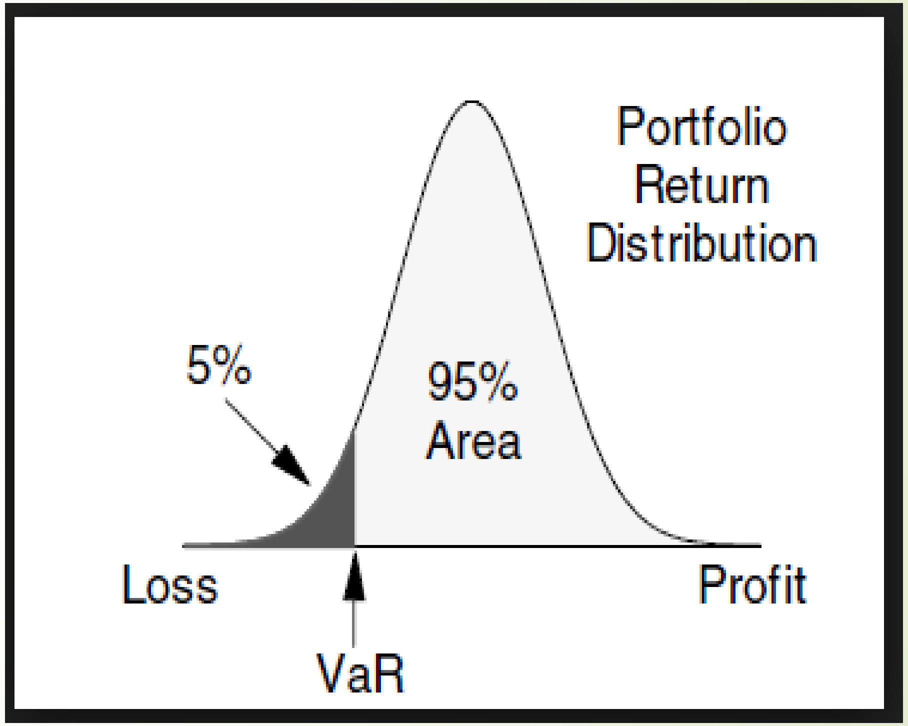
- The formulas in John Hull's book or any other academic books all apply in the Capital Markets world
- I've seen John Hull's book on the desk of traders, quants and middle office personnel
- **Software/Programming:**
- Visual Basic For Applications (VBA) mostly used with MS Excel
- Linux shell scripting
- Awk , Perl, Python (for parsing data out of scenario files)
- OLAP Cubes, extracting data with MDX (Multi-dimensional expressions)
- C/C++

Value at Risk

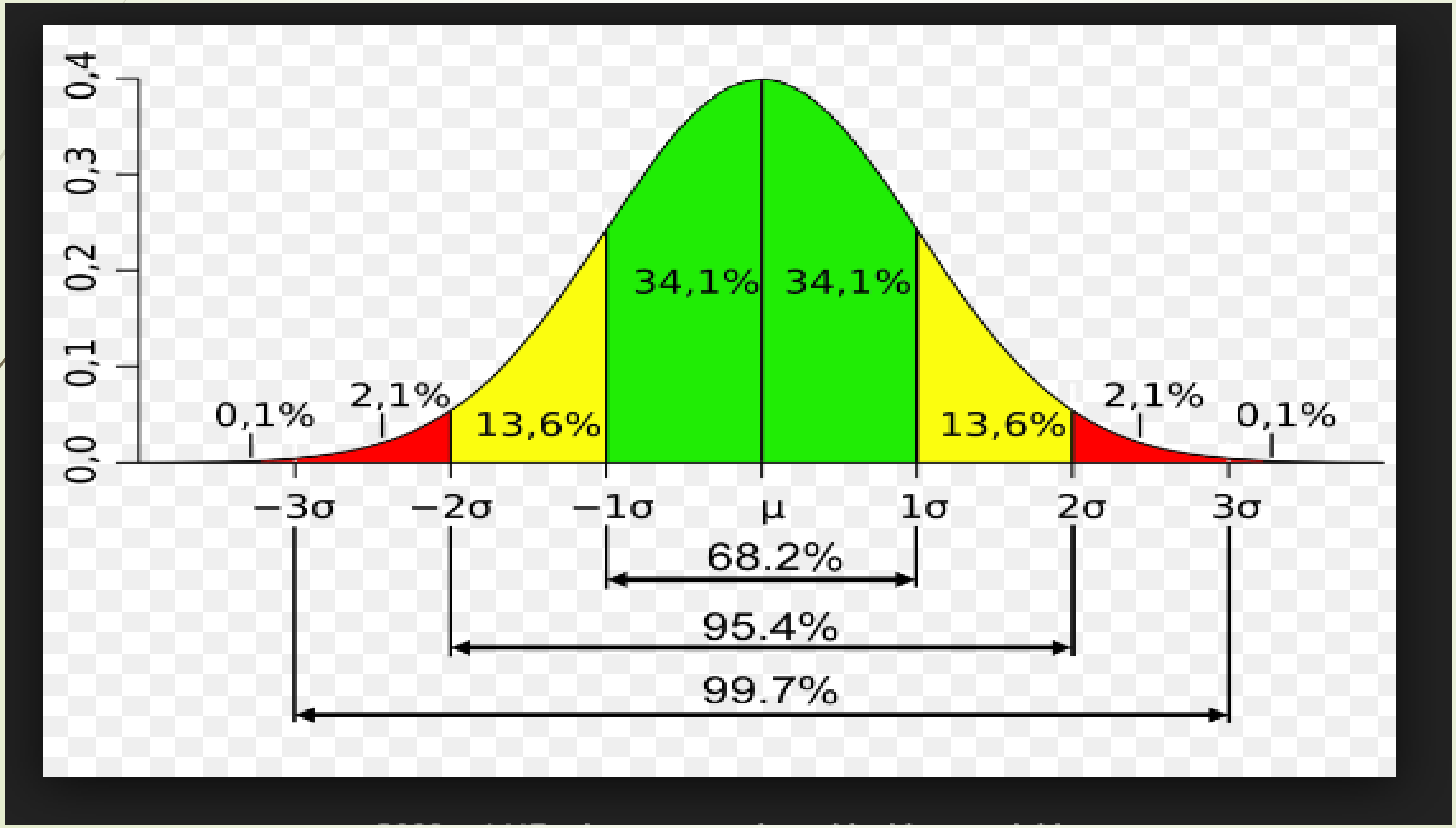


Value at Risk

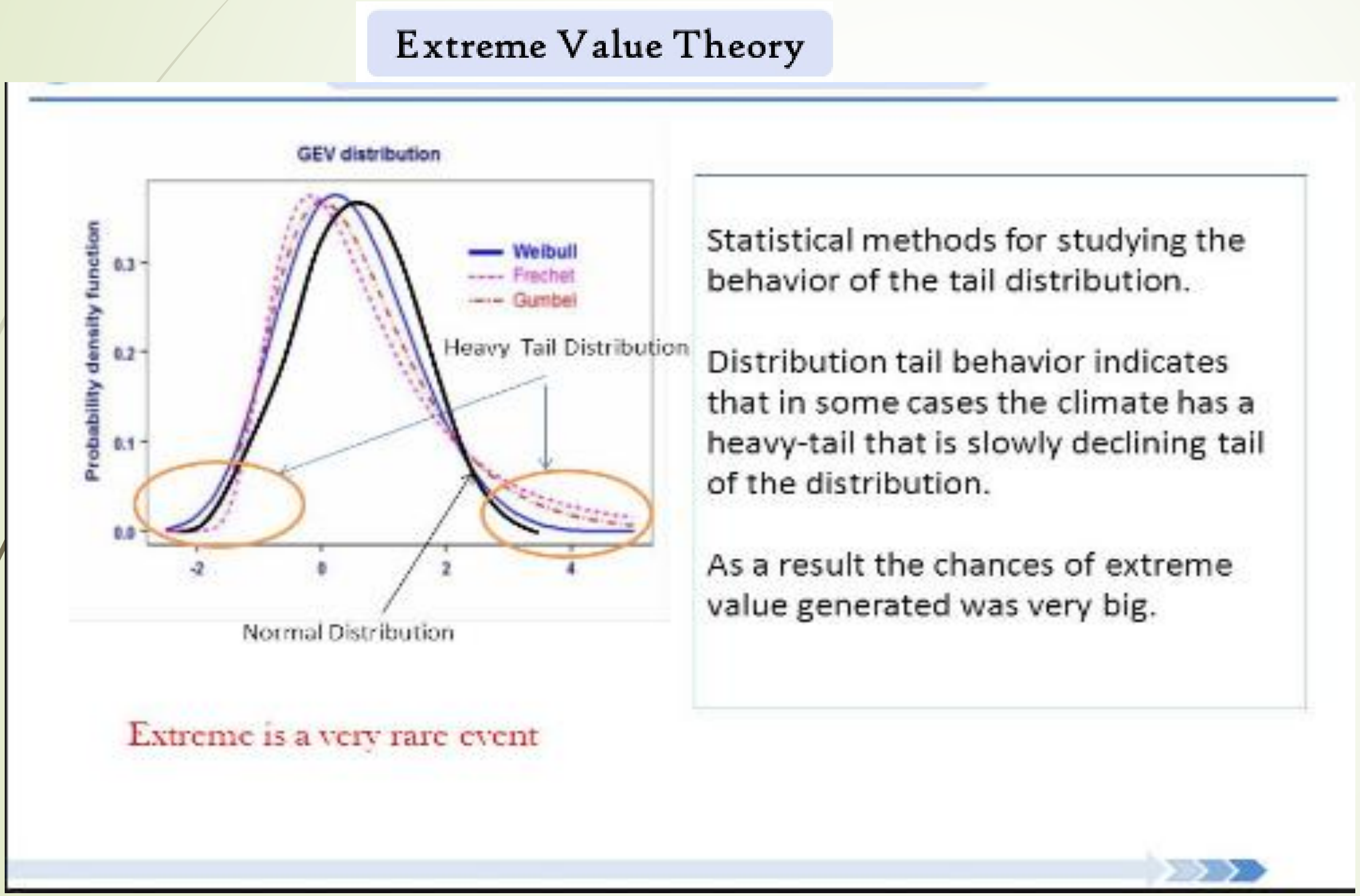




Normal Distribution (aka, the "Bell Curve")



Extreme Value Theory



Extreme Value Theory

- I used Extreme Value theory to predict the movements in CAD/USD, EUR/USD, JPY/USD exchange rates that could happen over a three-day period once every 15 years
- Would use the result to calculate losses in the trading book, if such moves happened
- Traders never agreed this could happen, however they had to adhere to the stress limits generated from the EVT scenarios
- Code was written in Matlab

MATHEMATICS USED IN MY CAREER

INTERNAL AUDIT

INTERNAL AUDIT

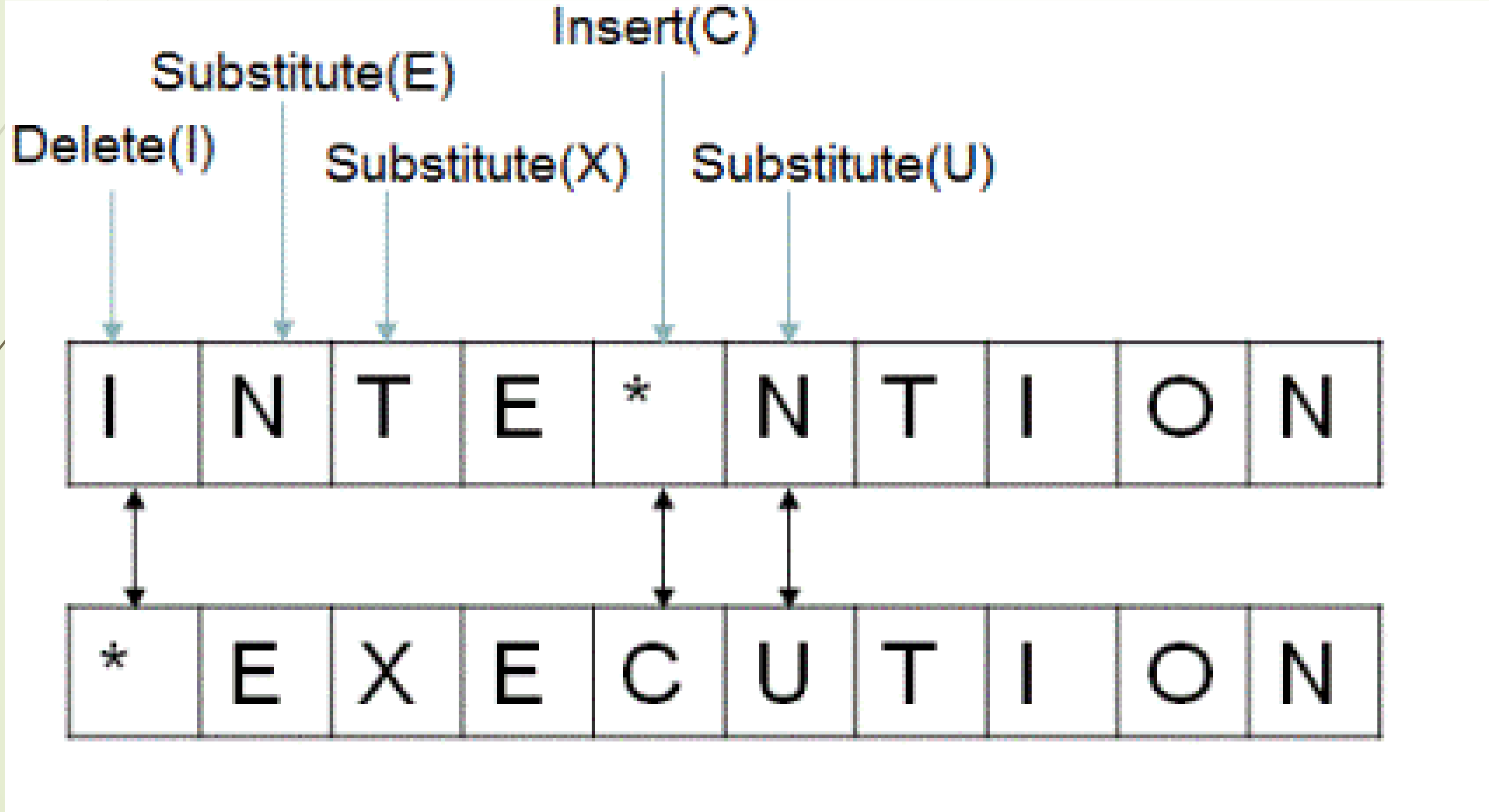
This is where I have used math in my analytics work while working in Internal Audit

- Predictive Models , e.g. Scoring of Vendor Invoice and Travel and Entertainment Expenses
- Statistical analysis of Risk Measures (Value at Risk, P&L etc…)
- Benford's law & other metrics(NFF, RSF, Levenshtein distance, Fuzzy Matching) to look for fraud / duplication
- Sentiment Analysis (used python and R packages – explored the math being used)
- Text Analytics (word frequencies, correlations, classification)

LEVENSHTEIN DISTANCE (edit distance)

- Minimum Edit distance between two strings str1 and str2 is defined as the minimum number of insert/delete/substitute operations required to transform str1 into str2
- For example if str1 = "ab", str2 = "abc" then making an insert operation of character 'c' on str1 transforms str1 into str2
- Therefore, edit distance between str1 and str2 is 1
- You can also calculate edit distance as number of operations required to transform str2 into str1
- For above example, if we perform a delete operation of character 'c' on str2, it is transformed into str1 resulting in same edit distance of 1
- Looking at another example, if str1 = "INTENTION" and str2 = "EXECUTION", then the minimum edit distance between str1 and str2 turns out to be 5 as shown below
- All operations are performed on str1

LEVENSHTEIN DISTANCE(cont)



Fuzzy Matching

- Fuzzy matching – approximate matching

Company Names
Tata
tata
Tataz
TataM Jan2015
Tata Motor
Reliance World
Reliance Ltd
Reliance #Petro
Reliance Global
Vanucoverltd 12 Company

Company Names
Tata
Tata Motors
Reliance
Vanucover

Record #	Name
A	Hewlett Packard
B	HP, Inc.
C	Hewlett-Packard Corp
D	HP Corporation
E	Hewlet Pakkard
F	Hewwlet Packerd

Hewlett Packard

Fuzzy Matching

- Similar_String = 2xLongestCommonSubsequence(String1, String2)
- Length(String1) + Length(String2)

Name 1	Name 2	%Match
John Smith	Johnny Smith Inc	85.7%
P.O. BOX 201	PO BOX 201	87.0%
55 King Street West	55 King Street W.	86.5%

- **Exercise: What is the Simil_String similarity between Alpha and AlphaGo?**

Number Frequency Factor

- Number Frequency Factor is a measure of the level of duplication in a set of Numbers
- Formula for NFF = sum(ci^2)/n^2 , i = 1,2..,n
- Where ci = number of times a value is repeated, n is the total number of values in the set, ci =0 if a value appears only once
- For example, calculate NFF for the set of values below:
- S = {1,1,1,4,5,8,8}
- Distinct values are 1,4,5,8
- 1 is repeated three times so c1 is 3^2 (i.e. 3*3)
- 8 is repeated twice so c8 is 2^2 (i.e. 2*2)

Number Frequency Factor

- 4 appears only once so c4 = 0
- 5 appears only once so c5 = 0
- n =7 is the number of values in S
- NFF = (3^2+2^2)/7^2 = 13/49 which is approximately 0.265
- If S = { 1,1,1,1,1,1,1} , NFF = 7^2/7^2 = 1 (all duplicates)
- If S = {1,2,3,4,5,6,7} Number , NFF = 0^2/7^2 = 0 (all distinct)

Relative Size Factor(RSF)

- The relative size factor (RSF) test is an important error-detecting test."
- Largest Record in a Subset
- Relative Size Factor = --
- Second Largest Record in a Subset
- If the largest vendor invoice is a large multiple of the second largest, we would investigate

BENFORD's LAW

Benford's Law

- Benford's law is sometimes referred to as digital/frequency analysis
- It's reasonable to expect that the first digit of any value in an dataset (i.e. Vendor Invoices) value to be random
- That is , there is an equal chance of the first digit being an number between 1 and 9
- But that isn't necessarily the case
- Although it's counterintuitive, some numbers appear more frequently than others in many datasets
- In fact the digit 1is the leading digit 30% of the time, 2 is the leading digit 18% of the time, 3 is the leading digit 12% of the time and larger numbers decreasingly so

Benford's Law

- **History of Benford's Law**
- In 1881, Simon Newcomb, an astronomer and mathematician, noticed something peculiar
- While looking through much-used log tables at the library, he found that earlier pages were more worn than later pages
- He concluded that his fellow scientists looked up numbers beginning with the digit **one** more often than numbers beginning with digit **two** and so on
- He concluded that the probability distribution of the first digit was
- P(d) = log10(1+1/d)
- Example: P(d=1) = log10(1+1/1)= log10(2) [get out your calculators]= 0.301029996

Benford's law

- **Who is Benford ?**
- In 1938, physicist Frank Benford, who was unaware of Newcomb's observation, also discovered same phenomena with his Logarithm Book used by scientists & engineers
- Unlike Newcomb, Benford attempted to test his theory with empirical data
- Frank Benford analyzed 20,229 data sets by hand(no MS Excel back then)

Benford's law

- **Examples:**
- Baseball statistics
- Areas of rivers
- Molecular weights of atoms
- Electricity bills
- Stock market quotes
- Populations of towns
- Physical and mathematical constants
- He discovered that appearance of each digits (1 – 9) is not equally distributed, instead some digits appear more frequently than others

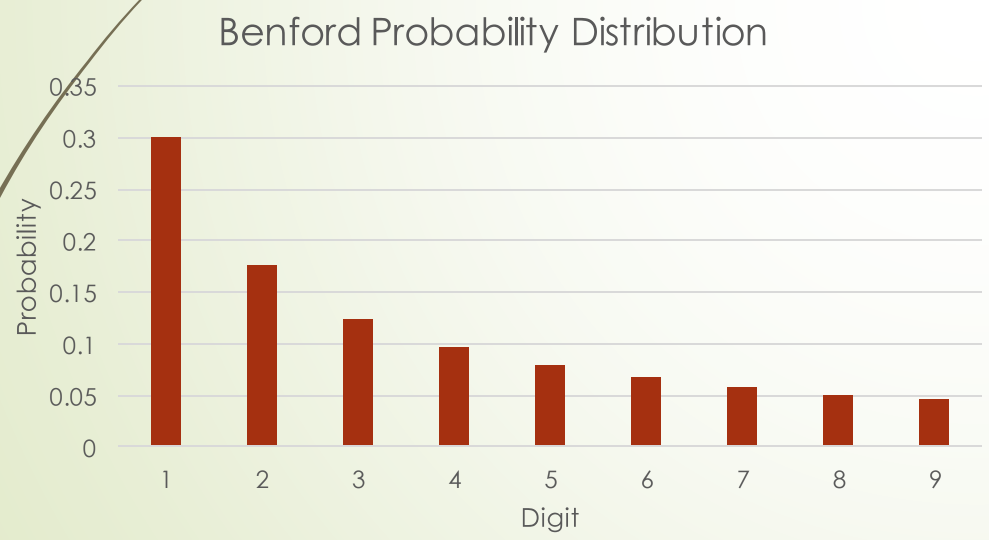
Benford's law

- **How is Benford's Law possible?**
- If a data entry begins with the digit 1, it has to double in size (100%) before it begins with the next digit – digit 2
- If a data entry begins with the digit 9, it only has to be increased by only 11% in order for the first digit to be digit 1 again
- Hence, chances of digit 1 is more likely than digit 9
- You will have more smaller numbers than larger numbers
- Check out the example in attached Excel worksheet

Benford’s Law

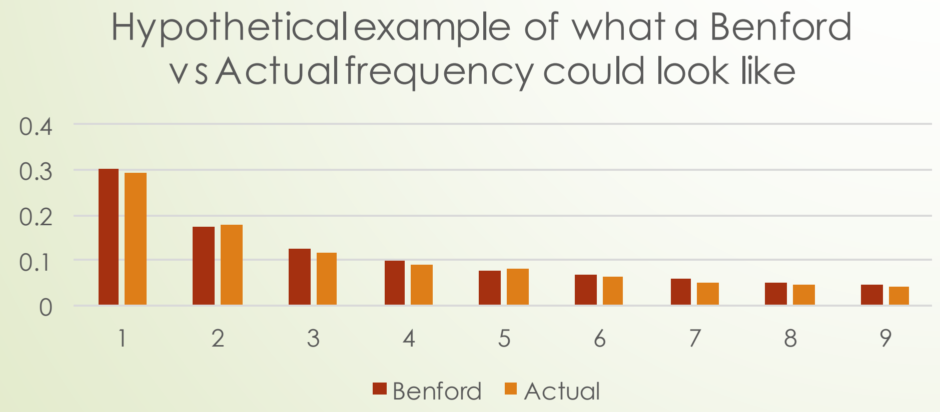
- A set of numbers is said to satisfy Benford's law if the leading digit d (d ∈ {1, ..., 9}) occurs with probability

$$P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}\left(\frac{d+1}{d}\right) = \log_{10}\left(1 + \frac{1}{d}\right)$$



Benford's Law

- Many different types of data sets follow this rule, including the
  - Number of twitter users by followers
  - Populations of cities across the USA
  - Heights of buildings
  - File sizes on your hard drive
- Auditors/Forensic accountants often use Benford's law to detect fraud
- When people "massage numbers" in an attempt to defraud, they tend to use more 8s and 9s as the first digit than expected from Benford's law



Benford's Law

State of Arizona vs Wayne James Nelson

- Invented or altered numbers are not likely to follow Benford's law
- Human choices are not random
- Let's look at an actual legal case that happened in 1992
- State of Arizona vs Wayne James Nelson
- In 1993 , Wayne James Nelson was accused of trying to defraud the state of Arizona of two million dollars (USD)
- Nelson, wrote 23 cheques to a fictitious vendor in seemingly random amounts
- But his plan had a major flaw, his amounts weren't random enough
- In the trial the defendant was accused of issuing cheques to a vendor that did not exist

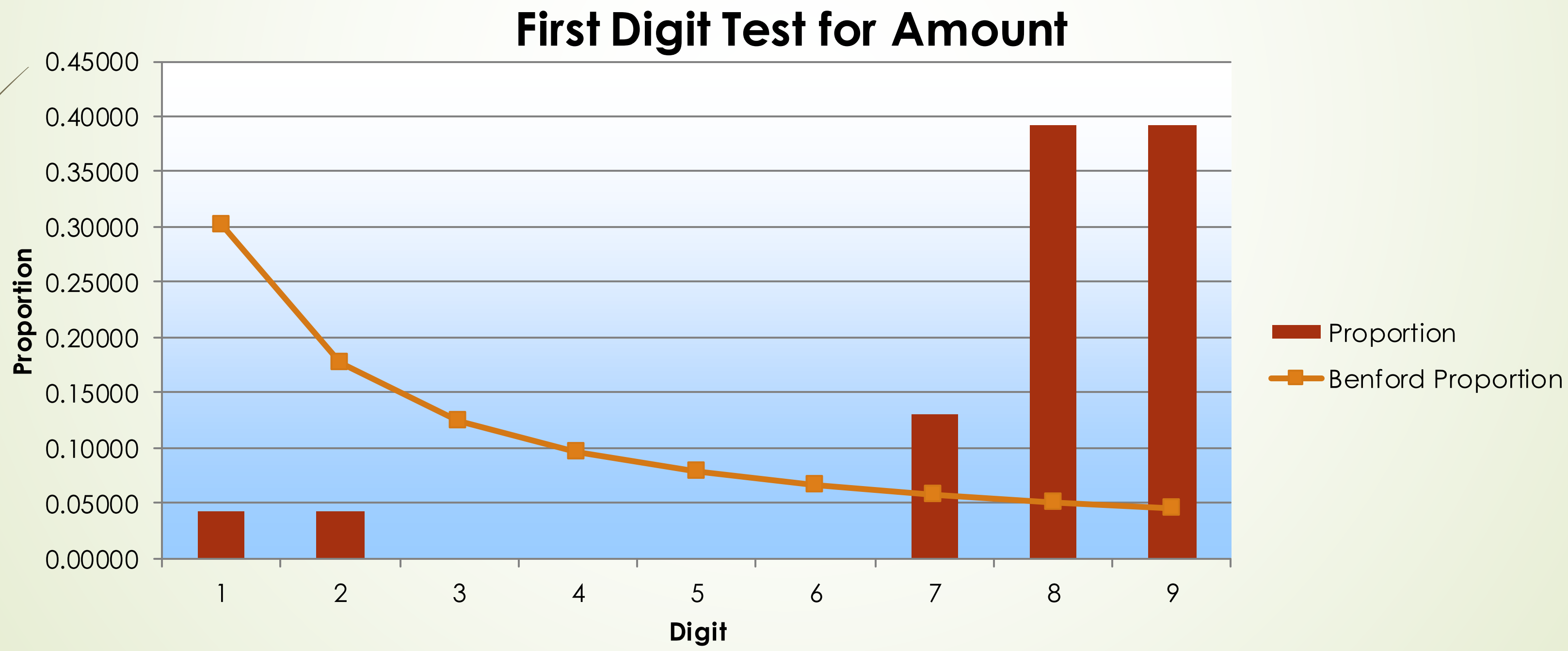
Benford's Law

- Here are the 23 cheque amounts:

Date	Amount
1992-10-09	1927.48
1992-10-09	27902.31
1992-10-14	86241.9
1992-10-14	72117.46
1992-10-14	81321.75
1992-10-14	97473.96
1992-10-19	93249.11
1992-10-19	89658.17
1992-10-19	87776.89
1992-10-19	92105.83
1992-10-19	79949.16
1992-10-19	87602.93
1992-10-19	96879.27
1992-10-19	91806.47
1992-10-19	84991.67
1992-10-19	90831.83
1992-10-19	93766.67
1992-10-19	88388.72
1992-10-19	94639.49
1992-10-19	83709.28
1992-10-19	96412.21
1992-10-19	88432.86
1992-10-19	71552.16

Benford's law

- First Digit Analysis

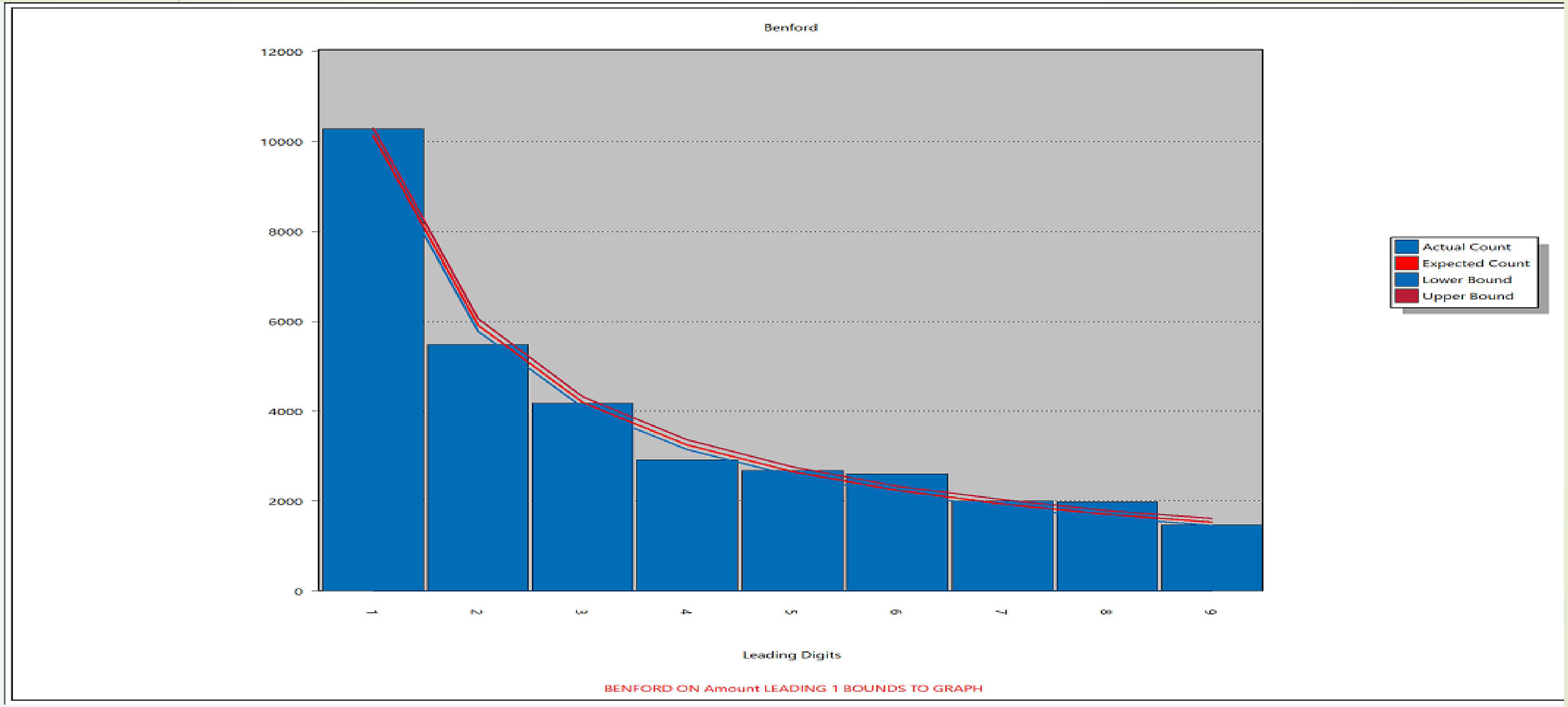


Benford's law

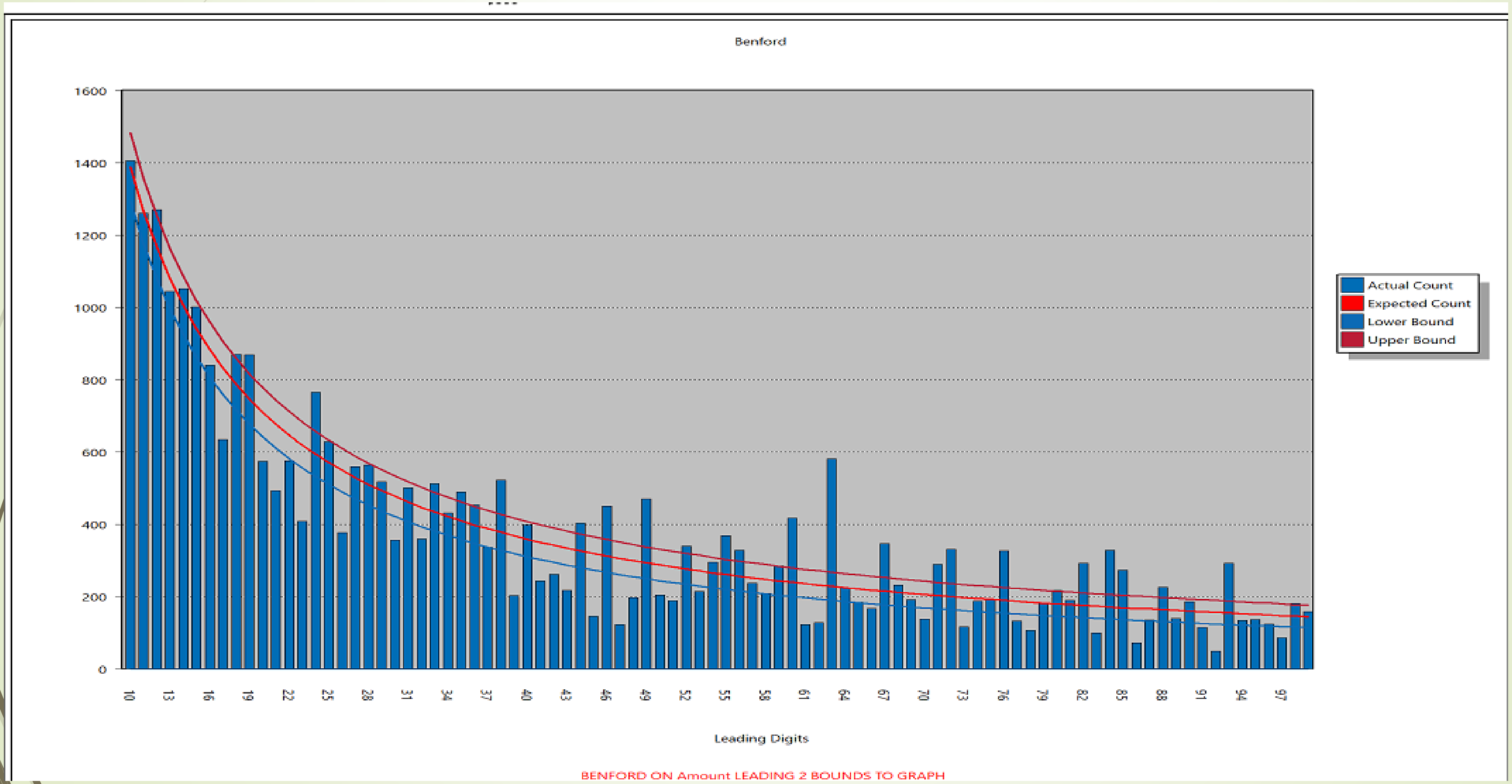
- You can clearly see that the fictitious cheque amounts do not conform to Benford's law
- In the next session, we will do the same frequency with the Zetaphor corporation's (fictitious corporation) invoice data

Example -- Vendor Invoices 1st digit

- Purchase card transactions

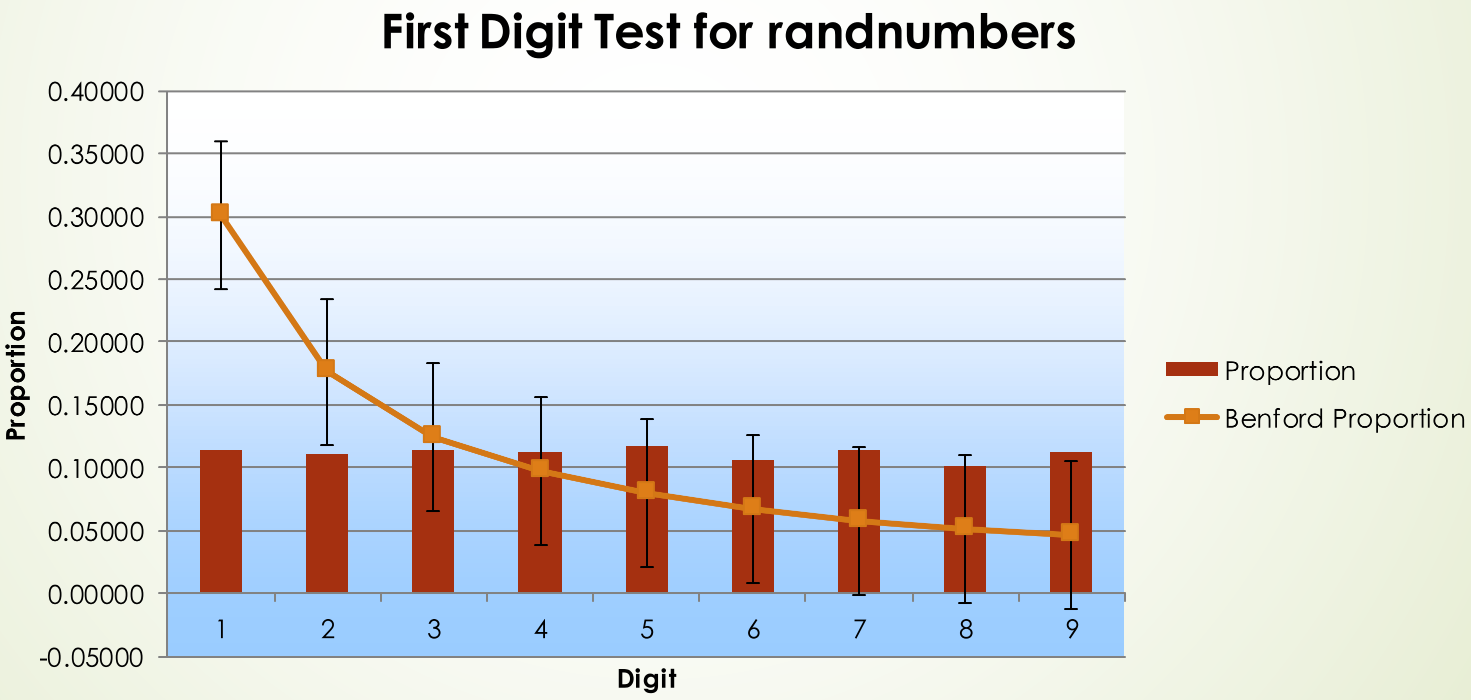


Example – Vendor Invoices first two digits

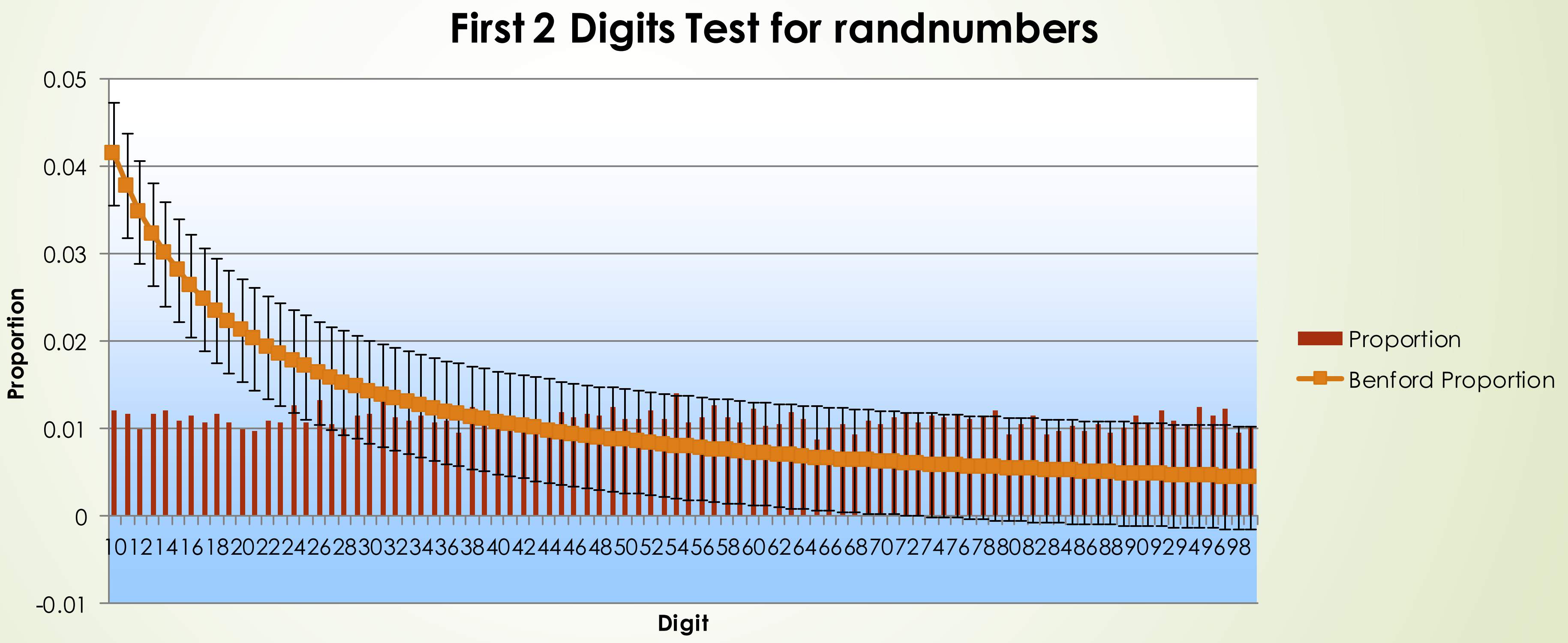


Benford and Random numbers

- Would Benford apply to random numbers?
- I generated 10000 random numbers using Excel and here is the plot for the first digit test and first two digits test

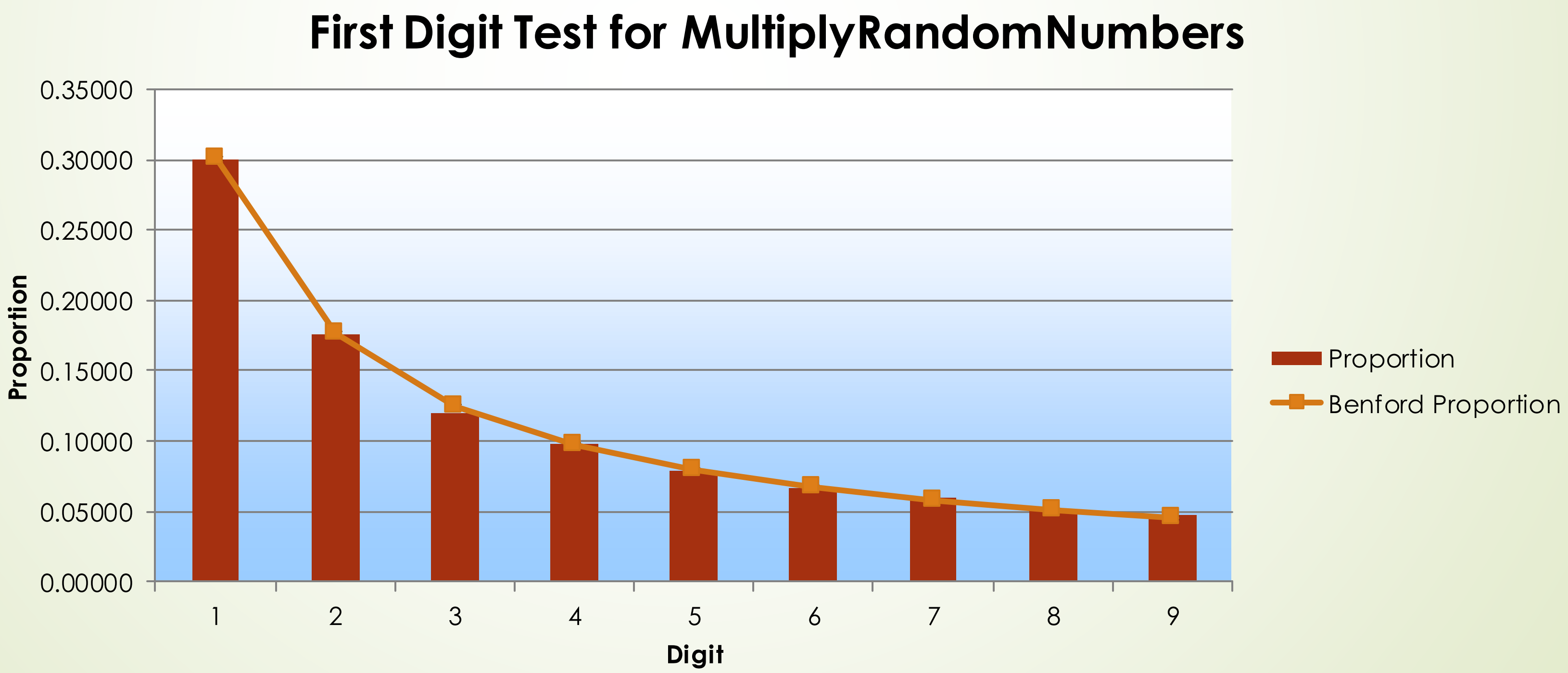


Benford and Random Numbers

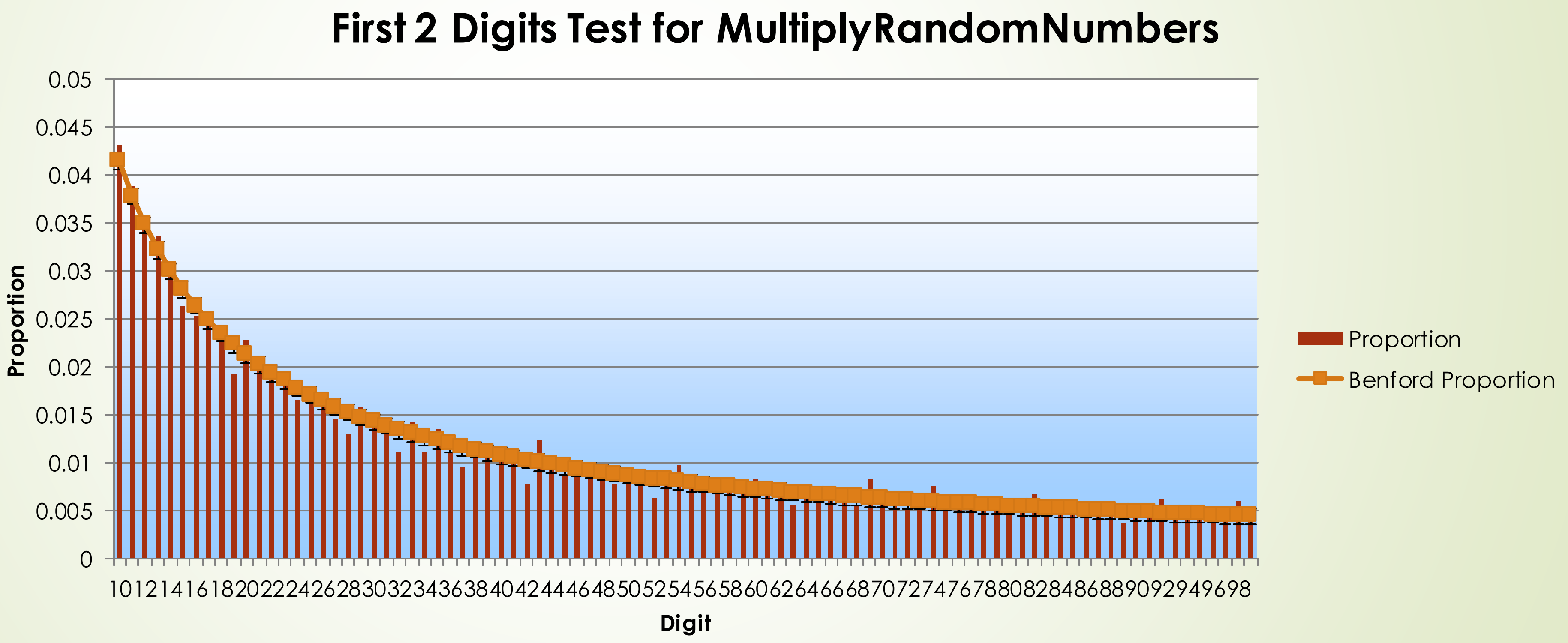


Benford and Random Numbers

- You can see that it doesn't follow Benford's law
- What I'll do now is generate 6 sets of 10000 numbers and multiply them together to get one set of random numbers
- Let's do a Benford Analysis on the multiplied random numbers



Benford and Random Numbers



Benford and Random Numbers

- What can you observe from this?
- It’s that Benford is a “limiting distribution” of digit frequencies

Benford and Random Numbers