Applying Statistics 101

Vong Jun Yi • Elyssa Ayu

1 Monty Hall dilemma

2 Benford’s Law

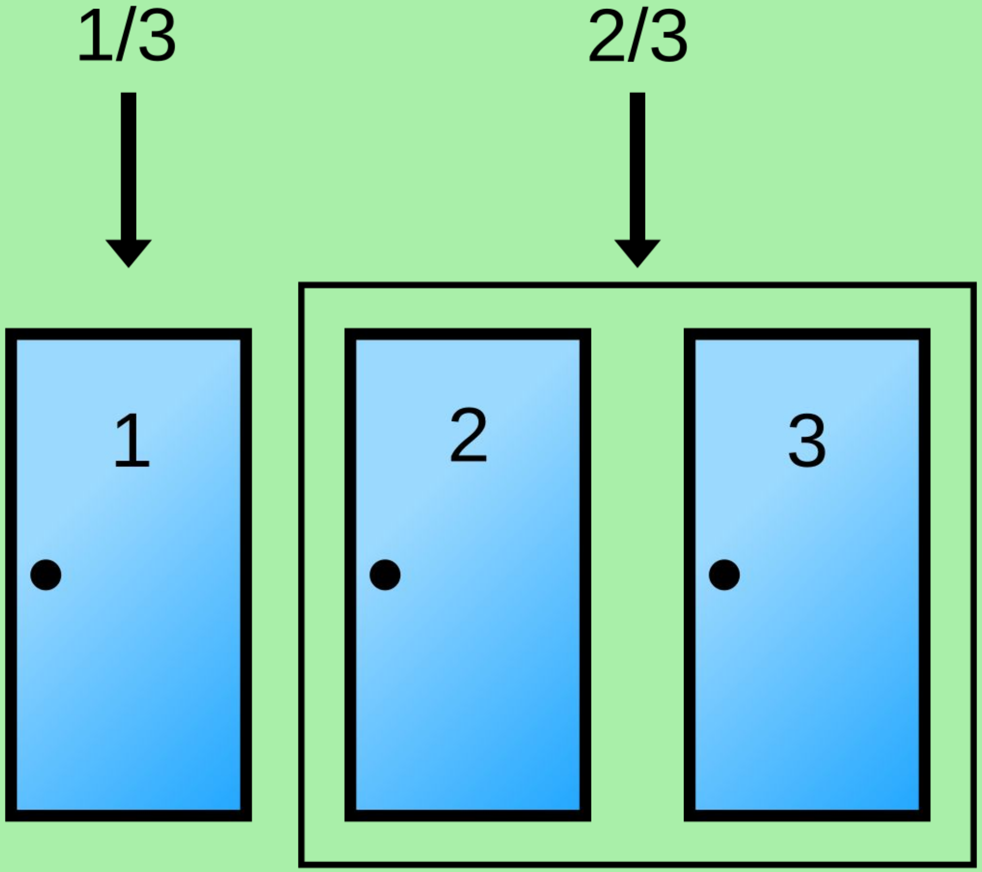
3 Other applications

1

Monty Hall dilemma

Monty Hall dilemma

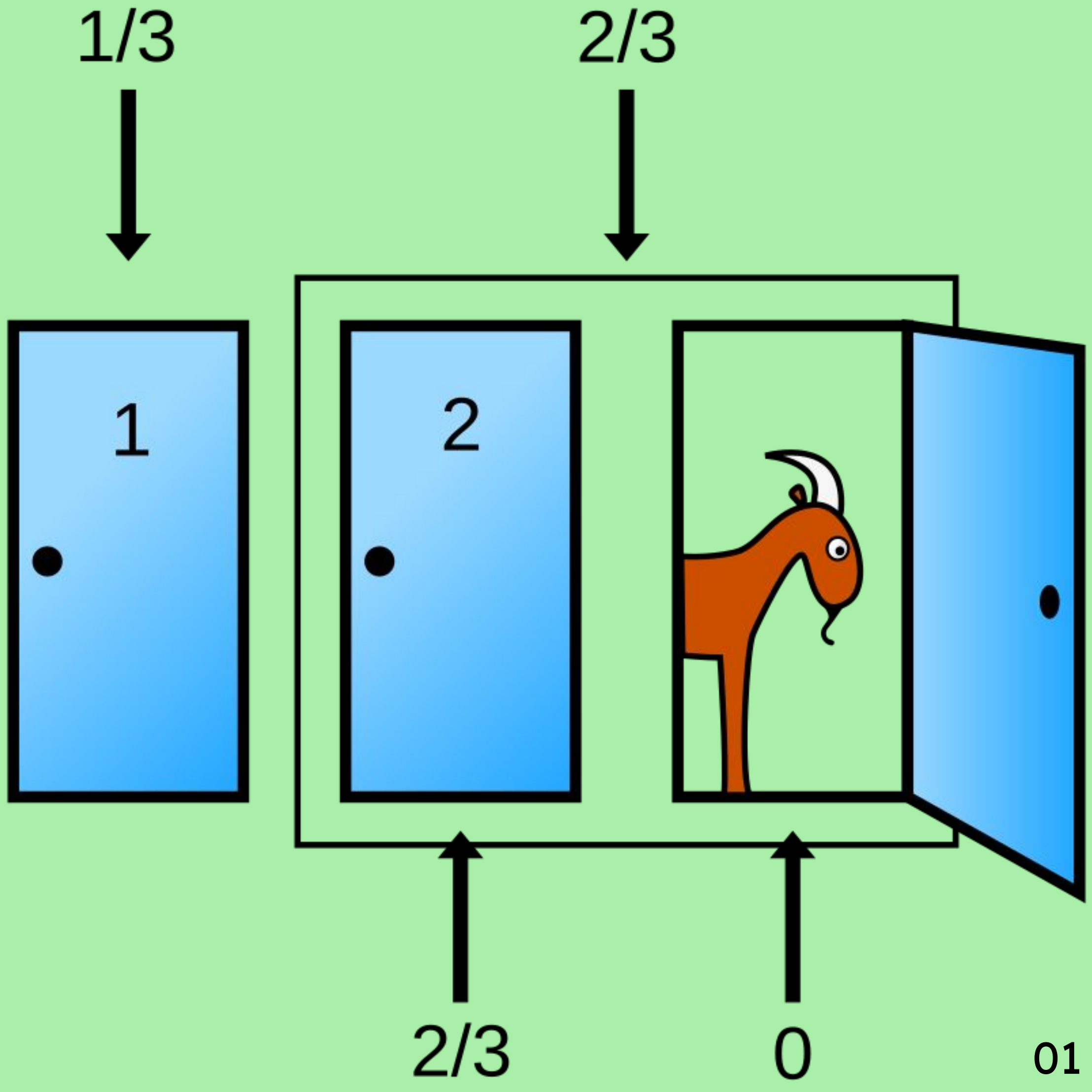
You're in a game show (Let's Make a Deal). The host (Monty Hall) leads you to wall with 3 closed doors. Behind one of the doors is the car of your dreams, and behind each of the other two is a goat. The 3 doors all have even chances of hiding the car.



01 Monty Hall dilemma

Monty Hall dilemma

The host knows precisely what is behind each of the three doors, explains how the game will work. First, you will choose a door without opening it, knowing that after you have done so, the host will open one of the 2 remaining doors to reveal a goat. When this has been done, you will be given the choice of switching doors or not.



01 Monty Hall dilemma

Question:

What strategy should you use to probabilistically optimise your chances?

Draw a table! If you choose Door 1:

Behind door 1	Behind door 2	Behind door 3	Result if staying at door #1	Result if switching to the door offered
Goat	Goat	**Car**	Wins goat	**Wins car**
Goat	**Car**	Goat	Wins goat	**Wins car**
Car	Goat	Goat	**Wins car**	Wins goat

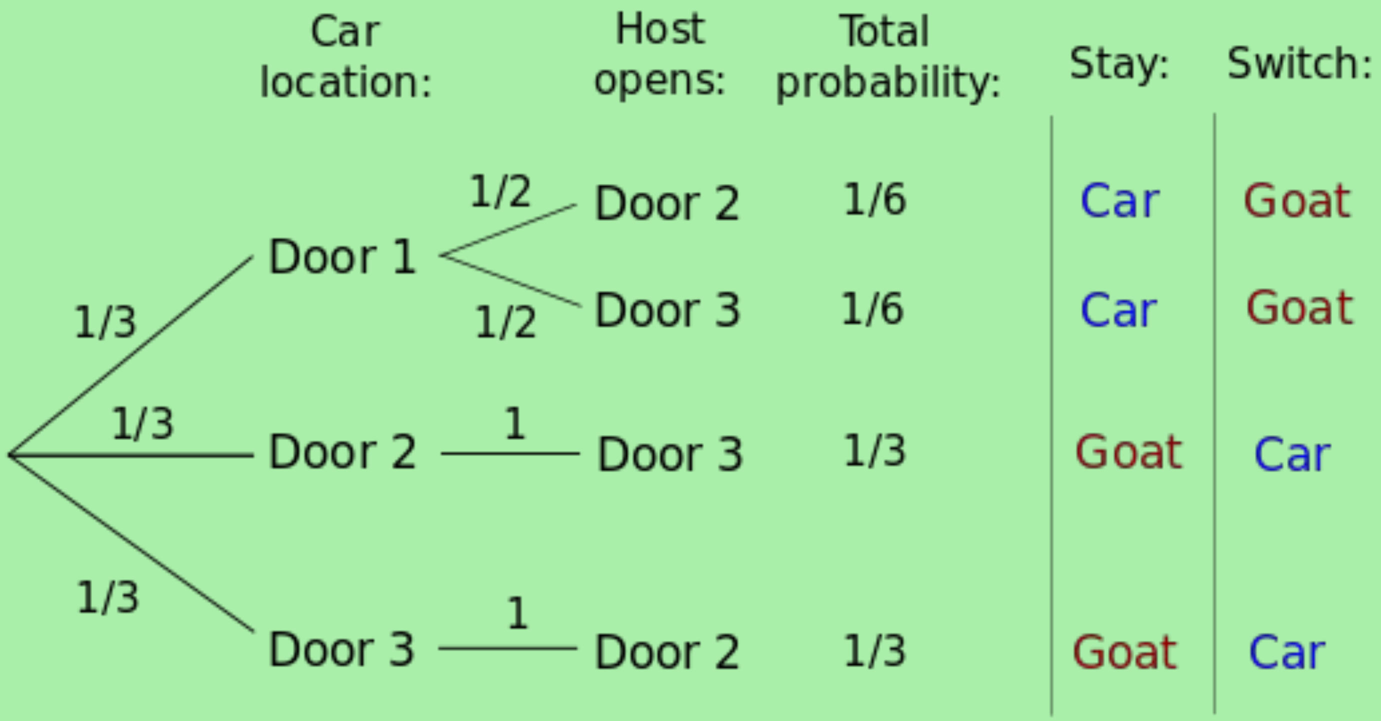
What can we conclude?

Define *win* as winning the car and *loss* as winning the goat instead.

Probability that you *win* **given** that you choose to switch → $\mathbb{P}(\text{win}|\text{switch}) = \frac{2}{3}$

Probability that you *lose* **given** that you choose to switch → $\mathbb{P}(\text{loss}|\text{switch}) = \frac{1}{3}$

We can use **Bayes' Theorem** too



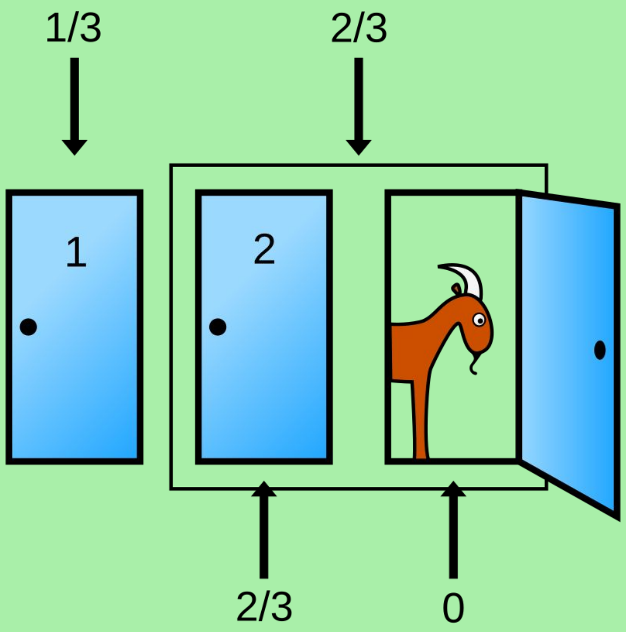
We can use **Bayes' Theorem** too

Note: A *win* and switching are NOT INDEPENDENT events.

$$\mathbb{P}(\text{win} \cap \text{switch}) = \frac{2}{3} \times \frac{1}{2} = \frac{1}{3}$$

$$\mathbb{P}(\text{switch}) = \frac{1}{2}$$

$$\therefore \mathbb{P}(\text{win}|\text{switch}) = \frac{1/3}{1/2} = \boxed{\frac{2}{3}}$$



Related questions to ponder on

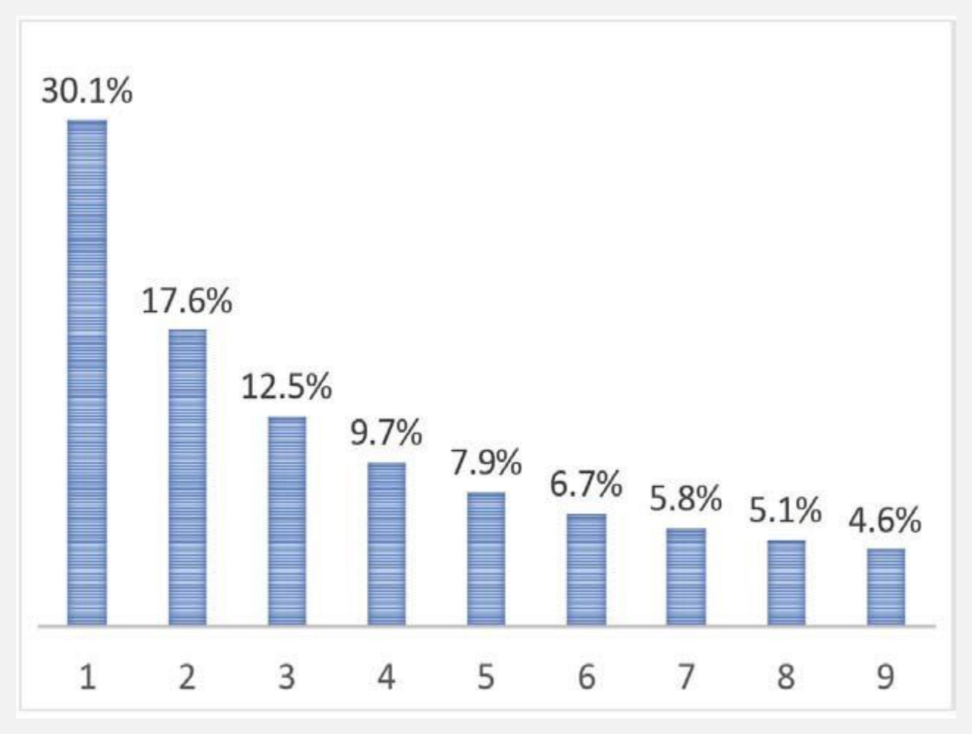
1. Is there a strategy which guarantees that you win the car **100% of the time**? What about 50%?

1. What if there were **more doors**? Suppose you are given n doors and the host opens n-2 other doors simultaneously after you pick a door. (Hint: The answer is more obvious.)

2

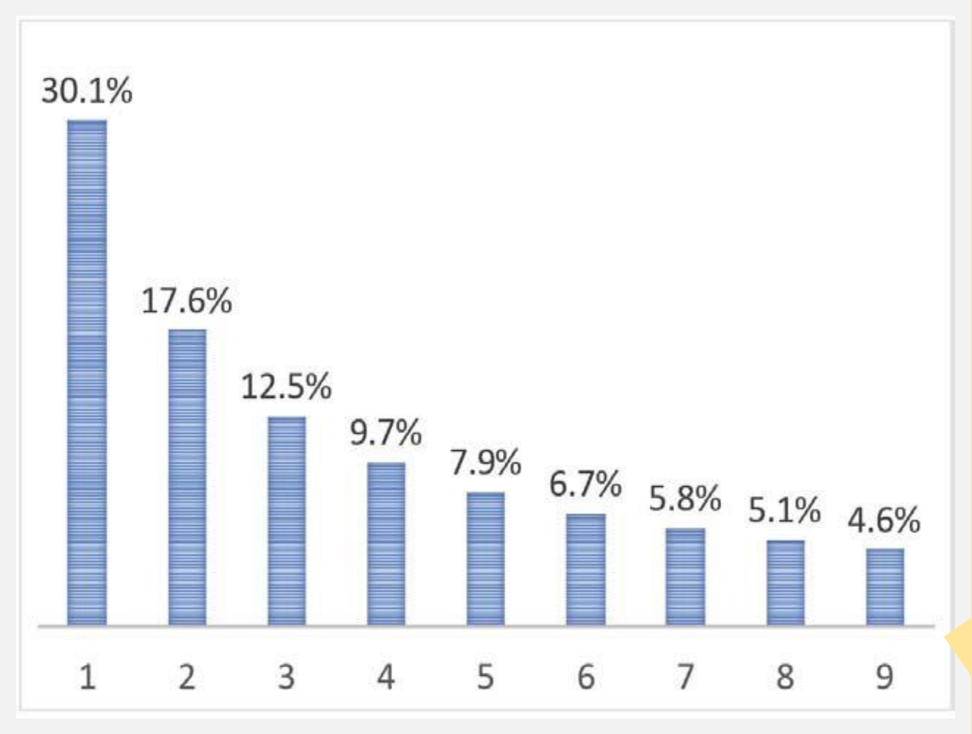
Benford's Law

What is Benford's Law?



A law which describes the **frequency distribution of leading digits** of numbers in a naturally-occuring dataset.

Examples of datasets used: populations, area of rivers, pressure, birth/death rate and anything else!



Therefore, the law states that:

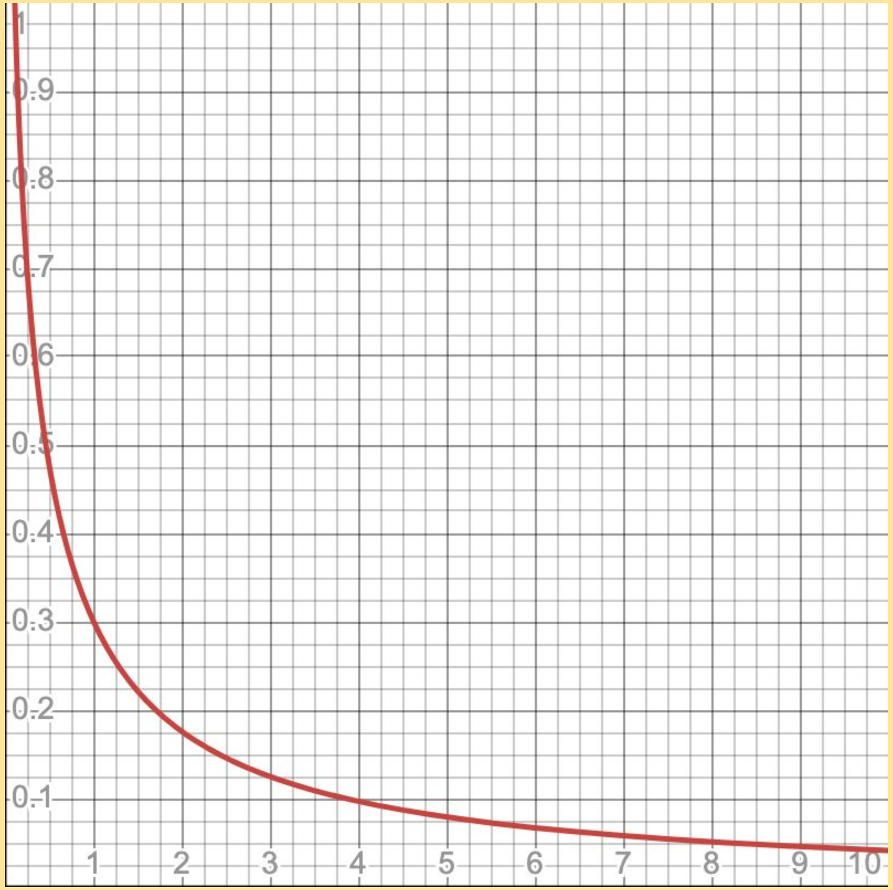
Leading digits with smaller values occur much more frequently than larger values.

Benford’s Law Formula

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

This logarithmic law shows a near perfect approximation of the probabilities after combining different data sets!

Benford’s Law Formula



Question:
How is Benford’s Law relevant and how can it be applied to real-life?

Fraud Detection

- Fraudulent data often do not follow Benford's Law as it is difficult to manually construct distributions which follow the law.

- For example: Fraudsters might change $1000 transactions into $900 ones (which obviously will skew the distribution).

Multiplication Game

Instructions:

1. A game dealer generates a random four-digit positive integer, which you cannot see, from a slot machine.
2. At the same time, you write down a positive integer with as many digits as you like on a piece of paper.
3. The product of the two integers is then found.

Multiplication Game

Rule:

If the first digit of the product is from 4 to 9, you win.
If the first digit of the product is from 1 to 3, you lose.

Multiplication Game

Seems like a good deal right?

Theoretically, if you find all the products of any two 4-digit integers from 1000 to 9999:

$$P\ (win) = 0.430$$

Multiplication Game

However, thanks to Benford's Law:

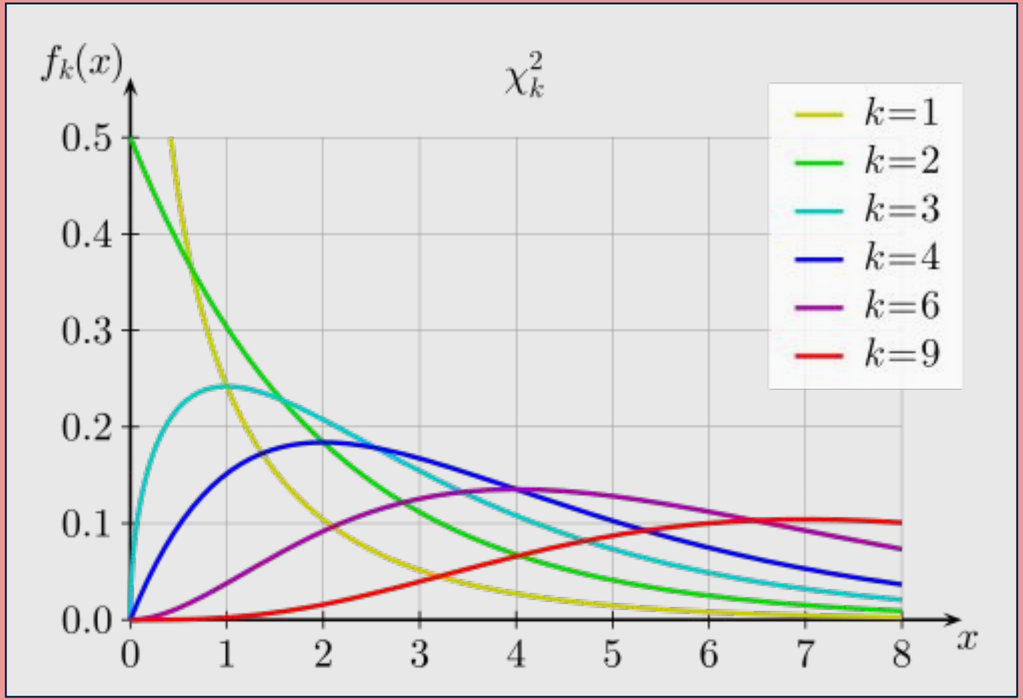
$$
\begin{aligned}
P(\text{lose}) &= P(\text{leading digit} = 1,2,3) \\
&= \log 10\,(1+1) + \log 10\,(1+\tfrac{1}{2}) + \log 10\,(1+\tfrac{1}{3}) \\
&= \log 10\,(2) + \log 10\,(3/2) + \log 10\,(4/3) \\
&= 0.602
\end{aligned}
$$

Therefore your rate of losing is much higher than your rate of winning despite how it seems at first!

3

Other applications

Chi-squared Distributions

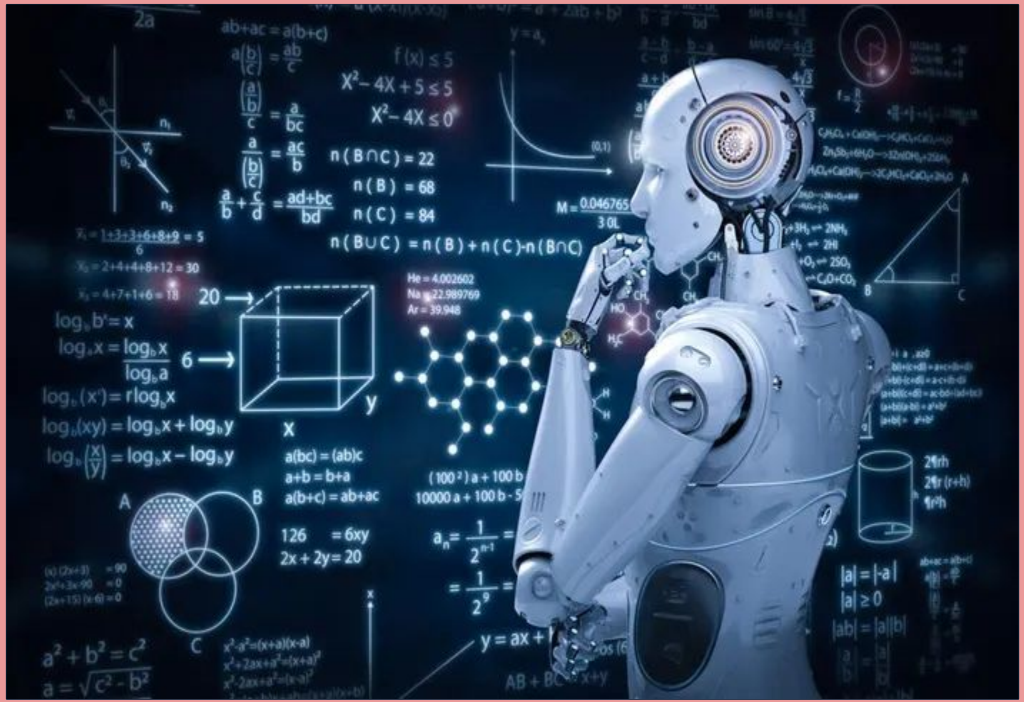


$$\mathbb{P}(\chi^2 \leq x) = \frac{1}{2^{\nu/2}\Gamma(\nu/2)} \int_0^x u^{(\nu/2)-1} e^{-u/2}\, du$$

Uses:

- Determine degree of **dependence** between various data
- **Detect fraud** and fake experimental results
- Find suitable data models to **simulate real-life events.**

Machine Learning



How is statistical knowledge applied?

- Anomaly detection
- **Bayesian inference** and conditional probabilities
- **Hypothesis testing** for comparison between different models
- Supervised learning with **regression analysis**

Thank you

SUNWAY MATHEMATICS CLUB