Forensic Accounting

Daryl DeFord

Dartmouth College
Department of Mathematics

Johns Hopkins University - Center for Talented Youth
Science and Technology Series
National Mathematics Awareness Month

Abstract

Abstract

Mathematics is frequently referred to as the study of patterns and relationships. Forensic accountants are mathematicians who study the patterns that appear in complex financial, business, and political transactions in order to discover whether the people involved are reporting accurate information. One of the most important concepts in this field is an understanding of mathematical randomness because human beings are particularly poor at constructing truly random data. In this workshop, we will discuss several of the ways that mathematicians are able to detect non–random behavior in large data sets and see how these methods can be applied to real world settings like tax audits and voter fraud.

Outline

1. Introduction
2. Probability and Randomness
3. Benford's Law
4. Applications
5. Conclusion

Forensic Accounting

Definition (Forensic Accountant)

Accountants and auditors that use mathematics to investigate fraud, analyze data, and help prevent financial crimes.

Necessary skills:

- Probability and Statistics
- Computer Programming
- Data Analysis
- Accounting/Auditing
- Communication Skills
- ...

Forensic Accounting

Definition (Forensic Accountant)

Accountants and auditors that use mathematics to investigate fraud, analyze data, and help prevent financial crimes.

Necessary skills:

- Probability and Statistics
- Computer Programming
- Data Analysis
- Accounting/Auditing
- Communication Skills
- ...

Forensic Accounting

Definition (Forensic Accountant)

Accountants and auditors that use mathematics to investigate fraud, analyze data, and help prevent financial crimes.

Necessary skills:

- Probability and Statistics
- Computer Programming
- Data Analysis
- Accounting/Auditing
- Communication Skills
- ...

Forensic

Figure : Exciting! The word forensic actually refers to evidence that is admissible in a legal court.

Accounting

Figure : Accountants help manage, collect, and report financial information to help investors and businesses.

Forensic Accounting: Career Path

- Bachelor's degree in accounting, actuarial science, or another related mathematical field
- Professional Certifications:
  - Certified Fraud Examiner (CFE)
  - Certified Public Accountant (CPA)
  - Certified in Financial Forensics (CFF)
  - Certified Internal Auditor (CIA)
  - ...
- Actuarial Exams
- Unemployment rate $\sim 2\%$
- Rapidly growing field
- Median Salary $\sim \$100,000$

Forensic Accounting: Career Path

- Bachelor's degree in accounting, actuarial science, or another related mathematical field
- Professional Certifications:
  - Certified Fraud Examiner (CFE)
  - Certified Public Accountant (CPA)
  - Certified in Financial Forensics (CFF)
  - Certified Internal Auditor (CIA)
  - ...
- Actuarial Exams

- Unemployment rate $\sim 2\%$
- Rapidly growing field
- Median Salary $\sim \$100,000$

Forensic Accounting: Career Path

- Bachelor's degree in accounting, actuarial science, or another related mathematical field
- Professional Certifications:
  - Certified Fraud Examiner (CFE)
  - Certified Public Accountant (CPA)
  - Certified in Financial Forensics (CFF)
  - Certified Internal Auditor (CIA)
  - ...
- Actuarial Exams

- Unemployment rate $\sim 2\%$
- Rapidly growing field
- Median Salary $\sim \$100,000$

Why Forensic Accounting?

- Large Scale Frauds
  - Enron
  - WorldCom
  - AIG
  - Tyco
  - ...
- Big Data
  - More storage
  - Greater processing power
  - Better algorithms
- Increased Regulatory Attention
  - Sarbanes–Oxley Act of 2002
  - Security and Exchanges Commission

Dartmouth

Why Forensic Accounting?

- Large Scale Frauds
  - Enron
  - WorldCom
  - AIG
  - Tyco
  - ...
- Big Data
  - More storage
  - Greater processing power
  - Better algorithms
- Increased Regulatory Attention
  - Sarbanes–Oxley Act of 2002
  - Security and Exchanges Commission

Why Forensic Accounting?

- Large Scale Frauds
  - Enron
  - WorldCom
  - AIG
  - Tyco
  - ...
- Big Data
  - More storage
  - Greater processing power
  - Better algorithms
- Increased Regulatory Attention
  - Sarbanes–Oxley Act of 2002
  - Security and Exchanges Commission

How to do Forensic Accounting

- Basic accounting techniques
- Probability and Statistics
- Computer Programming and Data Analysis
- **Digit Distributions**

Probability Basics

Definition (Probability)

The probability of an event is a description of how likely that event is to occur.

Kolmogorov Axioms:

1. The probability of any particular event is non-negative.
2. The probability of the collection of all possible events is 1.
3. The probability of two mutually exclusive sets of events is equal to the sum of their probabilities.

Probability Basics

Definition (Probability)

The probability of an event is a description of how likely that event is to occur.

Kolmogorov Axioms:

1. The probability of any particular event is non-negative.
2. The probability of the collection of all possible events is 1.
3. The probability of two mutually exclusive sets of events is equal to the sum of their probabilities.

Axiom Example: Coin Flipping

Imagine that we flip an unweighted coin. There are two possible events: either it lands heads up or tails up.

1. The probability of any particular event is non-negative.
    - Heads occurs with probability $\frac{1}{2} > 0$.
    - Tails occurs with probability $\frac{1}{2} > 0$.
2. The probability of the collection of all possible events is 1.
    - The coin either lands heads up or tails up.
3. The probability of two mutually exclusive sets of events is equal to the sum of their probabilities.
    - $\frac{1}{2} + \frac{1}{2} = 1$.

Notation

- We will denote the probability that a particular event occurs by $P(\text{event})$.
- For example:
  $P(\text{Heads}) = \frac{1}{2}$
- For example:
  $P(\text{Roll a 4 on a 6 sided die}) = \frac{1}{6}$.
- For example:
  $P(\text{A spade is on top of a standard 52 card deck}) = \frac{1}{4}$.

Independent Events (and)

- Two events are independent if they don't depend on each other.
- For example, if I roll a die and flip a coin the result of the roll doesn't depend on the flip and the result of the flip doesn't depend on the roll.
- When two events are independent, the probability that they both occur is equal to the product of their individual probabilities:

$$P(\text{A and B}) = P(A) \cdot P(B)$$

- Example:

$$P(\text{Roll a 4 and flip heads}) = P(\text{Roll a 4}) \cdot P(\text{Heads}) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$$

- Product Rule

Mutually Exclusive Events (or)

- Sum Rule
- For mutually exclusive events A and B, the probability that A or B occurs is equal to the sum of their probabilities.

$$P(\text{A or B}) = P(A) + P(B)$$

- Example:

$$P(\text{Draw a 5 or a face card}) = P(5) + P(\text{Face Card}) = \frac{1}{13} + \frac{4}{13} = \frac{5}{13}$$

- These are mutually exclusive because a card cannot be both a 5 and a face card.

Dartmouth

Non–Exclusive Events (or)

- Non-Exclusive Sum Rule
- When the events are not mutually exclusive we must subtract off the joint probability to avoid over counting

$$P(\text{A or B}) = P(A) + P(B) - P(\text{A and B})$$

- This is the same as the previous example when the events are exclusive because then $P(\text{A and B}) = 0$.
- Example:

$$P(\text{Draw a 5 or a spade}) = P(5) + P(\text{Spade}) - P(\text{5 and Spade})$$

$$= \frac{1}{13} + \frac{1}{4} - \frac{1}{52} = \frac{4}{13}$$

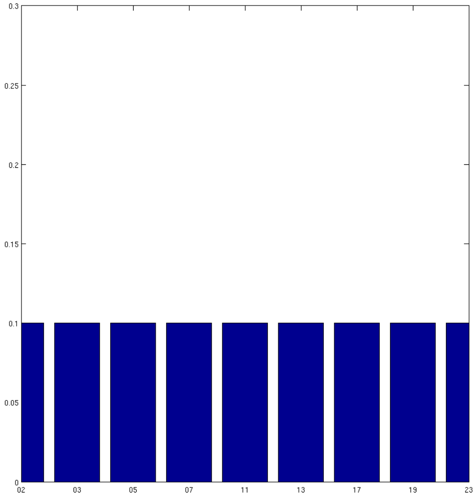
Probability Practice

Try the problems on the Discrete Probability Practice Sheet. Feel free to work with other students sitting nearby.

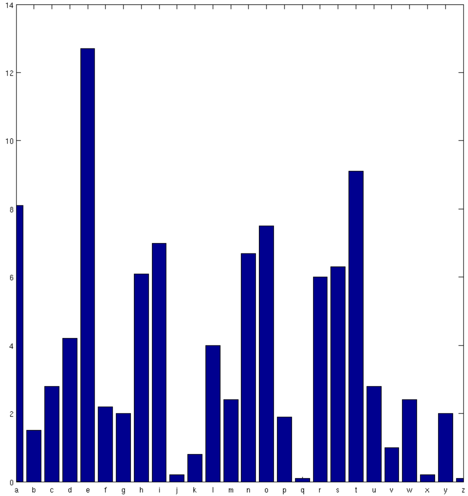
Discrete Distributions

- Today we will focus on discrete probability distributions.
- Our events will consists of a finite set, usually consisting of numbers.
- Each event will be assigned a probability between zero and one.
- We will usually represent the probabilities with a bar chart of percentages.

Discrete Example (uniform)



Discrete Example (Alphabet)



Discrete Example (Alphabet)

Example (Uniform Letters)

'fho', 'jbiz', 'dqjnq', 'ejzztiet', 'nu', 'yfmmlmy', 'o', 'ev', 'ljz', 'tixwhxf', 'rllxs', 'l', 'itovnrx', 'aqehip', 'zro', 'qwpipmg', 'qs', 'fbsqqyi', 'rsexsg', 'ncpftlug'

Example (English Distributed Letters)

'rai', 'rhfhr', 'eesef', 'tnep', 'lihyuehh', 'ouw', 'sedena', 'sssmeuh', 'i', 'lpyetwi', 'itimuh', 'hesimhae', 'atoepl', 'fcwou', 'etot', 'tcdne', 'ttierby', 'odsed', 'fid'

Probability Practice

Follow the directions on the Random Integer Sheet. Feel free to work with other students sitting nearby for parts 4, 5 , and 6.

Probability Paradoxes

- History
  - Puzzles
  - Gambling
- Formalism gives rise to intuition
- Forensic Accounting:
  - Invented data isn't truly random - even if it appears random at a first glance.

Probability Paradoxes

- History
  - Puzzles
  - Gambling
- Formalism gives rise to intuition
- Forensic Accounting:
  - Invented data isn't truly random - even if it appears random at a first glance.

Probability Paradoxes

- History
  - Puzzles
  - Gambling
- Formalism gives rise to intuition
- Forensic Accounting:
  - Invented data isn't truly random - even if it appears random at a first glance.

Birthday Problem

Problem (Birthday Paradox)

What is the probability that there is at least one pair of people in a group of 23 share the same birthday?

Solution

50.7%

If we assume that birthdays are uniformly distributed throughout the 365 days of the year, we can compute our answer as one minus the probability that no pair in our group shares a birthday:

$$1 - \underbrace{\left(\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{343}{365}\right)}_{23} = 50.7$$

Birthday Problem

Problem (Birthday Paradox)

What is the probability that there is at least one pair of people in a group of 23 share the same birthday?

Solution

50.7%

If we assume that birthdays are uniformly distributed throughout the 365 days of the year, we can compute our answer as one minus the probability that no pair in our group shares a birthday:

$$1 - \underbrace{\left(\frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{343}{365} \right)}_{23} = 50.7$$

Coin Examples

Example (1)

HTTHTHTHHTHTHHTTTHTHTHTTHTHTHTHHTHTHHTHT
HHTHHTHTTHTHTHTHTHTHHTHTHTTHTHTHHTHTHHTT
THTHTTHTHTHTHTHHTHTT

Example (2)

HTTHHHHHTHTHTHHHTTHTHHHTTTHTTHHHHTTTTTT
TTHHHHHTHHHHTHTTHTTTHHTTTHHTHHTTTHHHTHTH
HHHHTHHHHHTHHTHTHHHHTH

Coin Flipping

One of our simplest models of randomness is flipping a sequence of coins and recording the heads and tails that occur. Even in this seemingly simple setting unintuitive behavior can occur.

- Long constant subseqences
- Total transitions
- Substring Density

Coin Examples

Example (Fake)

HTTHTHTHHTHTHHTTTHTHTHTTHTHTHTHHTHTHHTHT
HHTHHTHTTHTHTHTHTHTHHTHTHTTHTHTHHTHTHHTT
THTHTTHTHTHTHTHHTHTT

Example (Real)

HTTHHHHHHTHTHTHHHTTHTHHHTTTHTTHHHHTTTTTT
TTHHHHHTHHHHTHTTHTTTHHTTTHHTHHTTTHHHTHTH
HHHHTHHHHTHHTHTHHHTH

Coin Example (Fake)

Example (Fake)

HTTHTHTHHTHTHHTTTHTHTHTTHTHTHTHHTHTHHTHT
HHTHHTHTTHTHTHTHTHTHTHHTHTHTTHTHTHHTHTHHTT
THTHTTHTHTHTHTHHTHTT

- 50 Heads and 50 Tails
- 11 HH, 40 HT, 39 TH, 9 TT
- Longest constant string TTT and HH
- Transitions 79

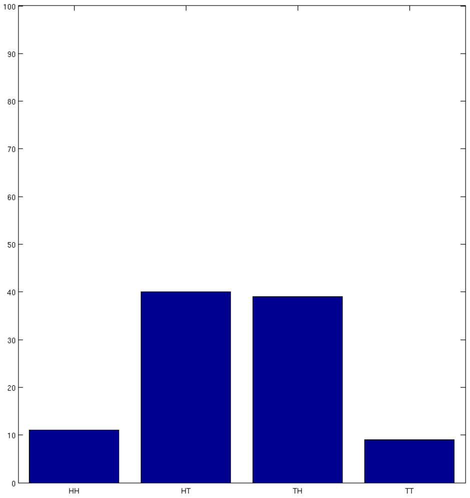
Coin Example (Real)

Example (Real)

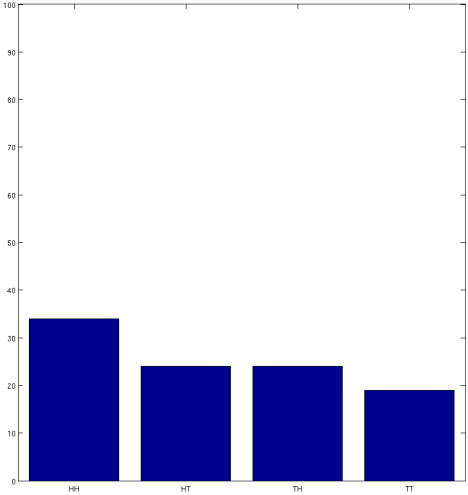
HTTHHHHHHTHTHTHHHTTHTHHHTTTHTTHHHHTTTTTT
TTHHHHHTHHHHTHTTHTTTHHTTTHHTHHTTTHHHTHTH
HHHHTHHHHTHHTHTHHHTH

- 58 Heads and 42 Tails
- 34 HH, 23 HT, 23 TH, 19 TT
- Longest constant string TTTTTTTT and HHHHHH
- Transitions 51

Coin Example (Fake)



Coin Example (Real)



Coin Flipping Practice

Try the activities on the Coin Flipping Sheet. Feel free to work with other students sitting nearby.

Coin Explanation

- Although all four substrings are equally likely to occur in a random sequence of flips, the expected number of flips until the first occurrence differs.
- For example, TH is likely to win against HH since once the first tails has been flipped TH has a 50% chance of occurring immediately, while HH must have at least two more flips.
- When you allow longer strings even stranger things can happen. The challenge problem sheet hints at one of these.

Monty Hall Practice

Try the activities on the Monty Hall Game Sheet. Feel free to work with other students sitting nearby.

Monty Hall Explanation

- The probability that you win if you switch is 2/3 while the probability that you win if you stay is only 1/3.
- This problem challenges both our mathematical and psychological intuition and has confused many professional mathematicians.
- One way to see that this is true is that after you have selected your door, the probability that the car is behind one of the other two doors is 2/3. Because the host is giving you new information about one of those doors, all of that probability is concentrated in the door that you could switch to.

Random Numbers

Before we move on, write 20 random integers of any[1] length on the More Random Integers Sheet.

[1]fairly reasonable

Digit Distributions

- Hopefully you are convinced that our intuition doesn't always work very well for probability.
- However, studying distributions over all possible numbers breaks our rule about finite distributions.
- Instead of looking at a distribution over infinitely many numbers, we will consider the distribution of digits (0–9) within the numbers.
- This turns out to be enough to discover fraudulent behavior.

Digit Example

Example (Fibonacci)

1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, 6765, 10946, 17711, 28657, 46368, 75025, 121393, 196418, 317811, 514229, 832040, 1346269, 2178309, 3524578, 5702887, 9227465, 14930352, 24157817, 39088169, 63245986, 102334155, 165580141, 267914296, 433494437, 701408733, 1134903170, 1836311903, 2971215073, 4807526976, 7778742049, 12586269025, 20365011074, 32951280099, 53316291173, 86267571272, 139583862445, 225851433717, 365435296162, 591286729879, 956722026041, 1548008755920, 2504730781961, 4052739537881

Digit Example

Example (Powers of Two)

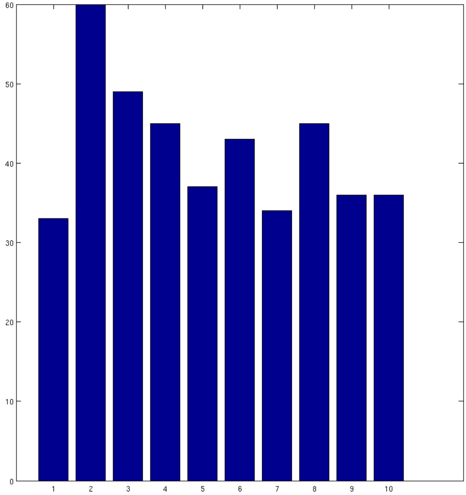
1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, 131072, 262144, 524288, 1048576, 2097152, 4194304, 8388608, 16777216, 33554432, 67108864, 134217728, 268435456, 536870912, 1073741824, 2147483648, 4294967296, 8589934592, 17179869184, 34359738368, 68719476736, 137438953472, 274877906944, 549755813888, 1099511627776, 2199023255552, 4398046511104, 8796093022208, 17592186044416, 35184372088832, 70368744177664, 140737488355328, 281474976710656, 562949953421312, 1125899906842624, 2251799813685248, 4503599627370496, 9007199254740992, 18014398509481984, 36028797018963968, 72057594037927936, 144115188075855872, 288230376151711744, 576460752303423488, 1152921504606846976, 2305843009213693952, 4611686018427387904, 9223372036854775808,

Digit Example

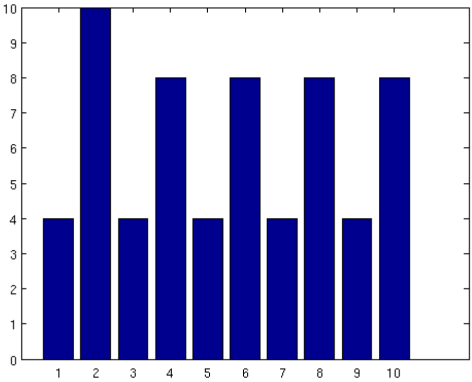
Example (US City Population)

8363710, 3833995, 2853114, 2242193, 1567924, 1447395, 1351305, 1279910, 1279329, 948279, 912062, 808976, 807815, 798382, 757688, 754885, 703073, 687456, 669651, 636919, 613190, 609023, 604477, 598707, 598541, 596462, 591833, 558383, 557706, 557224, 551789, 541811, 537958, 521999, 476050, 463794, 463789, 463552, 451572, 438646, 433748, 433746, 413201, 404155, 392552, 385635, 382605, 380307, 374676, 374417, 366046, 354361, 340882, 339130, 335288, 333336, 321078, 319057, 311853, 310037, 295357...

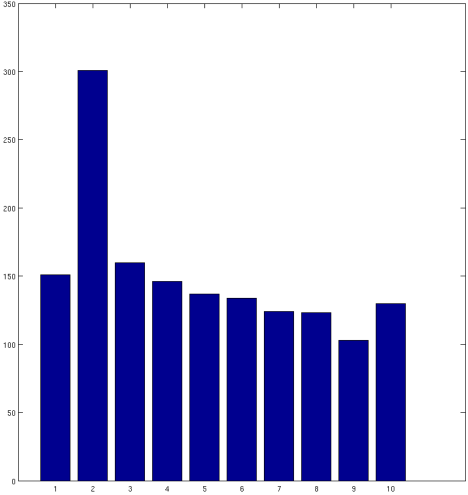
Fibonacci All Digits



Fibonacci Last Digit



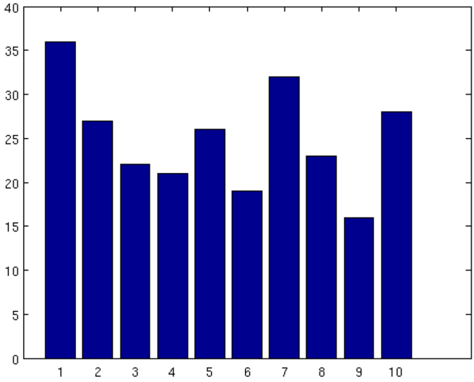
City All Digits



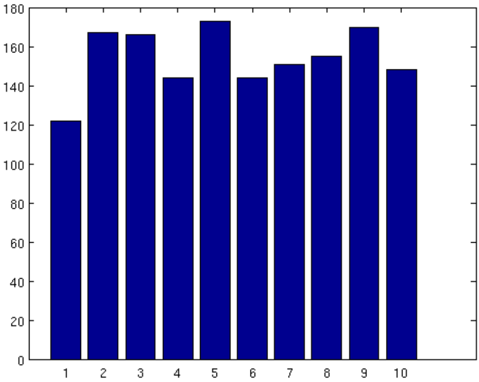
Dartmouth

City Last Digit

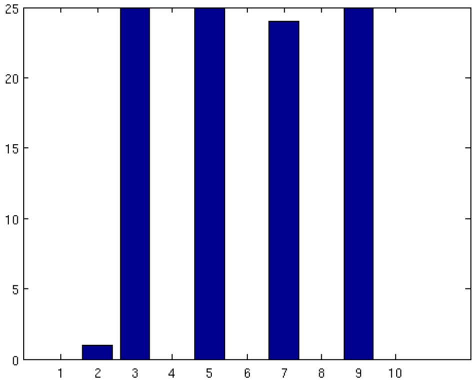
–



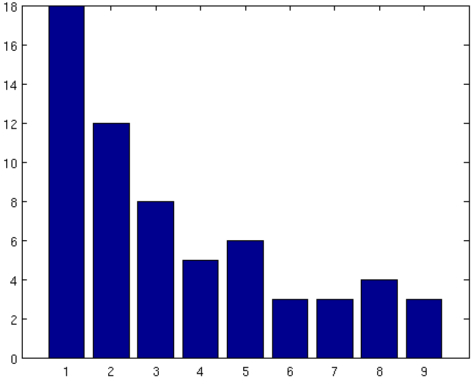
Powers of Two All Digits



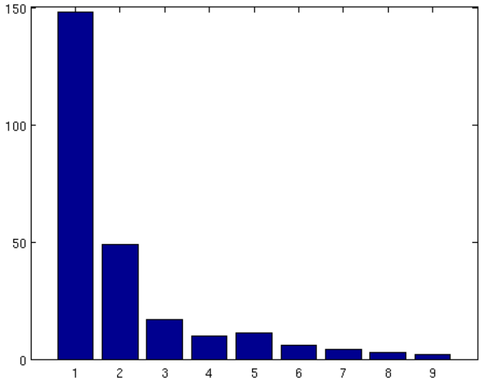
Powers of Two Last Digit



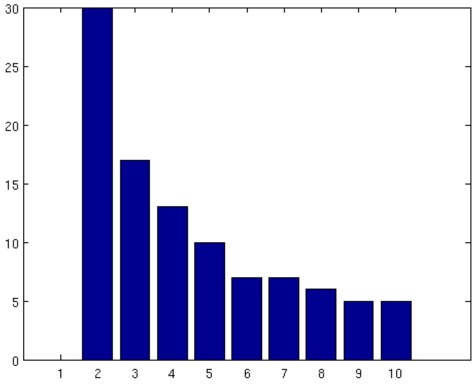
Fibonacci First Digit



City First Digit



Powers of Two First Digit



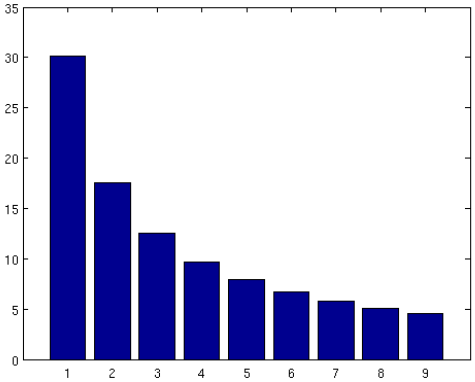
Benford's Law

- History
  - Newcomb (1881)
  - Benford (1938)
- Formula: each digit $d \in \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$ occurs with probability $P(d) = \log_{10}(d+1) - \log_{10}(d) = \log_{10}(1 + \frac{1}{d})$
- Scale invariance
- Base Invariance
- Later digits

Benford's Law

Benford's Leading Distribution	
Digit	Percentage
1	30.1%
2	17.6%
3	12.5%
4	9.7%
5	7.9%
6	6.7%
7	5.8%
8	5.1%
9	4.6%

Leading Digit Probabilities



Alphabet Distribution

The most common first letter in a word in order of frequency
T, O, A, W, B, C, D, S, F, M, R, H, I, Y, E, G, L, N, O, U, J, K

Benford Data

- Stock Market Trade Volume
- Population Numbers
- Building Heights
- River Lengths
- Street Adresses
- Financial Reports
- Scientific Data
- Power Laws
- Mathematical Constants
- Twitter Friends
- Bacterial Growth
- Power Laws
- Internet Links
- ...

Why Benford?

- Theoretically, we expect this distribution of digits when the fractional part of the logarithms of the numbers is uniformly distributed. (nearly) Equivalently, $P(x) \sim \frac{1}{x}$ over a large range.
- It is also desirable for the numbers to span several orders of magnitude
- Exponential Growth
- Products of Distributions
- Draws from different distributions.

Why Benford?

Leading Digit	Percentage Increase to next digit
1	100%
2	50%
3	33%
4	25%
5	20%
6	16.7%
7	14.3%
8	12.5%
9	11%

Log Leading Digits

Dartmouth

Benford Activity

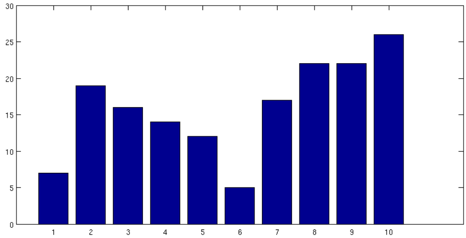
Compute the distributions for the 20 numbers you wrote previously. How do they compare to the expected distributions?

Fraudulent Checks

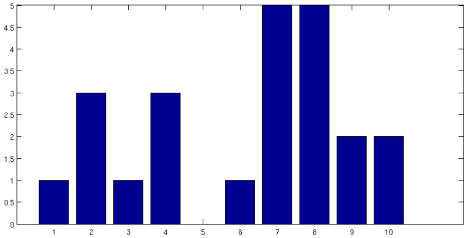
Example (Arizona v. Nelson (1993))

Wayne Nelson deposited 23 suspicious checks with the following amounts: 1927.48, 27902.31, 86241.90, 72117.46, 81321.75, 97473.96, 93249.11, 89658.17, 87776.89, 92105.83, 79949.16, 87602.93, 96879.27, 91806.47, 84991.67, 90831.83, 93766.67, 88338.72, 94639.49, 83709.28, 96412.21, 88432.86, 71552.16

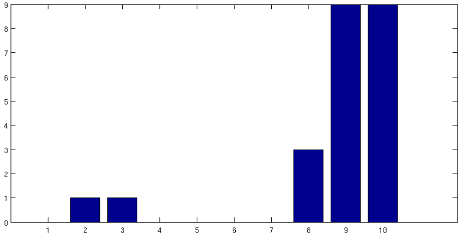
Nelson All Digits



Nelson Last Digit



Nelson First Digit



More Random Numbers

Compute the first digit distribution for the 20 random integers that you generated.

Accounting

- Enron 42% 1 and nearly 10% for both 8 and 9
- Madoff pyramid scheme has a similar pattern
- Medical Reimbursements 45% 1 money going to shell companies
- County financial records prior to bankruptcy, varied between nearly uniform and nearly peaked
- Detect significant rounding New Zealand accounting firms
- Call center rebates payouts of $49.99 to avoid $50 reporting threshold.

Voting

- Vote Switching
- Iranian, Swiss, Russian, and Mexican election results
- United States Polling companies
- Controversial Results – Science is not yet settled

Taxes

- Al Capone's tax returns had only 15% leading digit 1
- Rounding detected in Bill Clinton's taxes
- IRS Manual 4.1.10.3.1 part L
- Cost bases reporting for stocks and mutual funds
- Aggregate national tax data

Benford Activity

Try to construct a set of 20 random integers that approximately satisfies Benford's Law on the Benford Random Sheet. How close did you get?

Thanks to ...

- Dartmouth College
  Department of Mathematics

- Johns Hopkins University
  Center for Talented Youth

Further Reading

- Benford's Law: Applications for Forensic Accounting, M. Nigirini and J. Wells (2012).
- The Financial Numbers Game, C. Mulford and E. Comiskey (2002).
- Financial Shenanigans (3rd ed.), H. Schilit and J. Perler (2010).
- Courtroom Use and Misuse of Mathematics, Physics, and Finance, A. Lipton (2013).

THANK YOU!!!