Probability
概率

Wai-Ki CHING
Department of Mathematics
The University of Hong Kong

http://hkumath.hku.hk/~wkc/talks/probability.pdf

The Amazing Probability (奇異的概率)

- Some Story of Probability (概率典故)
- Calculation of Probability (概率計算)
- Mathematical Expectation (數學期望)
- Random Walk (隨機遊蕩)
- Benford's Law (本福定律)

香港大學 ·明德格物

- 「*明德*」是人與生俱來的光明皎潔的德性 。
- 「*格物*」指窮究事物的原理，而「格物」最基本的方法就是**讀書**。

「四書」中的《大學》

「大學之道, 在明*明德*, 在親民, 在止於至善。」

「……致知在*格物*，物格而後知至，知至而後意誠，意誠而後心正，心正而後身修，身修而後家齊，家齊而後國治，國治而後天下平。 」

為何習數？

- Famous star architects **Jacques Herzog** and **Pierre de Meuron** of Switzerland (Beijing Olympic Stadium) : Architecture is to make you feel **Comfortable** so that you don’t feel too cold in winter or too hot in summer.
- Mathematics is to make me feel **happier** so that I can see the **beauty** of **Life** and understand the **order** of the **Universe**.

古式・古骰

A matched Platonic-solids set of five dice, (from left) tetrahedron (4 sides), cube (6), octahedron (8), dodecahedron (12), and icosahedron (20).

Gambling and Related Activities

▪ Die	•About 2000 B.C.	•Egypt
▪ Lottery	•1st Century	•Roman
▪ Chess	•7th Century	•India
▪ Cards	•10th Century	•China
▪ Wheel	•About 1800	•France

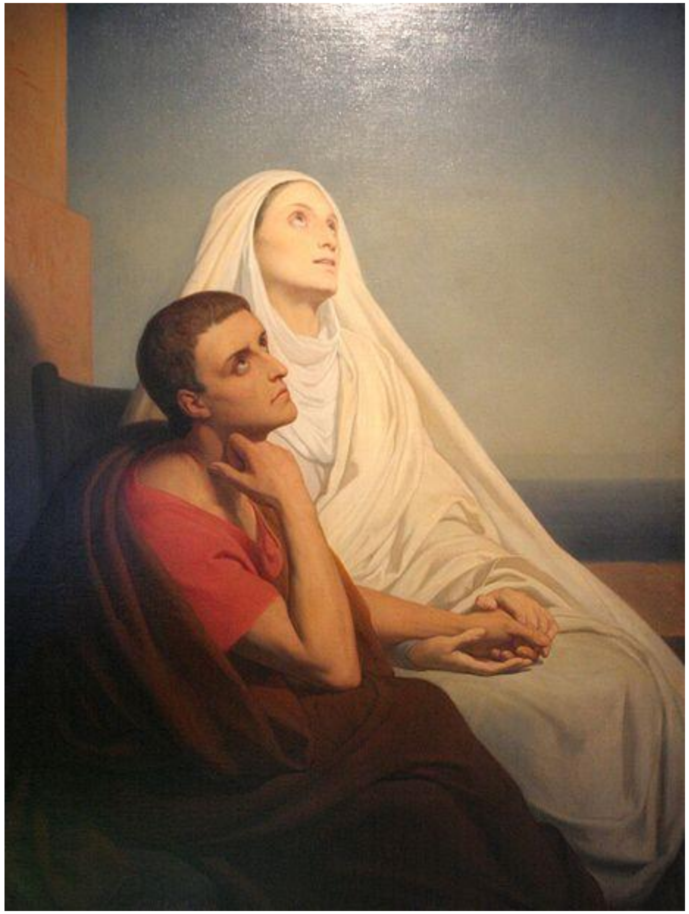
1 The Science of Probability

- With the advent of **Christianity**, the **concept of random events** developed by philosophers was rejected in the early time.
- 1. According to ***St. Augustine*** **(354-430)**, **nothing occurred by chance**, everything being minutely controlled by the will of God.

  2. If events **appear to occur at random**, then it is because of **our ignorance** and not in the nature of events.

  3. One should only seek for the will of God instead of looking at **patterns of behavior in aggregates of events**.

  (Taken from Poker faces: the life and work of professional card players by David M. Hayano, UCP Press, 1982.)



St Augustine and Monica by Ary Scheffer (1846). (维基百科全書)

1.1 The Game of Throwing Die

- The amazing contents and applications of probability theory owes its origin to **two questions on gambling** (game).
- The first question was raised by **Chevalier de Mere (雪佛萊・米爾)** (1607-1684) on his **problem of throwing a die**. He had a **title Chevalier (Knight)** and **educated at Mere**.
- The problem was solved by **Pascal (巴斯卡)**
- The second question was the **problem of points** solved by **Pascal-Fermat (巴斯卡 ・ 費瑪)**.

1517 -1648	Religion reformation (宗教改革);
12^{th} – 16^{th} century	Renaissance (文藝復興);
1637-1789	Enlightenment (啟蒙時代).

Blaise Pascal

(1623-1662)

1.2 The First Question

- De Mere made considerable money over the years in betting **double odds** on rolling at **least one ‘6’ in 4 throws** of a fair die.
- He then thought that the same should occur for betting on at **least one double-six** in **24 throws of two fair dice** (**This was their ancient believes**). It turned out that it did not work well.
- In 1654, he challenged his friend Blaise Pascal (1623-1662) for the reasons.

1.3 Answers from Pascal

- The probability of getting no ‘6’ in four **independent** throws of a fair die: $(5/6)*(5/6)*(5/6)*(5/6)=625/1296$.
- Therefore the probability of having at least one ‘6’ in 4 throws will be equal to $1 - 625/1296 = 671/1296 = 0.5177 > 0.5000$.
- This explained why de Mere got a good amount of money on double odds on his bet.

1.4 The Second Case Analysis

(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	**(6,6)**

Continue

- The probability of getting no double'6'in the throw of two fair dice is 1-(1/6*1/6)=35/36.
- The probability of getting no double 6 in 24 independent throws is $(35/36)^{24}$. Therefore the probability of having at least one double 6 in 24 throws is equal to $1 - (35/36)^{24} = 0.4914 < 0.5$.
- This explained why de Mere did not get a good amount of money on double odds on this bet.

1.5 Problem 1:

The Second Question Asked by De Mere

- Two Players A and B are playing a series of games which requires to score **5 points** (games) in order to win. In each game there is no draw.
- At the moment that Player A is leading **4 points** to **3 points**, the game was interrupted and cannot continue.

Continue

How should the players divide the stakes on the unfinished game?

	1	2	3	4	5	6	7
A	**WIN**	**WIN**	**WIN**	**WIN**	**LOSE**	**LOSE**	**LOSE**
B	**LOSE**	**LOSE**	**LOSE**	**LOSE**	**WIN**	**WIN**	**WIN**

1.6 The Response by Pascal-Fermat

- For the remaining two points, we may have

8th Point	9th Point	Final Winner
A Wins	A Wins	Player A
A Wins	B Wins	Player A
B Wins	A Wins	Player A
B Wins	B Wins	Player B

- Assume all the 4 outcomes are **equal likely** then the stake should be divided by the ratio 1:3 (B:A).

1.7 Another Response

- If P(A wins) = 4/7 and P(B wins)=3/7 according to their previous performance, we have
-

8th Point	9th Point	Final Winner	Probability
A Wins	A Wins	**Player A**	(4/7)*(4/7)
A Wins	B Wins	**Player A**	(4/7)*(3/7)
B Wins	A Wins	**Player A**	(3/7)*(4/7)
B Wins	B Wins	**Player B**	(3/7)*(3/7)

- Probability that Player B is the final winner is **9/49**.

  The stake should be divided in the ratio **9:40 (B:A).**

1.8 The Hidden Secrets of the Creative Mind
R. Keith Sawyer
(創意)

- Places where creative ideas suddenly emerged— the **Bathtub**, the **Bed** and the **Bus (3Bs)**.

1. **Take risks (承擔風險)**, and expect to make lots of **mistakes (容許犯錯).**
2. **Work hard (努力工作),** and take frequent breaks, but **stay with it over time (永不放棄) .**
3. **Do what you love (做你所愛),** because creative breakthroughs take years of hard work **(突破需時).**
4. **Develop a network of colleagues (同志網絡)** , and **schedule time for unstructured discussions (定期吹水).**

- Don't forget those **romantic myths** that creativity is all about being **gifted** and not about hard work.

2.1 The Concept of Independent (獨立) Event

- Two events A and B are said to be ***independent*** if we have

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

- Let A_1, A_2, A_3 … be independent events then the above result can be further extended to

$$P(A_1 \text{and } A_2 \text{ and } A_3 \text{ and } \ldots.) = P(A_1)P(A_2)P(A_3)\ldots.$$

2.2 A Question for Fun

- It is known that the probability an individual may engage in a bus accident is **0.0015** in Hong Kong.
- Therefore it is always better to get on a bus full of passengers than an empty bus . Because the probability of having an accident when you are the only one in the bus is **0.0015**. But the probability that you have an accident with ***n*** other people is $\mathbf{0.0015^{n+1}}$. It will be **very small** when ***n* is large**!
- Do you agree with it? Why?

2.3 The Answer

- The events are **not independent**, therefore the probability cannot be obtained by just multiplying the probabilities together.
- In fact, if the bus got an accident then everyone should be involved in the accident. These are **dependent events**.

2.4 Problem 2: Another Question for Fun

- Two fair dice of six-face are thrown, the possible outcomes of total number of dots are

  2,3,4,5,6,7,8,9,10,11,12.

- From above we notice that the number of elements in the set of even dots {2,4,6,8,10,12} is more than that in the set of odd dots {3,5,7,9,11}. Therefore we conclude that

  P(Even dots) > P(Odd dots).

- Do you agree? Why?

2.5 The Answer

- The probability of getting 2,3,….,11,12 are not equal. For example

  P(2 dots)=P({1,1})=1/6*1/6=1/36.

While probability

P(3 dots)=({(1,2),(2,1)})=2/36.

- In fact we have **P(Even dots) = P(Odd dots)**.

2.6 Mark Six

Mark Six is a popular lottery game in Hong Kong. Similar lottery game can be found all over the world. There are 49 balls (number 1 to 49) in the urn. Six balls are first drawn without replacement. The 7th ball is then drawn as the special number.

The following is some ten draws of the Mark Six Lottery in reverse order.

Date	Draw Number	Draw Results							
20/12/2002	02/110	13	18	23	24	26	33	+	15
17/12/2002	02/109	6	18	39	40	41	42	+	9
12/12/2002	02/108	7	15	16	23	31	35	+	8
10/12/2002	02/107	5	36	37	38	46	49	+	17
05/12/2002	02/106	11	21	27	31	37	44	+	1
03/12/2002	02/105	9	11	14	17	24	28	+	46
28/11/2002	02/104	17	19	26	31	37	43	+	38
26/11/2002	02/103	19	21	40	42	46	47	+	33
21/11/2002	02/102	4	16	18	25	29	41	+	21
19/11/2002	02/101	3	15	22	23	42	47	+	18

2.7 A Phenomenon for Explanation

- One observes that at **least 2** drawn numbers have the same 'first digit' in all the cases.
- This can be explained by using **Pigeonhole Principle.**
- How about the 'last digit' ?
- Observe that in about **80%** of the draws, at **least 2** drawn numbers having the **same last digit**.
- Is the machine bias?

2.8 **A Heuristic Explanation**

- We want to find p, the probability of having at **least 2 equal last digits**.
- Suffice to know q, the probability of having all the last digits being distinct (p=1-q).
- For **simplicity of calculation**, we assume that there are **50** balls and the probability of getting each ball is the same.

Continue

- **$q = (50/50)*(45/49)$**
  $*(40/48)*(35/47)$
  $*(30/46)*(25/45)$
  $= 0.2064.$
- Therefore, we have
  $p = 1- 0.2064$
  $= 0.7935 \sim 80\%$.

01	11	21	31	41
02	12	22	32	42
03	13	23	33	43
04	14	24	34	44
05	15	25	35	45
06	16	26	36	46
07	17	27	37	47
08	18	28	38	48
09	19	29	39	49
10	20	30	40	50

2.9 A Suggestion?

- The probability matches with the observation.
- When you buy Mark Six, remember to choose the six numbers such that at least two of them have the same last digit!? Do you agree?

2.10 Assignment: What is the True Probability?

- There are 7 exclusive situations.
- X1, X2, X3, X4, X5 and X6 are distinct last digit and not equal to zero.
- P(E1)=(45/49)*(40/48)*(35/47)*(30/46)*(25/45)*(20/44).
- P(E2)=(4/49)*(45/48)*(40/47)*(35/46)*(30/45)*(25/44).
- ….
- 1 – (P(E1) + P(E2) +…+P(E7))
- = 0.7935? Why?

E1	x1	x2	x3	x4	x5	x6
E2	0	x2	x3	x4	x5	x6
E3	x1	0	x3	x4	x5	x6
E4	x1	x2	0	x4	x5	x6
E5	x1	x2	x3	0	x5	x6
E6	x1	x2	x3	x4	0	x6
E7	x1	x2	x3	x4	x5	0

2.11 Problem 3: Monty Hall Problem

- Behind one of the three doors, there is a car. You are asked to open one of them and I know the correct door and can help you to open one door in the following two manners.
- (I) I open an empty door for you before you choose one.
- (II)You choose one door first without opening it, I then open an empty door from the remaining two doors. You are allowed to change your choice (you then change your choice).
- Which one will you choose (I or II)? Why?

http://www.youtube.com/watch?v=syxrYR5kYYM

2.12 Solution

- It is clear that the probability of winning the car is 1/2 for Case 1.
- For Case 2, if the player doesn't change his choice, then the probability of winning the car is 1/3.
- In Case 2, if the player changes his choice then his winning probability will be 2/3. Why? Because in this situation, the player wins if and only if he chose a wrong door at the very beginning (the chance is 2/3).

3 Mathematical Expectation

- Let x be a **random variable** taking values in {0,1,2,….,} and $P(x=0)=p_0$, $P(x=1)=p_1$,….,. Then

  $$E(x)= 0*p_0 + 1*p_1 + \dots. +$$

- The meaning of E(x), 'the expected value of x' is as follows. **We perform a large of number of identical and independent experiments**, each time the result of x is recorded. Then E(x) ***equals*** to the average value of the recorded x.

- **Expected value** is **NOT** the **value expected**!

- E.g. Let x be the number of heads obtained in tossing a fair coin one time. Then P(x=1)=P(x=0)=0.5.

  $$E(x) = 0*P(x=0) + 1*P(x=1)=0*0.5+1*0.5 = 0.5.$$

3 Continue

- Throwing a fair die and record the number of dots obtained for large number of times.
- The expected number of dots is equal to

  $1/6(1) + 1/6(2) + 1/6(3) + 1/6(4) + 1/6(5) + 1/6(6) = 3.5.$
- Here 3.5 is the average number of dots obtained when we perform the experiment of throwing a fair die for large number of times.

3.1 Pascal's Expectation

- Let p be the probability that God exists and assume that $p > 0$ (no matter how small it is). Each one must make a decision: **believe** or **not believe**.
- For an **non-believer**, if God exists, the payoff is **negative infinity** because of the serious result in the hell.
- For a **believer**, if God dose not exist, they will waste the time, money and energy etc, (say -z).

3.1 Continue

	Existence (Probability = p)	**Non-existence (Probability = 1-p)**
Believer	x	-z
Non-believer	Negative infinity	y

- E(Believer) = p*x+(1-p)*(-z) = p*(x+z)-z (finite).
- E(Non-believe) = p*(negative infinity) + (1-p)*y
  = negative infinity.
- Should we therefore be a believer? Why?

3.2 Problem 4: A Waiting Time Problem

- A Bus Company running a 24-hour bus service and the pattern of bus arrivals is as follows:
- | ← 1.5 hour→ |←0.5 hour→|← 1.5 hour →| ….
- They claim that the average waiting time of a passenger is **0.5 hour** =**30 minutes**.
- The average inter-arrival time of buses is **0.75 or 0.25**. When you arrive, you can be in any point in a time interval of length **1.5 or 0.25**. The average waiting time is **0.5*(0.75+0.25)=0.5 hour**. Do you agree? Why?

3.2 Answer to Problem 3

- For the pattern:| ← 1.5 hour→ |← 0.5 hour→| …
- The probability of getting into a longer time interval is **1.5/(1.5+0.5)=3/4** and the probability of getting into a shorter time interval is **0.5/(1.5+0.5)=1/4**.
- When you are in a longer interval, the average waiting time is **0.75** while when you are in a shorter interval, the average waiting time is **0.25**.
- Therefore the average waiting time is
- 3/4 (0.75)+1/4 (0.25)=**5/8** (37.5 minutes) > 1/2 (30 minutes).

4 The Chinese Order of Life

Destiny, fortune, fengshui, virtue, and study are the Chinese order of life.

中国人生の秩序

一命二運三風水
四積陰德五讀書
六名七相八敬神
九交貴人十養生
十一擇業與擇偶
十二趨吉及避兇

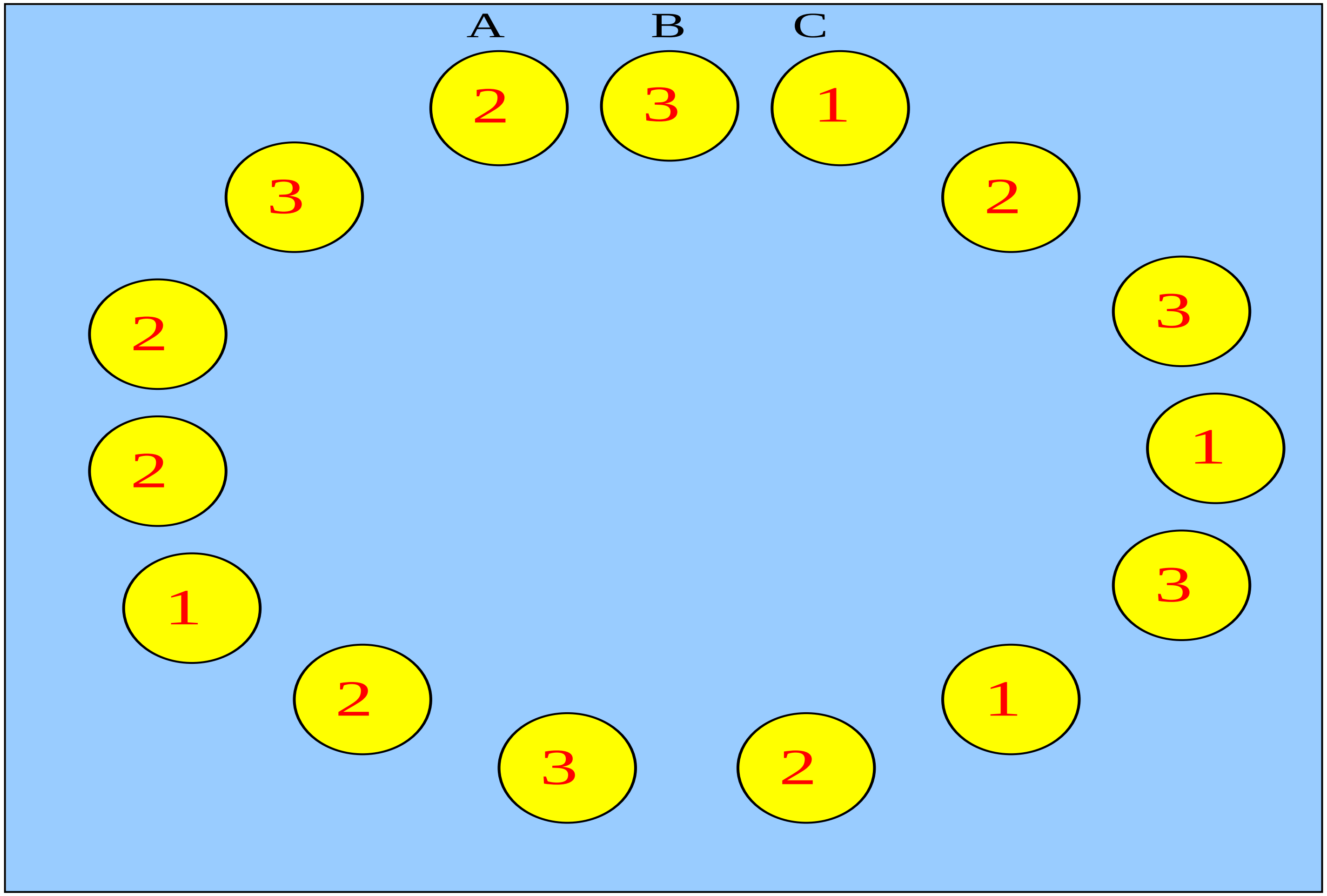
4.1 Problem 5: Cards and Random Walk

- We play a game of 3N cards. Each card has a label of 1,2 or 3 and there are N cards of 1,2,3 respectively.
- We spread the cards randomly on a table and form a circle. Randomly choose 3 consecutive cards on the table and these 3 cards form a special ZONE. For the cards in this zone, we give them extra marks A,B and C.

Continue

- Choose randomly one of the cards in the zone and perform a walk (clockwise) with the step size equals to the number on the card. Continue the walk in this manner until you come back to the zone. Then record the mark (A,B,C) where you stop. Repeat the process again.

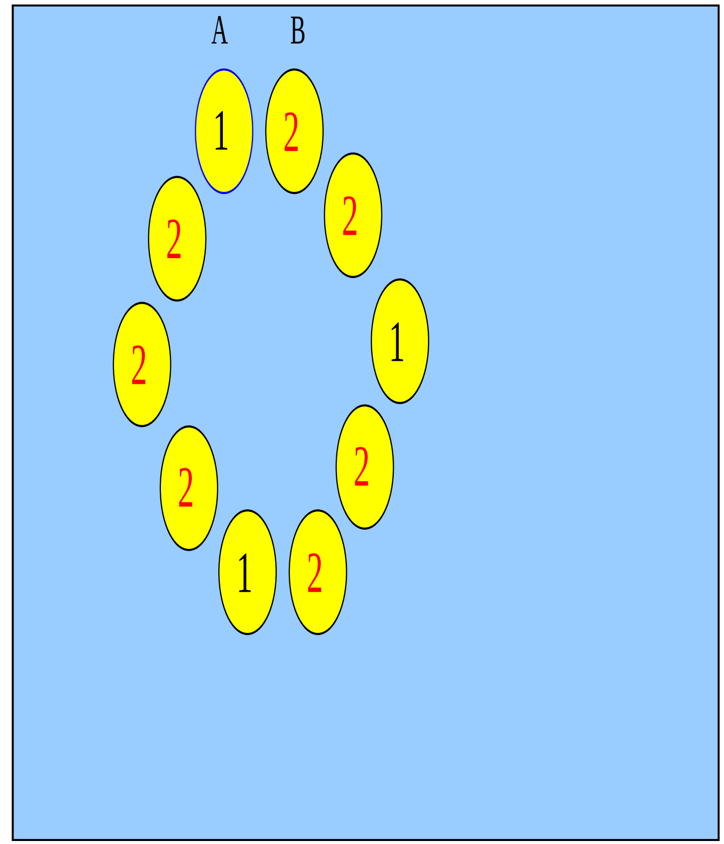
Continue



http://www.youtube.com/watch?v=q70o3c8t3D8

4.2 The Random Walk with Cards (The case of m=2)

Please put 1 or 2 randomly in the circles. Then begin with one of the circles in the zone (A,B) and perform a walk (clockwise) with the step size equals to the number on the circle. Continue the walk in this manner until you come back to the zone. Then record the mark (A,B) where you stop. Repeat the process again for the other one in the zone.



4.2 Some Observations

- The step size in each move is at most 2. Therefore it is useful to look at patterns of 2 consecutive cards.
- The total number of possible patterns = $2^2=4$.
- Among the patterns, there are 3 'convergent patterns' in blue. When the walker is in any one of the positions, his future path will be converged. The exceptional pattern is 22.
- Therefore for a random walk of 2*N cards with numbers either 1 or 2, the only case that it has no convergent path is that all the cards have the same number 2 on them.

11 12 21 22

4.3 Some Observations

- The step size in each move is at most 3. Therefore it is useful to look at patterns of **3 consecutive** cards.
- The total number of possible patterns = 3^3=27. Why?

Continue

- Among the patterns, there are **12 'convergent patterns'** in blue. When the walker is in any one of the positions, **his future path will be converged**.

111	112	221	121	212	132	211	113	232
213	321	311	222	333	123	331	313	122
133	231	223	332	312	131	323	322	233

4.4 Convergent Probability

- We note that if the **3 walker paths** do not converge with each other then it implies that we CANNOT find 3 consecutive cards taking a blue pattern in the 3N cards.
- This then implies that, cards 1 to 3, 4 to 6, 7 to 9, …., 3(N-1)+1 to 3N take no **blue patterns**.
- P(no convergent paths) < $(1\text{-}12/27)^N = (5/9)^N$.
- For N=5, the probability is less than 0.0529 ~5%.

5 Benford's Law

- In 1881, a Mathematician **Simon Newcomb** noticed that the pages of logarithm tables with **small initial digits** were **dirtier** than those with **larger initial digits**, such that

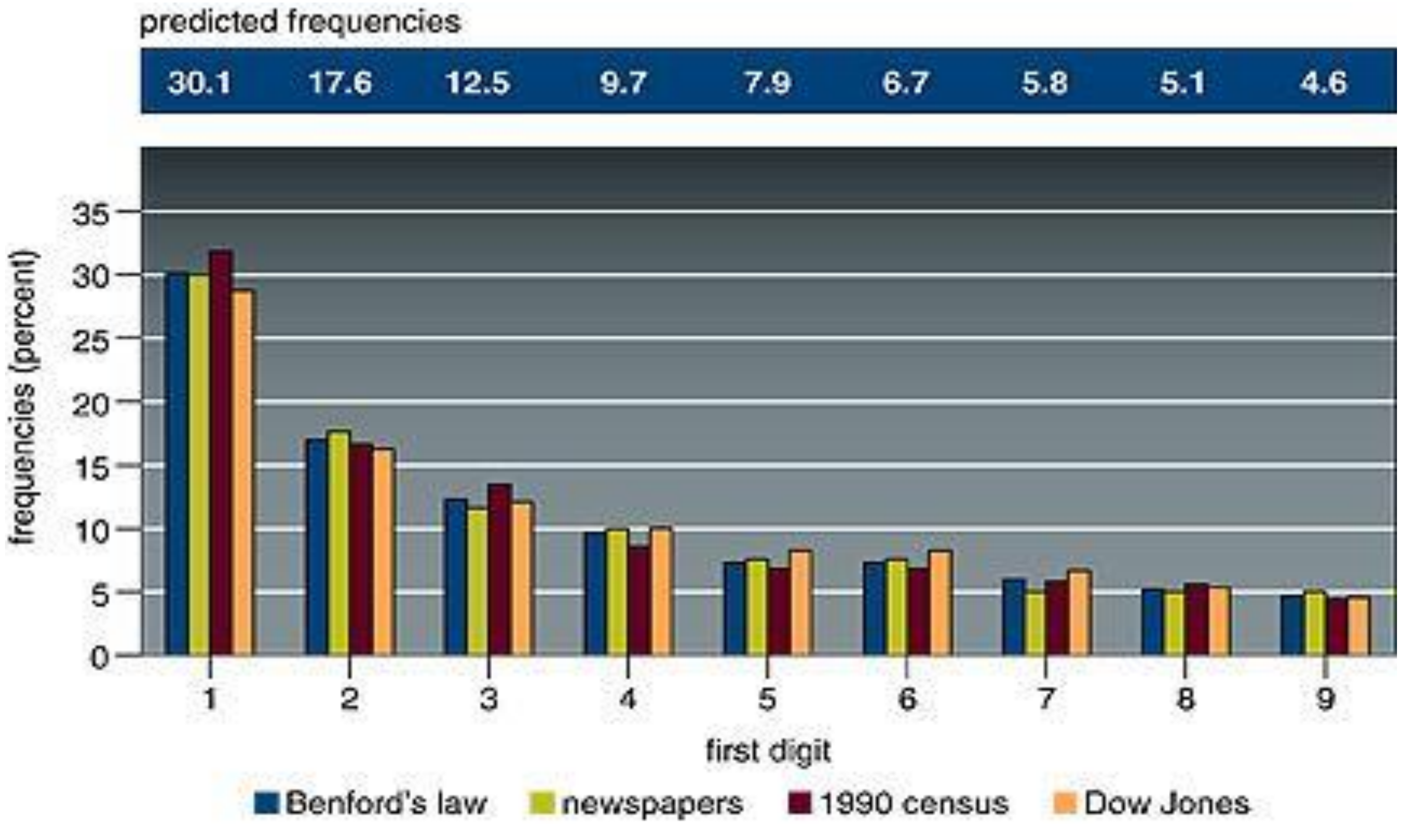
$$1>2>3>4>5>6>7>8>9.$$

- In 1938, a Physicist Frank Benford proposed the Benford's law based on the **empirical evident**:

  P(The first significant digit = d) = $\log_{10}(1+1/d)$

  for $d=1,2,3,4,5,6,7,8,9$.

5.1 Frequency of First Digits, From 1 to 9.



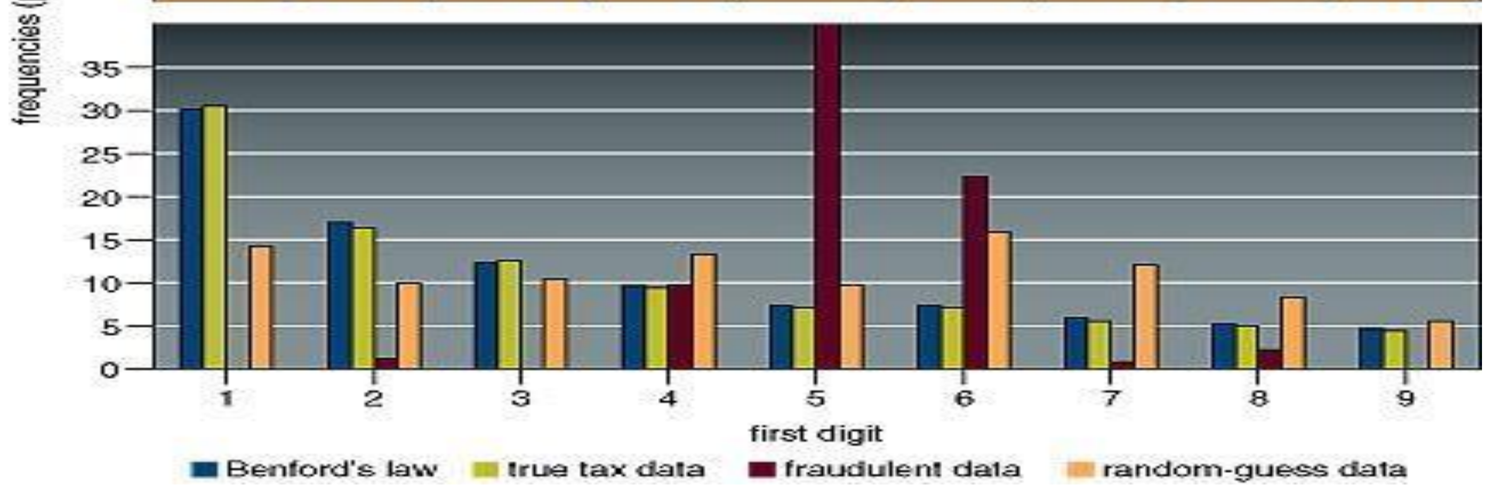
From "The First-Digit Phenomenon" by T. P. Hill, *American Scientist*, July-August 1998. [*The New York Times*, Tuesday, August 4, 1998]

5.2 Detecting Fraud

- An interesting application of Benford's law is to help in detecting possible **fraud** in **tax returns**.
- Empirical research in US has shown, the interest paid and received are very good fit to Benford's law.

5.3 First Significant Digits Tax Data

	1	2	3	4	5	6	7	8	9
Benford's law	30.1	17.6	12.5	9.7	7.9	6.7	5.8	5.1	4.6
true tax data	30.5	17.8	12.6	9.6	7.8	6.6	5.6	5.0	4.5
fraudulent data	0	1.9	0	9.7	61.2	23.3	1.0	2.9	0
random-guess data	14.7	10.0	10.4	13.3	9.7	15.7	12.0	8.4	5.8



From "The First-Digit Phenomenon" by T. P. Hill, *American Scientist*, July-August 1998. [*The New York Times*, Tuesday, August 4, 1998].

5.4 A Heuristic Analysis

- Suppose in month 0, Hang Seng Index is 100.
- We assume that it **increases at a rate of 10 percent** (it can be r%) per year.
- Let f(1) be the number of years for the index to reach 200 from 100, then we have

$$100*(1.1)^{f(1)}=200$$

or

$$f(1)=(\log(200)-\log(100)) / \log(1+1/10)$$

Continue

- Let f(2) be the number of years for the index to reach 300 from 200, then we have

$$200*(1.1)^{f(2)} = 300$$

or

$$f(2)=(\log(300)-\log(200)) / \log(1+1/10)$$

- Inductively we have for d=1,2,3,4,5,6,7,8,9

$$f(d) = (\log(100(d+1))-\log(100d)) /\log(1+1/10)$$

- We note that $\log(100d)=\log(100)+\log(d)$ and $\log(100(d+1))=\log(100)+\log(d+1)$.

Continue

- Thus f(d) can be simplified as follows:

$$f(d) = \log(1+1/d) / \log(1+1/10)$$

- We also note that log(10)=1 and

  $F=f(1)+f(2)+\ldots+f(9) = \log(10)/\log(1+1/10).$

- Therefore the probability of observing (d=1,2,3,4,5,6,7,8,9) as the first digit is

$$P(d) = f(d)/F = \log(1+1/d).$$

PROBLEM 6: Which of the following receipts is a fake one? Receipt A (Green) or Receipt B (Blue)

125	225	610	156	781	263	355	665	188	455
441	186	503	335	295	885	163	559	905	103

611	885	236	526	920	760	441	950	654	758
995	162	165	365	578	856	356	513	234	425

一期一会 (いちごいちえ)

一期一會，日本茶 道 的用語。

「一期」，表示人的一生；

「一會」，意味着僅有一次相會，勸勉人們應知所珍惜身邊的人。