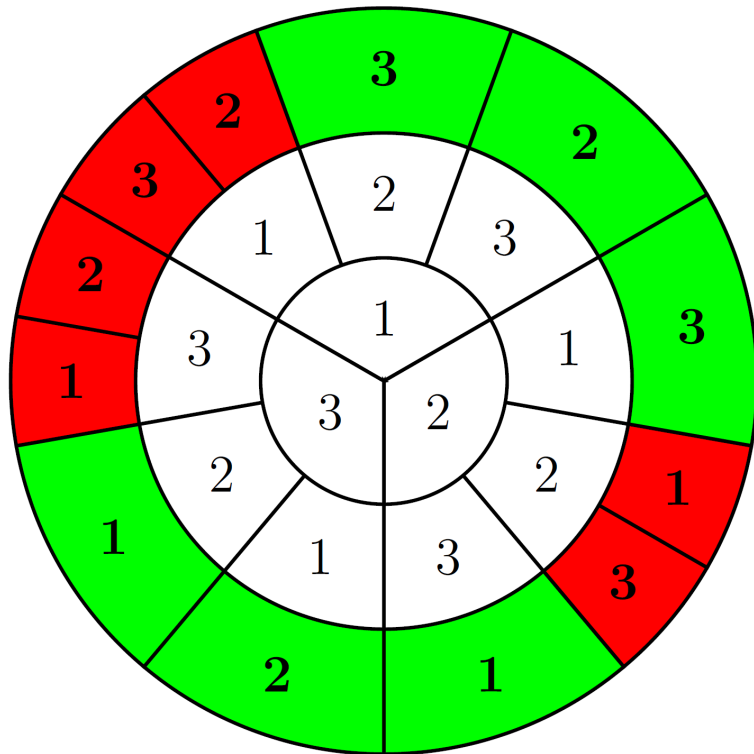


Monkeys and Coincidences

The Index of Coincidence

Monty Hall (host knows)



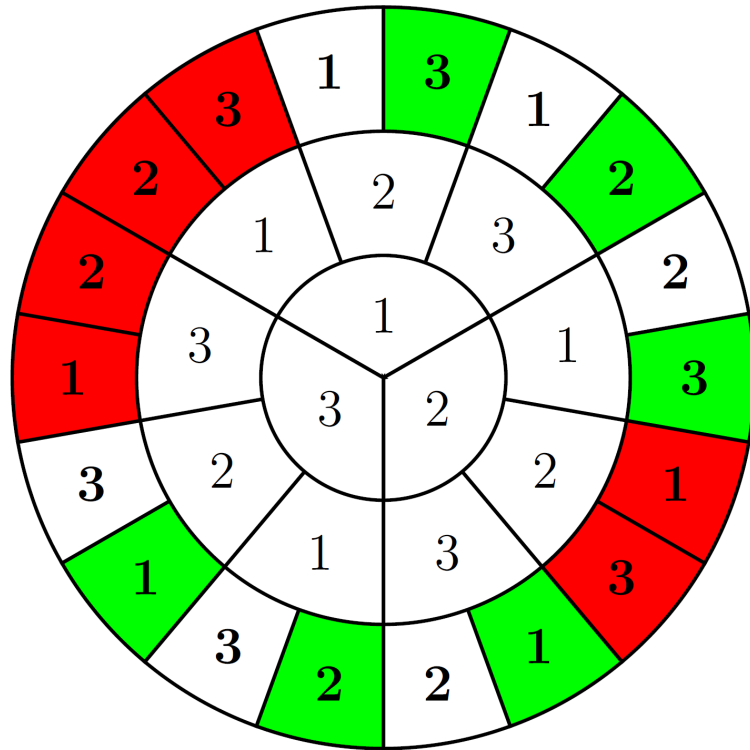
Inner wheel: where the car is
Middle wheel: door that you choose
Outer wheel: door that Monty opens

Green: switching wins
Red: staying wins

$$P[\text{red}] = 1/3$$

$$P[\text{green}] = 2/3$$

Monty Hall (host doesn't know)



Inner wheel: where the car is
Middle wheel: door that you choose
Outer wheel: door that Monty opens

Green: switching wins

Red: staying wins

White: didn't happen

$P[\text{green} \mid \text{red or green}]$

$$= (6/18)/(12/18) = 1/2$$

$P[\text{red} \mid \text{red or green}]$

$$= (6/18)/(12/18) = 1/2$$

Monkey Words

- We used letter frequencies to recognize English.
- Allowed Eve to get key one letter at a time.
26 choices for e
25 choices for t
...
26+25+24+...+2+1 options to inspect = 27·13 = 351

not

26 · 25 · 24 ... 2 · 1 keys = 26!

- Stirling's Formula:

$$n! \approx \sqrt{2n\pi} \left(\frac{n}{e}\right)^n$$

Monkey Words

- The monkey can do better with roulette wheels
- <http://www.math.ucsd.edu/~crypto>

Index of Coincidence

	a	n	i	m	a	l	a	a	r	d	v	a	r	k
a	■				■		■	■				■		
n		■												
i			■											
m				■										
a	■				■		■	■				■		
l						■								
a	■				■		■	■				■		
a	■				■		■	■	■			■		
r									■				■	
d										■				
v											■			
a	■				■		■	■				■		
r									■				■	
k														■

Index of Coincidence

- text had 14 characters. Grid has 14·13 nondiagonal places.
- text had 5 "a"s, which caused 5·4 nondiagonal blue squares
- text had 2 "r"s, which caused 2·1 nondiagonal red squares
- text had 1 "n", which caused 1·0 nondiagonal colored squares
- text had 0 "j"s, which caused 0·(-1) nondiagonal colored squares
- Probability that a nondiagonal square is colored:
$$22/182=12.09\%$$

Index Of Coincidence

- How does the table change if we replace each letter with the one after it (Caesar shift)?
- How does the table change if we use monoalphabetic substitution?
- How does the table change if we use Vigenere encryption?
- How does the table change if we use Rectangular Transposition?
- How does the table change if we use Playfair encryption?
- How does the table change if we use ADFGVX?

Index of Coincidence

- If we make a "plaintext" of N characters by spinning a 26-slot roulette wheel (all slots the same size), what is the probability that two randomly chosen characters are the same letter?
- There are approximately $N/26$ letters "A", so there are

$$\frac{N}{26} \left(\frac{N}{26} - 1 \right)$$

ways to choose an "A" and then choose another one, and the same number of ways to choose a pair of "B"-s, ..., and the same number of ways to choose a pair of "Z"-s. That is, there are

$$N \left(\frac{N}{26} - 1 \right)$$

ways to choose an ordered pair of characters which are the same letter. There are $N(N - 1)$ ways to choose two characters.

Index of Coincidence

The probability of getting the same letter twice is

$$\frac{N(N/26 - 1)}{N(N - 1)} = \frac{N - 26}{26N - 26} \rightarrow \frac{1}{26}$$

This should make sense. Whatever the first letter we picked turned out to be, the second one had probability $1/26$ of being the same. Note that $1/26$ is about 3.85%.

Index of Coincidence

- Fine for random text, but what about actual text?
- Suppose that our text has length N , with N_0 "A"-s, N_1 "B"-s, ..., N_{25} "Z"-s. The probability of getting the same letter twice is

$$\frac{N_0(N_0 - 1) + N_1(N_1 - 1) + \cdots + N_{25}(N_{25} - 1)}{N(N - 1)}$$

$$\frac{N_0^2 + N_1^2 + \cdots + N_{25}^2 - N}{N(N - 1)}$$

Index of Coincidence

- We can compute this using only the ciphertext!
- For unencrypted text, we can go further. The front page of the New York Times yields the frequency table

Frequency Table for Front Page of New York Times, April 16, 2004

A	8.41	J	0.15	S	6.66
B	1.56	K	0.68	T	8.98
C	3.66	L	4.06	U	2.28
D	3.61	M	2.60	V	0.96
E	12.84	N	7.64	W	1.72
F	2.05	O	7.04	X	0.24
G	1.99	P	2.28	Y	1.54
H	4.63	Q	0.15	Z	0.11
I	7.79	R	6.36		

These are the percentages of the 24580 characters in the news articles which the website describes as "front page news".

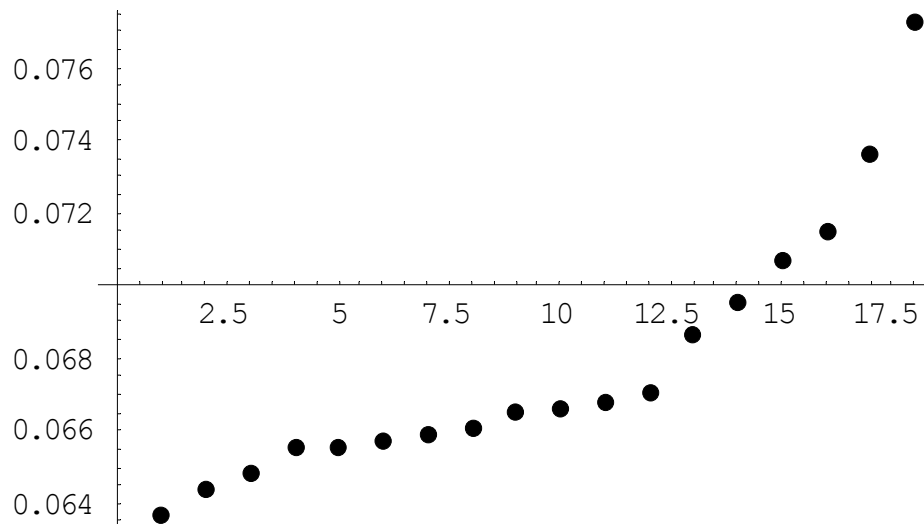
Index of Coincidence

- These frequencies are roughly the same for any text written in English, unless the author is deliberately trying to screw them up.
- Thus, in a plaintext of N characters, roughly $0.0841 N$ of them will be "A", $0.1284 N$ of them will be "E", etc.
- The sum $N_0^2 + N_1^2 + \dots + N_{25}^2$ will be roughly $0.06686 N^2$. The probability of choosing the same letter twice is

$$\frac{0.06686N^2 - N}{N(N - 1)} \rightarrow 0.06686$$

Variation in the IOC

- The IOC varies much less than single letter frequencies.
- The Void (book written without the letter "e") has IOC 7.72%.
- I have 18 standard plaintexts. Here are the IOCs:

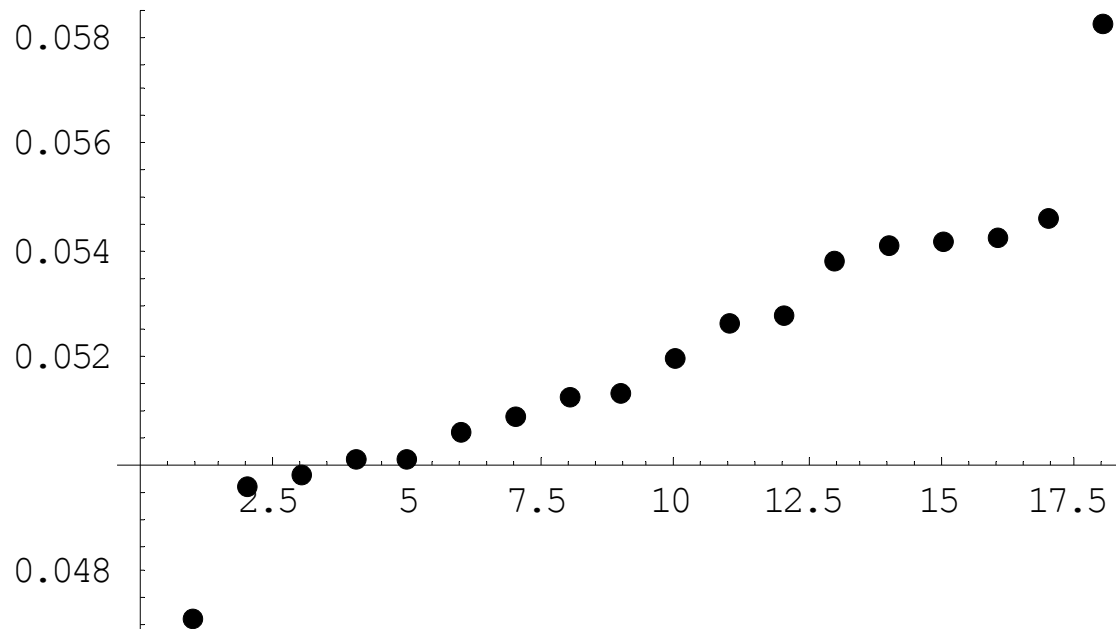


Two-Letter Vigenère

- Let c_0, c_1, \dots, c_{n-1} be a message encrypted with Vigenère with a 2-letter keyword.
- Let E be the event that $c_i = c_j$, with i, j chosen uniformly but not equal.
- $$P[E] = P[E \mid i, j \text{ both even}] P[i, j \text{ both even}]$$
$$+ P[E \mid i \text{ odd}, j \text{ even}] P[i \text{ odd}, j \text{ even}]$$
$$+ P[E \mid i \text{ even}, j \text{ odd}] P[i \text{ even}, j \text{ odd}]$$
$$+ P[E \mid i \text{ odd}, j \text{ odd}] P[i \text{ odd}, j \text{ odd}]$$
- $$P[E] = 0.06686 (1/4)$$
$$+ (1/26) (1/4)$$
$$+ (1/26) (1/4)$$
$$+ 0.06686 (1/4)$$
$$= 5.27\%$$

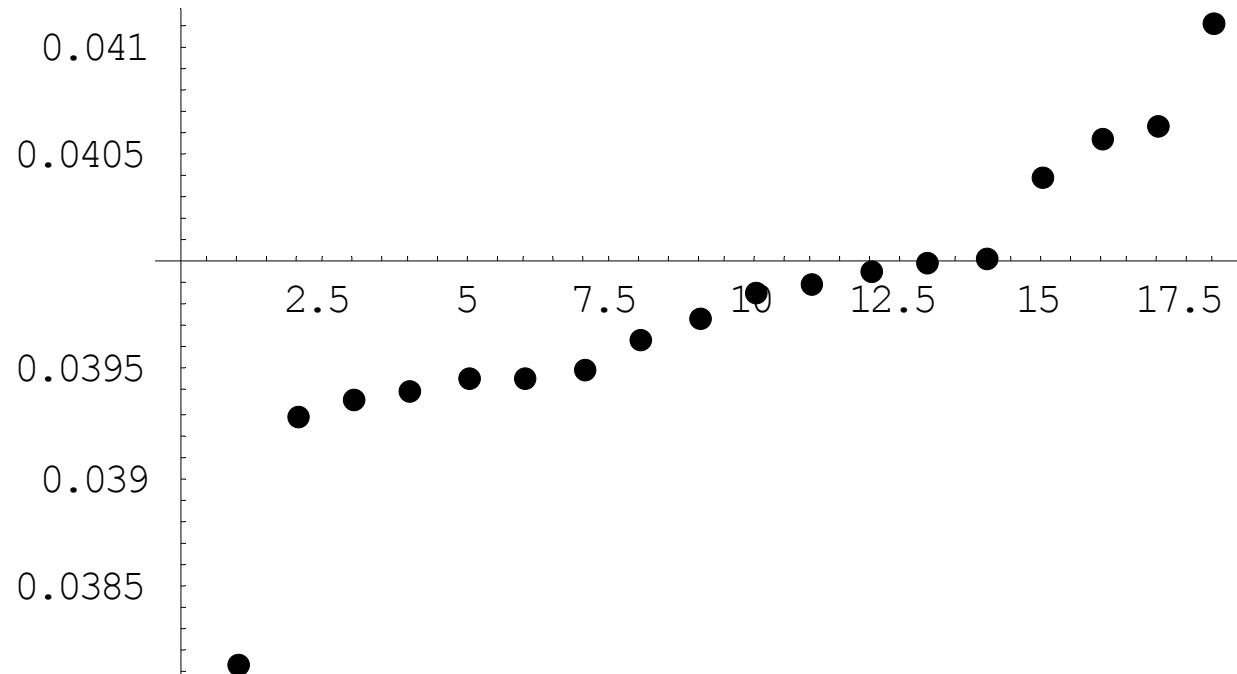
Two-letter Vigenère

Here are the indices of coincidence of my 18 plaintexts, after encryption with a two letter keyword.



20-letter Vigenère

- Here are the indices of coincidence of my 18 plaintexts after Vigenère encryption with a random 20-letter keyword.



Summary

- Goats and cars paradox is not paradoxical
- Index of coincidence is around 3.85% for random text
- IOC for English is around 6.69%
- Frequency analysis depends on high IOC
- Can use this to distinguish between cryptosystems