Introduction to probabilistic reasoning

Why do we use probability?

- “What is the probability it will rain today?” may help you decide what to wear today.

- "What is the probability that you will damage your car when driving?” will help the insurance company decide how much to charge you for insurance.

- "What is the probability of winning the lottery?” may help you decide whether to buy a ticket or not

Why do we use probability?

- Physics: Statistical mechanics and quantum physics
- Finance: Probability is fundamental to quantitative finance. Probability is used to model stock values and determine “fair” pricing for financial products.
- Medicine: In randomized clinical trials, participants are randomly allocated to receive therapy or placebo.

Why do we use probability?

- Space exploration: Probability is also used to come up with risk assessments to understand the various risks involved in carrying out a space exploration journey.

- Meteorology: With satellites data and other sophisticated tools, the seasonal probability of hurricanes and also the pathway of an ongoing hurricane can be predicted.

Why do we use probability?

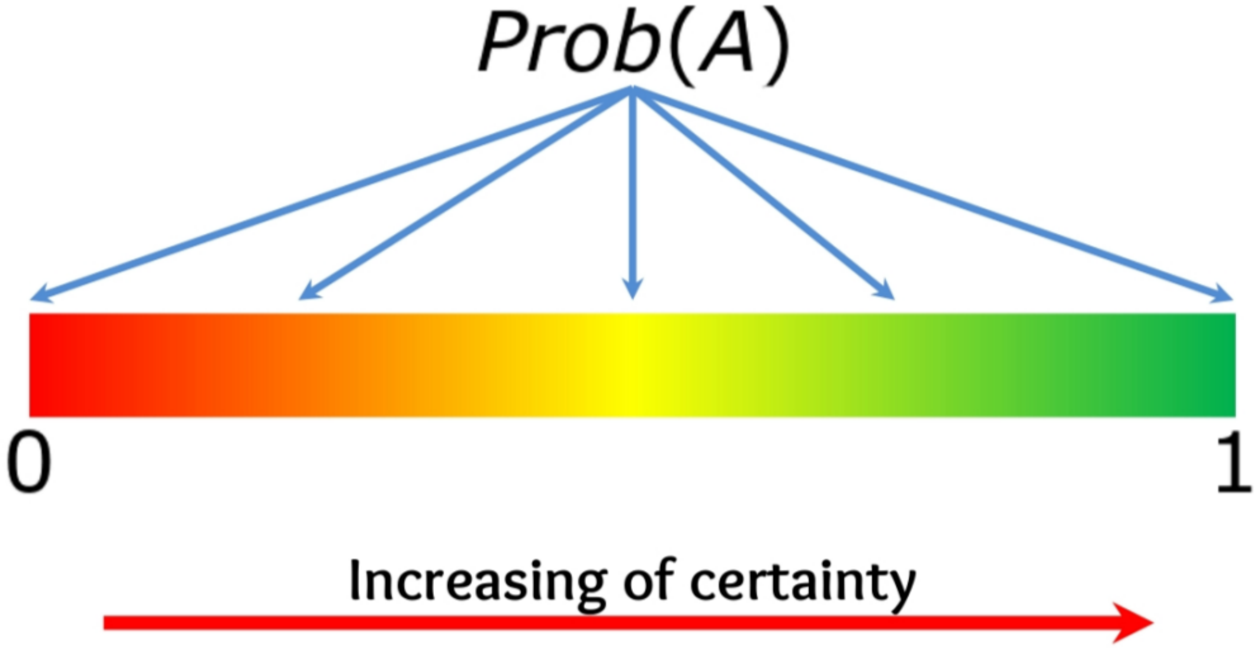
- Insurance: Insurers will use probability to draw up risk assessments in order to determine premium rates.

Why do we use probability?

- Machine learning
  - Class Membership Requires Predicting a Probability
  - Models Are Designed Using Probability e.g., Naive Bayes, logistic regression, graphical model.
  - Models Are Trained With Probabilistic Frameworks e.g., MLE, MAP, EM etc.
  - Models Are Tuned With a Probabilistic Framework
  - Models Are Evaluated With Probabilistic Measures

Probability scale

- Measurement of chance, or degree of uncertainty of a proposition.



Definition

- Event or Proposition: A statement that can be true or false.
- Prob(Event)
- Expressed as a fraction, decimal number or percentage

$$Prob(Rain) = \frac{1}{20} = 0.05 = 5\%$$

Probability scale

- $0 \leq \text{Prob(Event)} \leq 1$

- $\text{Prob(Event)} = 0 \leftrightarrow$ event can’t happen

- $\text{Prob(Event)} = 1 \leftrightarrow$ event must happen

How to measure probability?

1. Classical or theoretical probability: it is the likelihood of something happening based on all possible outcomes.

2. Objective (Frequentist): The probability is calculated from experiment.

3. Subjective (Bayesian): The probability is based on personal belief which can be adjusted by data.

Classical

- All outcomes are equally likely

- 10 possible outcomes <--> P (each outcome) = 1/10

- General case

$$Prob(Event) = \frac{Number\ of\ outcomes\ in\ which\ event\ occurs}{Number\ of\ \text{all}\ possible\ outcomes}$$

Frequentist

- All event are not equally likely
- Conduct experiment a **large** number of times
- Prob $\leftrightarrow$ proportion of time Event occurs.

$$Prob(Event) = \frac{Number\ of\ times\ event\ occurs}{Total\ number\ of\ times\ experiment\ done}$$

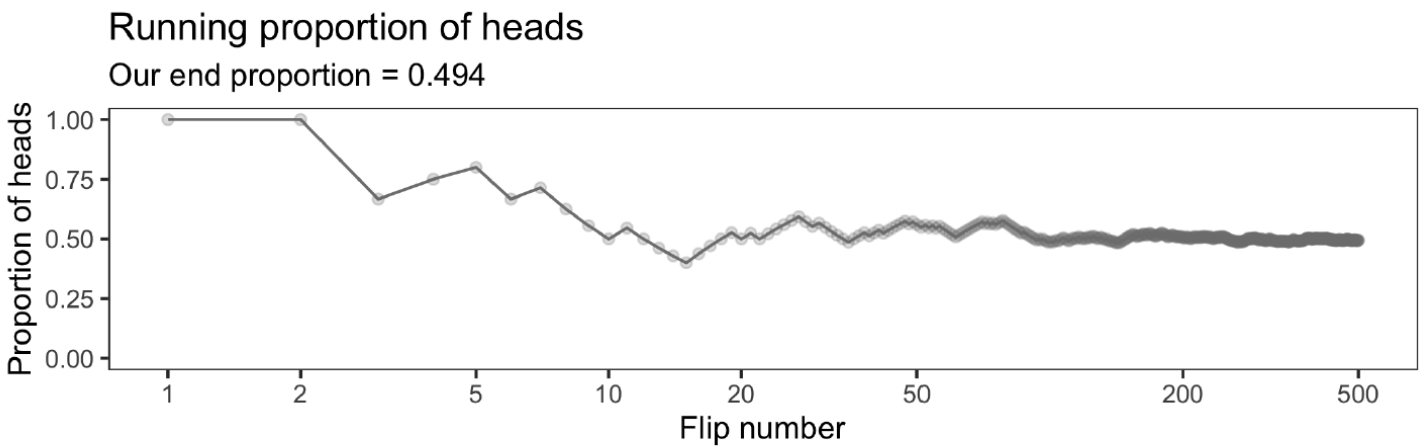
Frequentist

- The frequentist interpretation states that probability is essentially the long-term relative frequency of an outcome.

$$Prob(A) = \lim_{N \to \infty} \frac{\#(A)}{N}$$

Frequentist

- The Law of Large Numbers
- For example, to find the probability of getting a "head" when throwing a coin, one can repeat the experiment many times, as illustrated below:



Frequentist

- Drawback of the frequency approach:
  - Probability in this framework is only possible when the outcome can be repeated.
  - For example, what is the probability that Thailand will win the next FIFA world cup?

- In contrast, the subjectivist view of probability is based on one’s belief.

Subjective/Bayesian

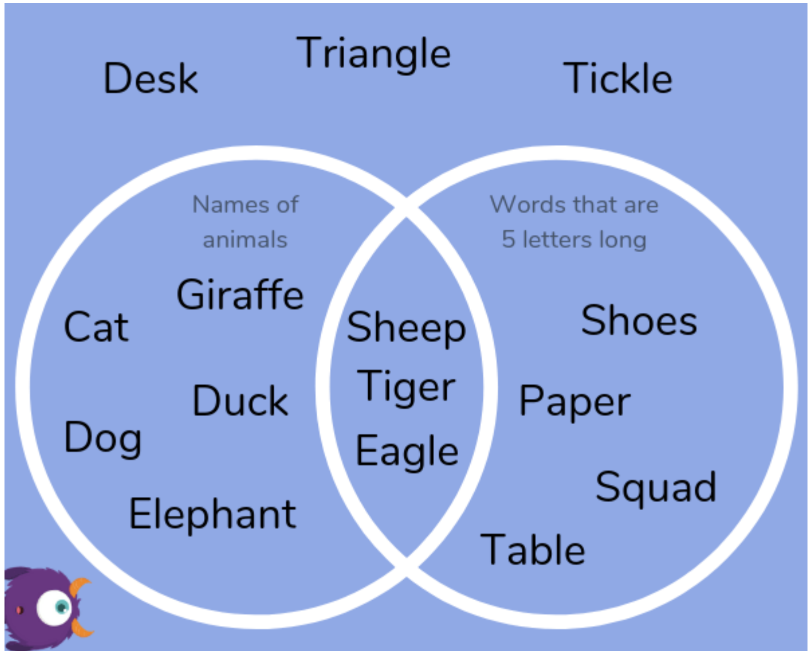
- Personal belief
- Consider Prob(Liverpool wins next game). What is your value?

How much would you pay for a bet which gives you 1 Pound if Liverpool wins next game?

Venn Diagrams

- Venn diagrams are a way to display events

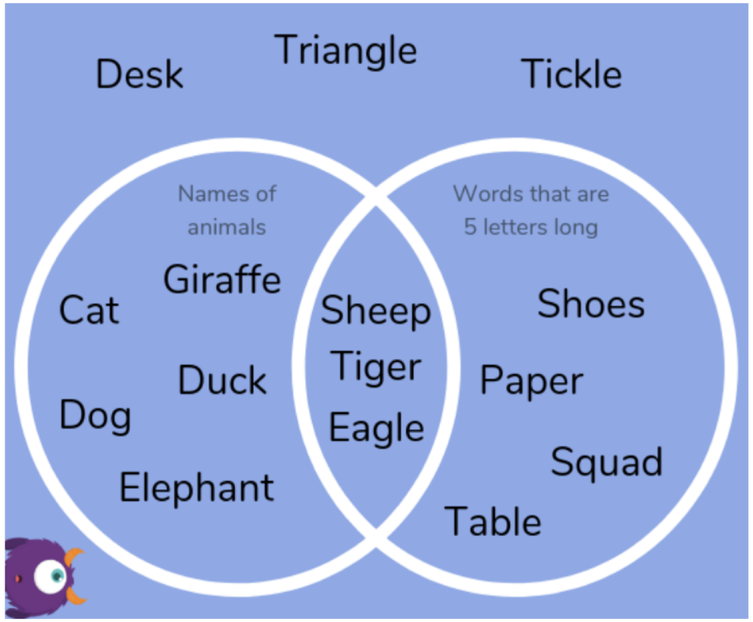
- Example: A collection of words
  - Event A: A word is the name of an animal
  - Event B: A word is 5 letters long



Joint probability

- The probability of events occurring together.

$$Prob(A\ and\ B) = Prob(A \cap B) = Prob(A, B) = Prob(B, A)$$



The Principle of Inclusion-Exclusion

- Consider the Venn Diagram. If you add n(A) and n(B), you count the intersection part twice. You should never count an element twice to find out how many you have. If you have counted it twice, there is an easy way to correct for that: subtract it once.

$$
\begin{aligned}
Prob(A\ or\ B) &= Prob(A \cup B) \\
&= Prob(A) + Prob(B) - Prob(A, B)
\end{aligned}
$$

Example

At the start of the flu season, a doctor examines 50 patients over two days. 30 have a headache, 24 have a cold, 12 have neither. Some patients have both symptoms.

- What is the probability that a random patient has both symptoms?

Example

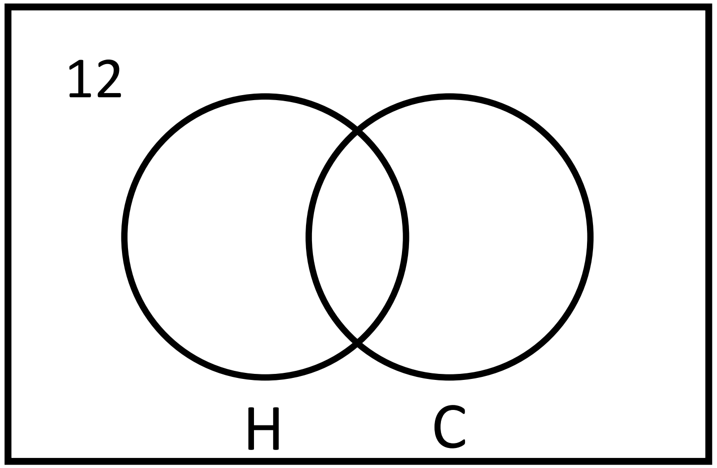
$$Prob(H\ or\ C) = Prob(H) + Prob(C) - Prob(\mathrm{H}, C)$$

$$\frac{(50-12)}{50} = \frac{30}{50} + \frac{24}{50} - Prob(H, C)$$

H = a patient has a headache

C = a patient has a cold

$$\therefore Prob(HandC) = \frac{16}{50}$$



$$Prob(H\ , \sim C) = \frac{30}{50} - \frac{16}{50} = \frac{14}{50}$$

$$Prob(\sim H, C) = \frac{24}{50} - \frac{16}{50} = \frac{8}{50}$$

Independent Events

- Independent events are events where the occurrence of one event has no effect on the other event.

- Dependent events are events that have different outcomes depending on what has already happened. The outcome of one event depends on the outcome of another event.

Independent Events

- For example, if we flip a coin twice, whether we get heads on the first flip has no effect on whether I get heads on the second flip.

$$Prob(A, B) = Prob(A)Prob(B) = Prob(B)Prob(A)$$

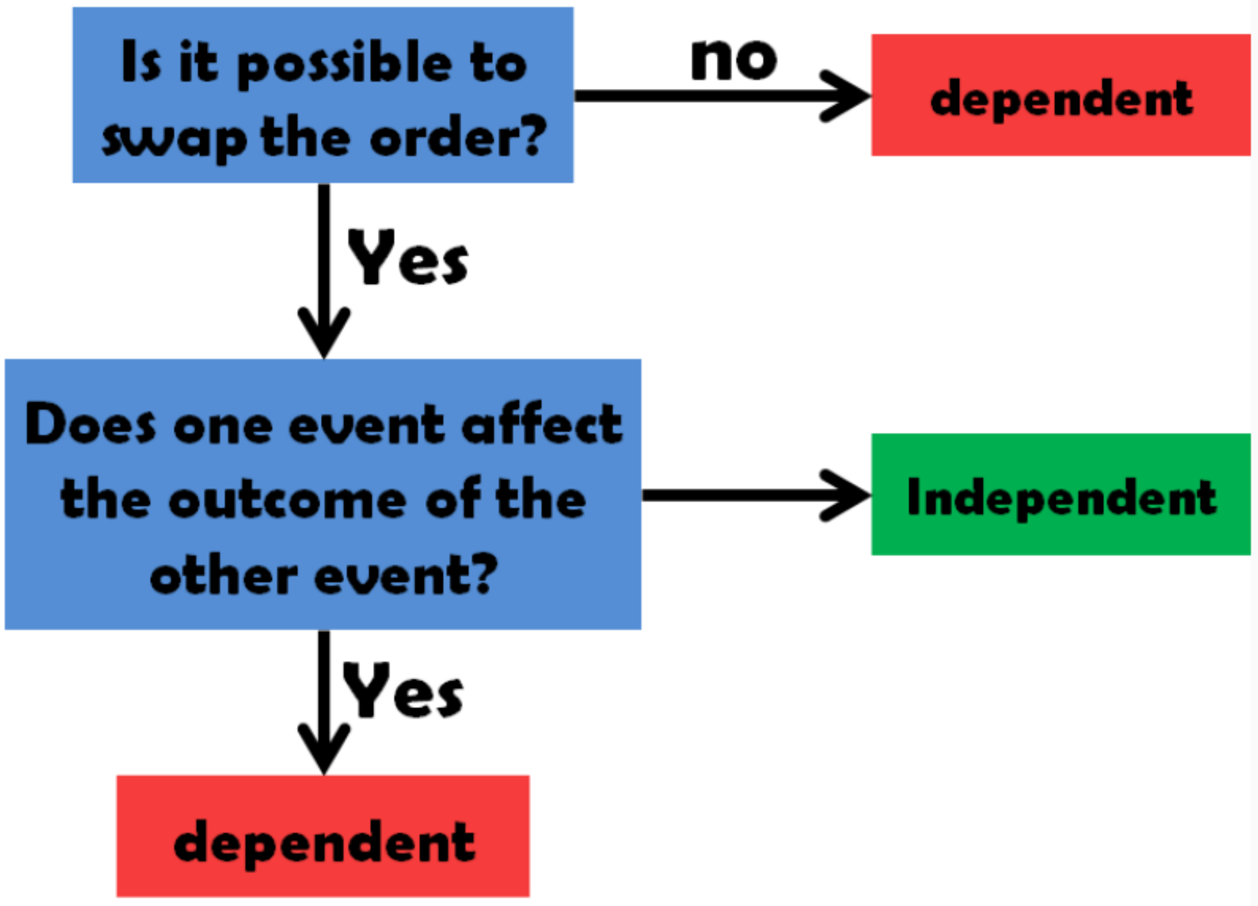
Independent Events

- From Flu example, are a headache and cold independent?

$$Prob(H,C) = \frac{16}{50} = \frac{8}{25}$$

$$Prob(H)Prob(C) = \frac{30}{50} \times \frac{24}{50} = \frac{36}{125}$$

Independent or not



Examples: Independent events

- Owning a dog and growing your own herb garden.
- Winning the lottery and running out of milk.
- Buying a lottery ticket and finding a penny on the floor

Examples: Dependent events

- Robbing a bank and going to jail.
- Not paying your power bill on time and having your power cut off.
- Catching a bus first and finding a good seat.
- Parking illegally and getting a parking ticket.
- Driving a car and getting in a traffic accident.

Mutually and Non-Mutually Exclusive Events

- **Mutually exclusive events** are events that can not happen at the same time. Examples include:
  - right and left hand turns
  - even and odd numbers on a die
  - winning and losing a game
  - running and walking.

Mutually and Non-Mutually Exclusive Events

- **Non-mutually exclusive** events are events that can happen at the same time. Examples include:
  - driving and listening to the radio
  - even numbers and prime numbers on a die
  - losing a game and scoring
  - running and sweating.

- Non-mutually exclusive events can make calculating probability more complex.

Mutually exclusive events

- If A and B are mutually exclusive then

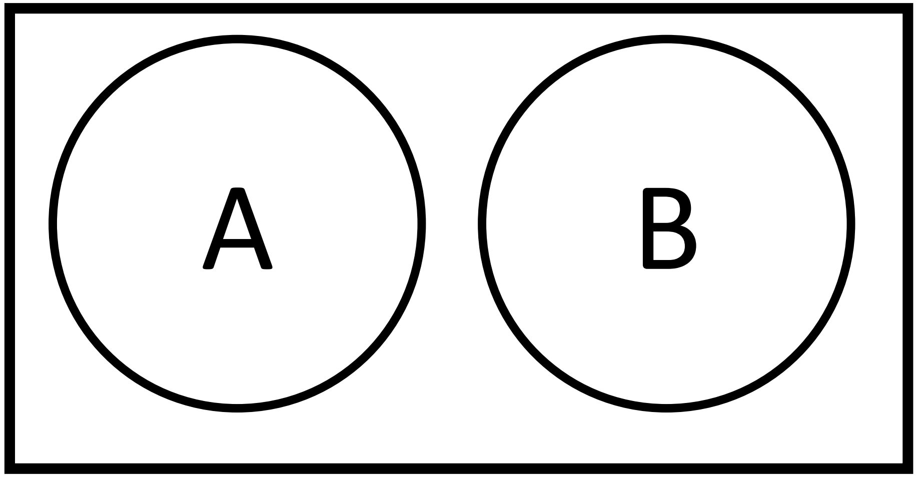
$$Prob(A\ or\ B) = Prob(A) + Prob(B)$$

because

$$Prob(A\,,B) = 0$$

Mutually exclusive events

- Example:
  - Event A: You are at university
  - Event B: You are bowling



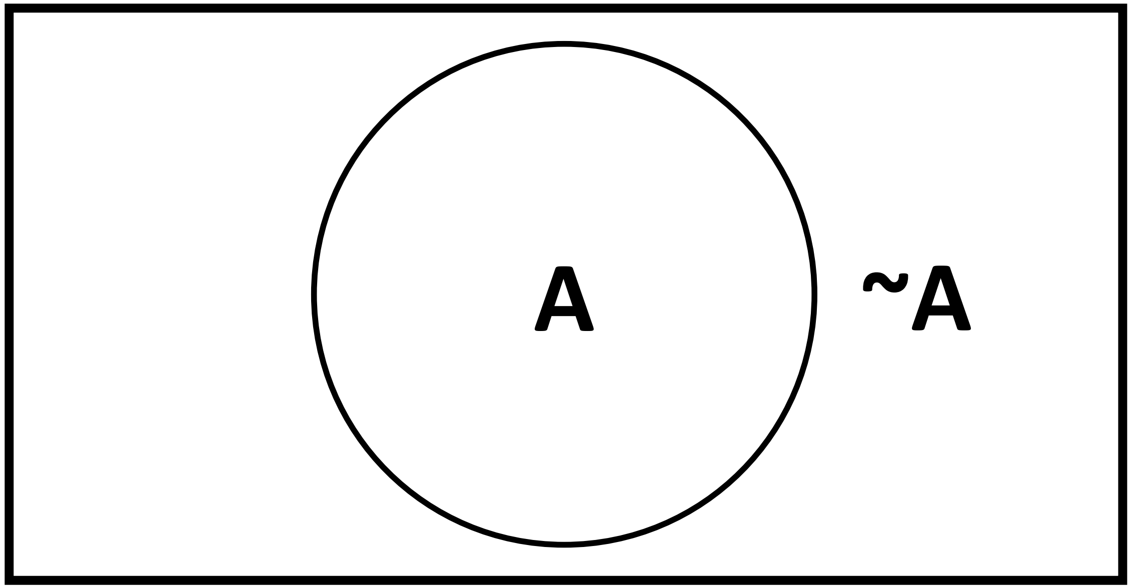
$$Prob(A, B) = 0$$

Comparison

Basis for Comparison	Mutually Exclusive Events	Independent Events
Meaning	Two events are said to be mutually exclusive, when their occurrence is not simultaneous.	Two events are said to be independent, when the occurrence of one event cannot control the occurrence of other.
Influence	Occurrence of one event will result in the **non-occurrence** of the other.	Occurrence of one event will have no influence on the occurrence of the other.
Mathematical formula	Prob(A and B) = 0	Prob(A and B) = Prob(A) Prob(B)
Sets in Venn diagram	Does not overlap	Overlaps

Complement events

$$Prob(A) + Prob(\sim A) = 1$$



Example: Electric lights

- A building has three rooms
- Each room has two separate electric lights
- Probability of 0.1 that a given light will have failed
- All light are independent

Find the probability that there is at least one room in which both lights have failed.

Solution

- For a given light, the probability that it has failed is 0.1.

- For a given room, the probability that both lights have failed is $0.1 * 0.1 = 0.01$

- The probability that at least one light is working in a room is $1 – 0.01 = 0.99$

The probability that at least one light is working in every one of the three rooms (that is, in Room A and in Room B and in Room C) is

$$0.99 \times 0.99 \times 0.99 = 0.970299$$

The probability that there is at least one room in which both lights have failed (that is the probability that it is not true that there is at least one light working in every room) is

$$1 - 0.970209 = 0.029701$$

Conditional probability

- We already knew that if events A and B are independent

$$Prob(A, B) = Prob(A) \times Prob(B)$$

- What if A and B are depedent?
- If A is the first part of a compound event and B is the second which depends on A

$$Prob(A, B) = Prob(B|A) \times Prob(A)$$

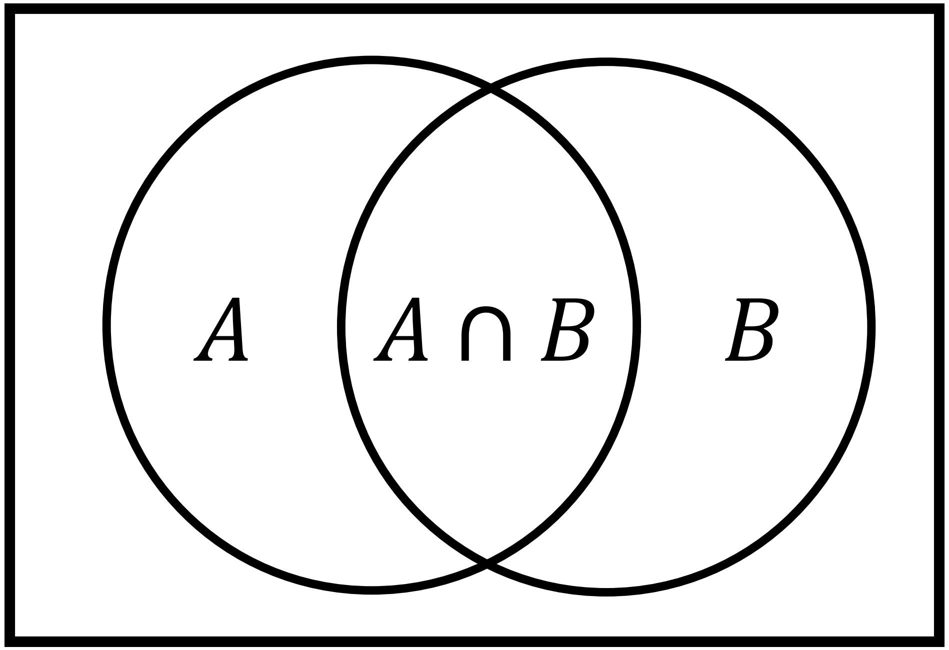
Conditional probability

- $Prob(B|A)$ is called conditional probability. This means "The probability of B given A has already occurred".
- We can also swap the order of conditional probability

$$Prob(A|B) = \frac{Prob(A,B)}{Prob(B)}$$

$$Prob(B|A) = \frac{Prob(A,B)}{Prob(A)}$$

Conditional probability



$$Prob(A|B) = \frac{Prob(A,B)}{Prob(B)}$$

$$Prob(B|A) = \frac{Prob(A,B)}{Prob(A)}$$

Conditional probability

Example: Utility companies – forecast periods of high demand

Prob(High demand | air temperature is below normal) = 0.6

Prob(High demand | air temperature is normal) = 0.2

Prob(High demand | air temperature is above normal) = 0.05

Conditional probability

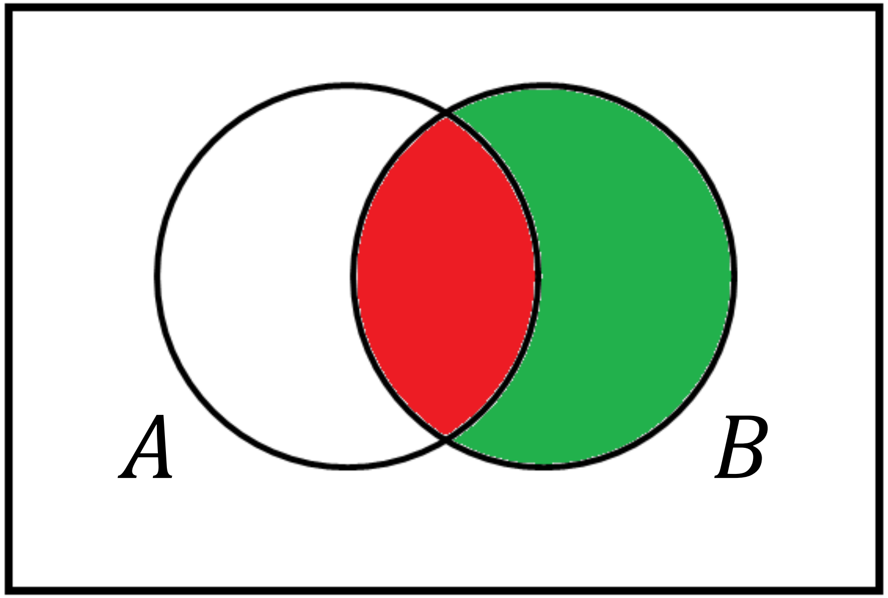
- If A and B are independent:

$$\begin{aligned} Prob(A,B) &= Prob(A)Prob(B) \\ Prob(A|B) &= Prob(A) \\ Prob(B|A) &= Prob(B) \end{aligned}$$

Conditional probability

- Complement events

$$Prob(A|B) + Prob(\sim A|B) = 1$$



Conditional independence

- If events $A_1, A_2, \dots, A_n$ are conditional independent conditioned on event B, then

$$
\begin{aligned}
Prob(A_1, A_2, \dots, A_n|B) &= Prob(A_1|)Prob(A_2|B) \dots Prob(A_n|B) \\
&= \prod_{i=1}^{n} Prob(A_i|B)
\end{aligned}
$$

Conditional probability

- Bayes' theorem

$$Prob(A, B) = Prob(A|B)Prob(B) = Prob(B|A)Prob(A)$$

$$Prob(A|B) = \frac{Prob(B|A)Prob(A)}{Prob(B)}$$

Marginalization

- Marginalisation in probability refers to “summing out” the probability of a random variable X given the joint probability distribution of X with other variable(s).
- It is a direct application of the law of total probability

$$Prob(A) = \sum_{i=1}^{N} Prob(A, B_i)$$

Marginalization

- Conditional marginalization

$$Prob(A|C) = \sum_{i=1}^{N} Prob(A, B_i|C)$$

Example: Whether and happiness

- Suppose we're interested in how the weather affects someone's happiness in the United Kingdom (UK).
- We can write this mathematically as $Prob(happiness|weather)$ i.e. what's the probability of someone's happiness level given the type of weather.
- The UK is made up of 3 countries: England, Scotland and Wales

Example: Whether and happiness

- Now it's feasible that people in Scotland are just generally happier than people in England.
- The problem is that people always have a nationality so we can't just get rid of it in the measurement.
- So what we're actually measuring is

$$P(\text{happiness}, \text{country} \mid \text{weather}),$$

i.e. we're looking at the happiness and the country at the same time.

Example: Whether and happiness

- Marginalisation tells us that we can calculate the quantity we want if we sum over all possibilities of countries

$$
\begin{aligned}
P(happiness|weather) \quad = \quad & P(happiness, country = England \mid weather) \\
& + P(happiness, country = Scotland \mid weather) \\
& + P(happiness, country = Wales \mid weather)
\end{aligned}
$$

Example: Soccer match

- If the Barcelona have an 80% chance of winning at home and a 60% chance of winning in away game (not real stats!), what is the probability that they win any random game in the season.

- If a random game has a 70% chance of being a home game and 30% chance of being a road game (10 games left in the schedule, 7 at home)?

Example: Soccer match

There are 3 events:

- W: win the match
- H: Home game
- A: Away game

Example: Soccer match

The probabilities:

- $Prob(W|H) = 0.8$
- $Prob(W|A) = 0.6$
- $Prob(H) = 0.7$
- $Prob(A) = 0.3$

Example: Soccer match

By marginalization,

$$
\begin{aligned}
Prob(W) &= Prob(W,A) + Prob(W,H) \\
&= Prob(W|A)Prob(A) + Prob(W|H)Prob(H) \\
& \quad 0.6 * 0.3 + 0.8 * 0.7 = 0.74
\end{aligned}
$$

Representing joint probability by table

- The relationship between events can be represented by table.
- Example: Defected electronic part from different factories

	Defect	Not-defect
Factory A	8	52
Factory B	12	68

Example: Defected electronic part from different factories

	Defect	Not-defect
Factory A	8	52
Factory B	12	68

$$
\begin{aligned}
Prob(\text{Defect} \mid \text{Factory A}) &= \frac{8}{8+52} = 0.13 \\
Prob(\text{Defect}\ ,\ \text{Factory A}) &= \frac{8}{8+52+68+12} = 0.05 \\
Prob(\text{Factory A}) &= \frac{8+52}{8+52+12+68} = 0.43 \\
Prob(\text{Factory A} \mid \text{Defect}) &= \frac{8}{8+12} = 0.4
\end{aligned}
$$

Example: Online sale of a music single

	< 30	30-50	50+	
Male	275	125	25	425
Female	325	175	75	575
	600	300	100	1000

$$Prob(Male\,, < 30) = \frac{275}{1000}$$

$$\begin{aligned} Prob(Male) &= Prob(Male\,, < 30) + Prob(Male\,, 30-50) + Prob(Male\,, 50+) \\ &= 0.275 + 0.125 + 0.025 = 0.425 \end{aligned}$$

$$\begin{aligned} Prob(Female) &= Prob(Female\,, < 30) + Prob(Female\,, 30-50) + Prob(Female\,, 50+) \\ &= 0.325 + 0.175 + 0.075 = 0.575 \end{aligned}$$

Example: Online sale of a music single

- The age distribution of the customers

$$
\begin{aligned}
Prob(<30) &= Prob(Male\ and < 30) + Prob(Female\ and < 30) \\
&= 0.275 + 0.325 = 0.6
\end{aligned}
$$

$$
\begin{aligned}
Prob(30-50) &= Prob(Male\ and\ 30-50) + Prob(Female\ and\ 30-50) \\
&= 0.125 + 0.175 = 0.3
\end{aligned}
$$

$$
\begin{aligned}
Prob(50+) &= Prob(Male\ and\ 50+) + Prob(Female\ and\ 50+) \\
&= 0.025 + 0.075 = 0.1
\end{aligned}
$$

Example: Online sale of a music single

- Conditional probability

$$Prob(Male|30-50) = \frac{Prob(Male\ and\ 30-50)}{Prob(30-50)} = \frac{0.125}{0.3} = 0.4167$$

$$\begin{aligned} Prob(Female|30-50) &= \frac{Prob(Female\ and\ 30-50)}{Prob(30-50)} = \frac{0.175}{0.3} = 0.5833 \\ &= 1 - Prob(Male\ |30-50) \end{aligned}$$

Example: Online sale of a music single

- Conditional probability

$$Prob(<30|male) = \frac{Prob(Male\ and < 30)}{Prob(Male)} = \frac{0.275}{0.425} = 0.6471$$

$$Prob(30-50|male) = \frac{Prob(Male\ and\ 30-50)}{Prob(Male)} = \frac{0.125}{0.425} = 0.2941$$

$$\begin{aligned} Prob(50+|Male) &= 1 - Prob(<30|Male) - Prob(30-50|Male) \\ &= 1 - 0.6471 - 0.2941 = 0.588 \end{aligned}$$

Tree diagram

- Experiment with multiple outcomes

- Represent each experiment by a circle

- Branches from it represent outcomes

- Each outcome has a probability associated with it

Tree diagram

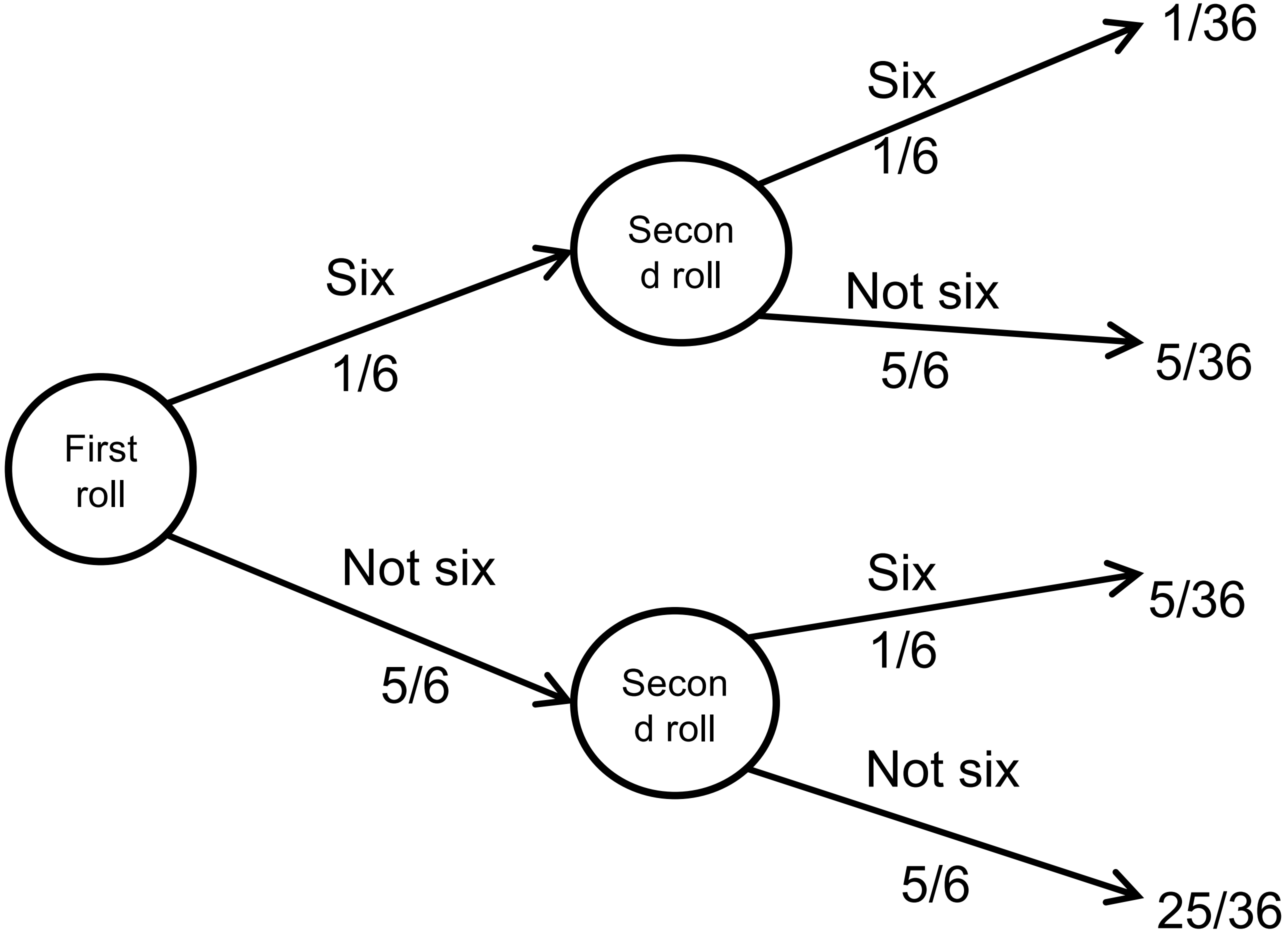
- Consider the probability of throwing two consecutive 6’s on a die.

$$Prob(six\ and\ six) = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$$

Tree diagram

- **Mutually Exclusive:**
No two outcomes can happen simultaneously

- **Collectively Exhaustive:**
Set of all possible outcomes represents the entire range of possible outcomes
Sum of probabilities equals ONE



Gambler's fallacy

- Believing that a random outcome is affected by previous outcomes, or believing that sequences of random events — such as rolling dice, and pulling slot machines — have memory.
- In sport, we also have "hot hand phenomenon" that a person who experiences a successful outcome has a greater chance of success in further attempts.

Gambler's fallacy

- In investment, investors tend to hold onto stocks that have depreciated and sell stocks that have appreciated. They call this a "general disposition to sell winners too early and hold losers too long."

Gambler’s fallacy

- When an institution fails to recognize the statistical independence of random events, unrelated events can be identified as causes in search for an explanation.
- The Gambler's Fallacy can lead to suboptimal decision-making.
- This fallacy can be problematic when two events are **not** causally related but we think they are.
- For example, storm clouds and rain are a good example for causal relation.

Gambler’s fallacy

Why it happens

- One of the reasons is that we don’t like randomness. So, we try to rationalize random events to create an explanation and make them seem predictable.
- So, when a random event occurs or is set to happen, we try to rationalize it by finding patterns or indications in the history of events similar to it-- even when they aren’t actually related.

Gambler’s fallacy

Why it happens

- In what’s known as the “law of small numbers,” we often take small samples of information to represent, or speak for, the larger population from which they are drawn. → “representativeness heuristic”
- This heuristic are mental shortcuts our brains use to help us make decisions quickly.

Gambler's fallacy

Why it happens

- We often choose past experiences that we want future events to be similar to, or that we think should be representative of an ideal outcome.
- A gambler may take a few successful turns at the slot machine to represent a longer winning streak that will continue (as it has sometimes done in the past), or conversely, to assume there will be a loss which will even out their wins

Gambler’s fallacy

Why it happens

- We think of a chance as a fair process or “self-correcting process” rather than a random one.
- This is to say, we think that chance aims at a fair and balanced equilibrium.
- Deviations away from this equilibrium are restored by an opposing outcome as a chance process unfolds.

Gambler’s fallacy

- Consider an examination, a student who thinks they have circled too many “A” options in a row on their multiple choice exam, and so they select a “C” to break the suspicious pattern.

- We do this all the time in both our personal and professional lives.

Gambler's fallacy

To counter the effect of this cognitive bias,

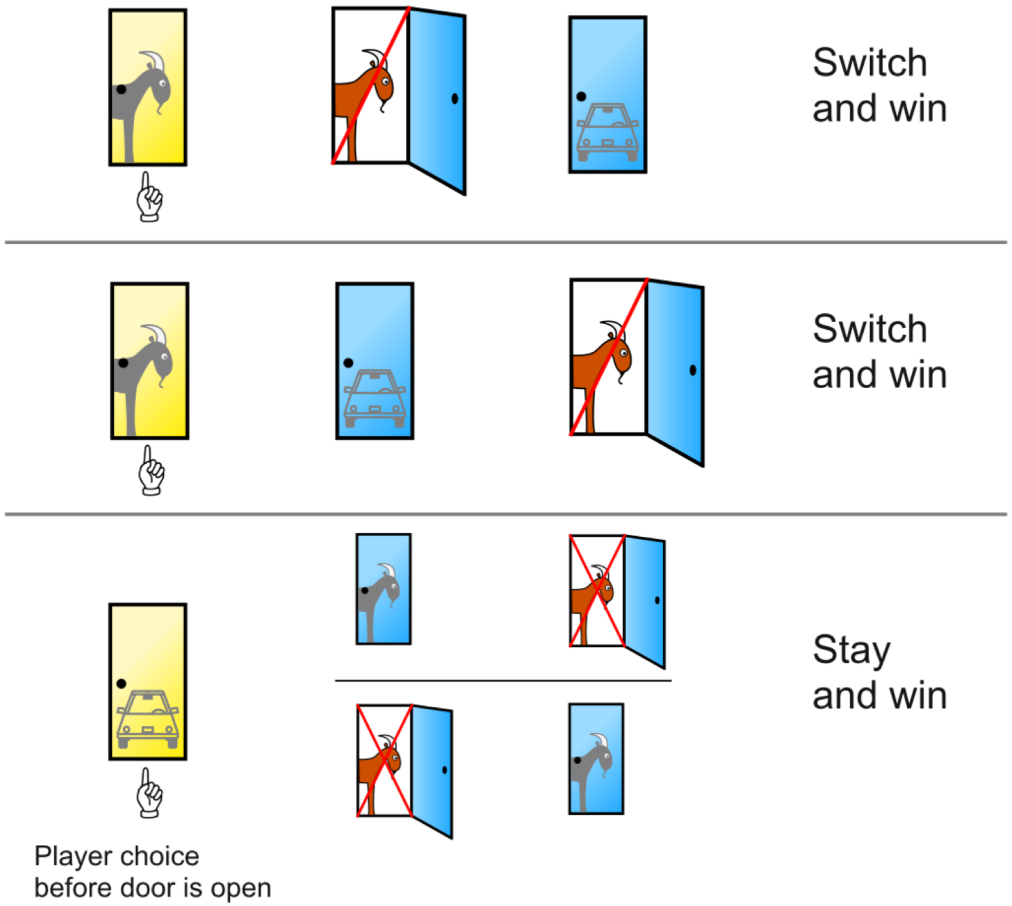
- we need to recognize the causal independence of the events in question.

- Thinking through the actual process by which an event occurs.

Example: The Monty Hall Problem

- The game is simple. Pick a door. Two of the doors have goats behind them, one has a car. After you pick your door, you will be shown one of the doors that had a goat behind them. Your choice is to stick or switch to the other closed door.

Example: The Monty Hall Problem



Example: The Monty Hall Problem

- To explain the analysis, let there be three doors i.e., yellow, blue green.
- Let assume that contestant choose yellow door
- We can fill in the probability table as following

Car is behind	Host open a door		
	Yellow	Blue	Green
Yellow	0	50%	50%
Blue	0	0	100%
Green	0	100%	0

Example: The Monty Hall Problem

- Host then opens blue door. We can then compute the probability

$$Prob(Car\ is\ behind\ yellow\ door \mid Host\ opens\ blue\ door) = \frac{50}{50+100} = 0.33$$

$$Prob(Car\ is\ behind\ green\ door \mid Host\ opens\ blue\ door) = \frac{100}{50+100} = 0.67$$

Car is behind	Host open a door		
	Yellow	Blue	Green
Yellow	0	50%	50%
Blue	0	0	100%
Green	0	100%	0

Expected Value

- Lets play the spinner. How much would you expect to win every time you spin it? Why do you think this might be?

Expected Value

- So for the spinner, what we expect to win won't be \$1 or \$2 or \$5 or \$20. It will be the average.
- However, could you calculate the average by saying:

$$\frac{1+2+5+20}{4} = \frac{28}{4} = 7$$

- Why would this not be correct the average?

Expected Value

- All outcomes are not equally likely.
- We need a way of giving more weight to the $1 prize and less weight to the $20 prize.
- “What is the expected outcome in random event”
- The question of expectations in games of chance require us to look at the outcomes and their probabilities.

Random variables

- A variable whose numerical values depend on outcomes of a random phenomenon.
  - Discrete random variable
  - Continuous random variable

Experiment or Survey	Random variable	Sample space
Tossing a coin	Coin face	Head(1) or Toe(0)
Rolling two dice	Sum of value	2,3,...,12
Measuring temperature	Air temperature	-50<t<200
Operating a coffee shop	Number of customers in a week	0,1,2,3,4,...

Random variables

Example: rolling a dice

- Random variable X – upper case X

  X = outcome of a roll of dice

- Observation x – lower case X
- Event is X = x
- We need to compute $Prob(X = x)$

x	$P(X = x)$
1	$1/6$
2	$1/6$
3	$1/6$
4	$1/6$
5	$1/6$
6	$1/6$
sum	1

Probability Distribution/Density Functions (PDF)

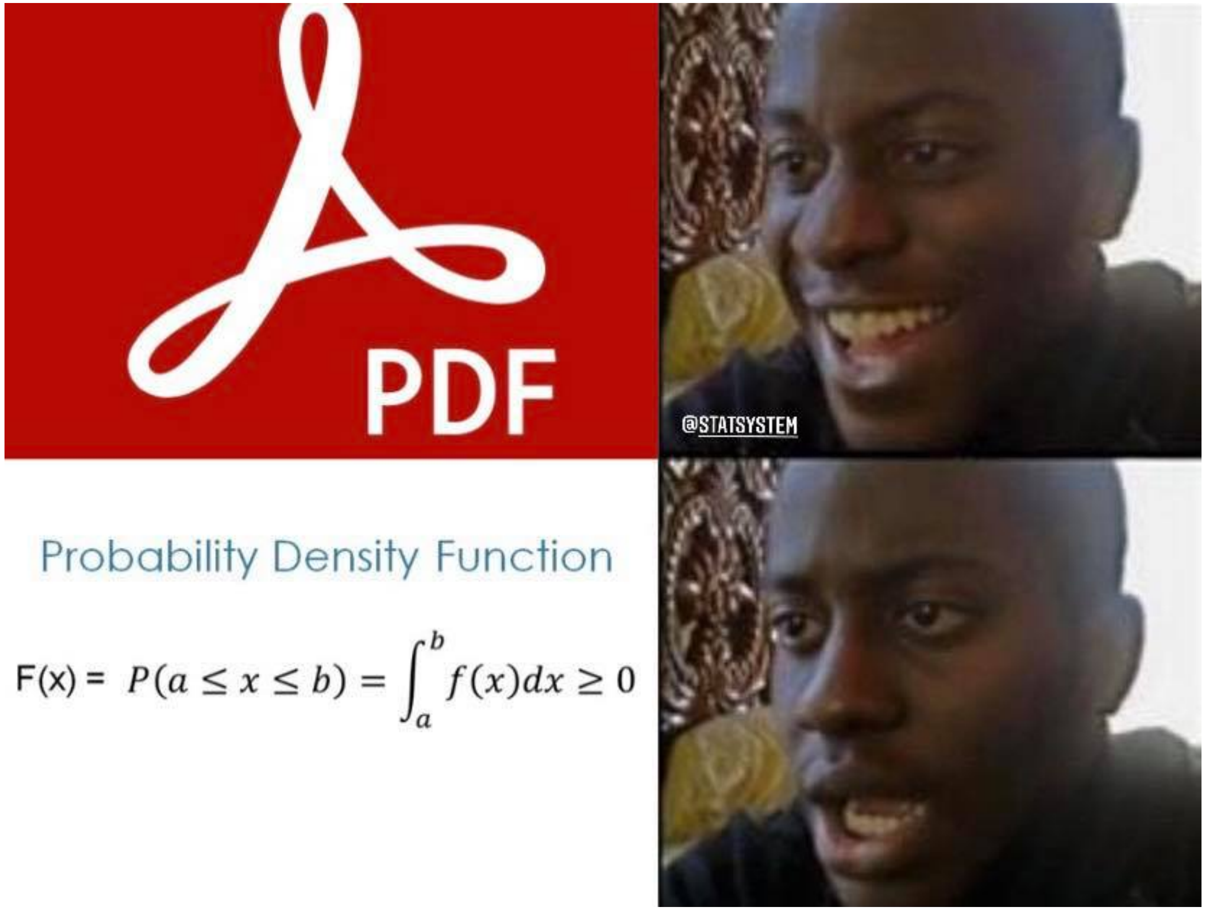
Example: Tossing a coin

- The random variable X = outcome of a toss of a coin

- We can easily list all values X can take and the probability they occur

Prob(X = head) = 0.7   Prob(X = toe) = 0.3

Probability Density/Distribution Functions (PDF)



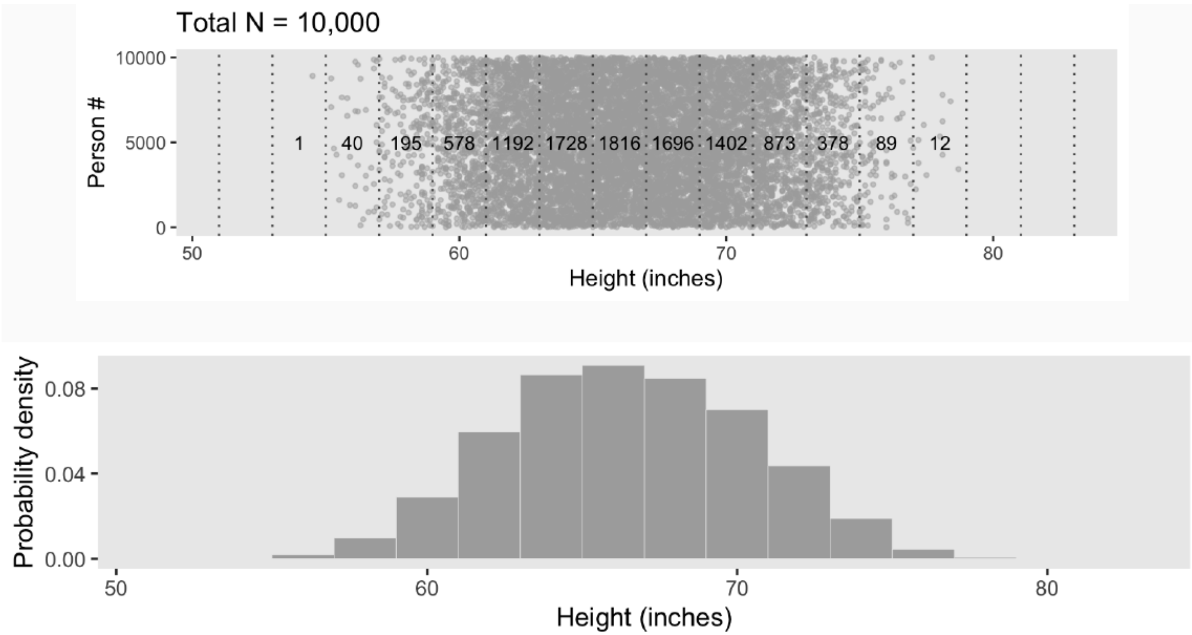
Probability Distribution/Density Functions (PDF)

- Some situations won't be that easy to see.
- Probability Density/Distribution Functions (PDF) is a function used for describes how the probabilities are distributed over the values of the random variable.
- If a random variable X is distributed with PDF f, we can state that

$$X \sim f$$

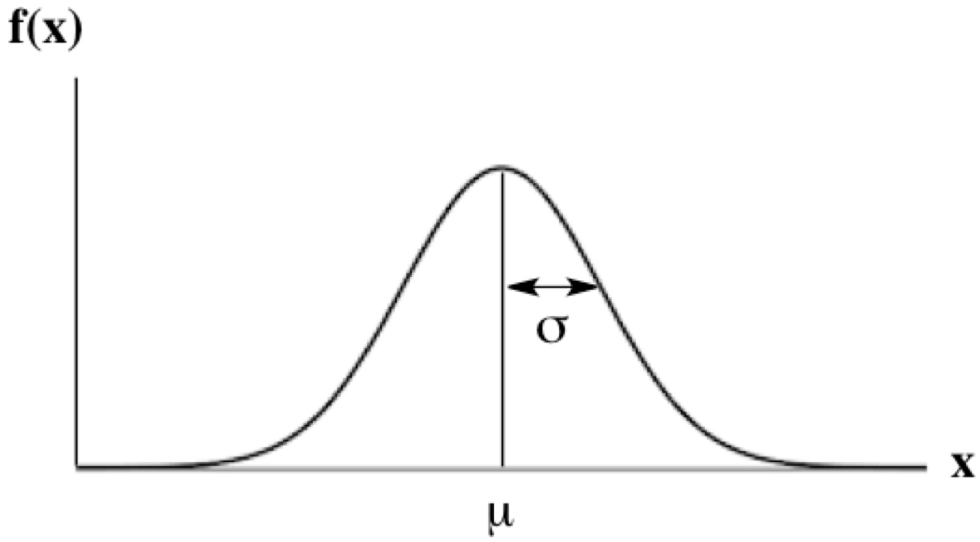
Probability Density/Distribution Functions (PDF)

- The probability distribution can also be obtained from repeated experiment.
- Example: height data



Probability Distribution/Density Functions (PDF)

Example: The famous normal or Gaussian distribution



$$f(x) = \frac{1}{\boxed{\sigma}\sqrt{2\pi}} e^{-(x-\boxed{\mu})^2/2\sigma^2}$$

$$e = 2.71828$$

Discrete distribution function

Discrete distribution function

- It is used with discrete random variable or when there are only a certain number of inputs.

- It can also be used with categorical data i.e., coin tossing.

- For example: When rolling a die, the inputs could be the number rolled on the die, and the output would be the probability of the roll.

Discrete distribution function

- Let a discrete random variable X has a distribution function f. That is, $X \sim f$

- The probability of the random variable $X = x$ is then

$$Prob(X = x) = f(x)$$

Expectation of a Discrete Random Variable

- Expectation of a discrete random variable is

$$E(X) = \mu = \sum x \, Prob(X = x) = \sum xf(x)$$

- This is called the weighted average and it gives us a formula for the expected value

Expectation of a Discrete Random Variable

- Expectation of a discrete random variable is

$$E(X) = \mu = \sum x\, Prob(X = x) = \sum xf(x)$$

Die-rolling experiment

$$E(X) = \sum x\ \ Prob(X = x) = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = 3.5$$

Variance of a Discrete Random Variable

- The variance of a discrete random variable is

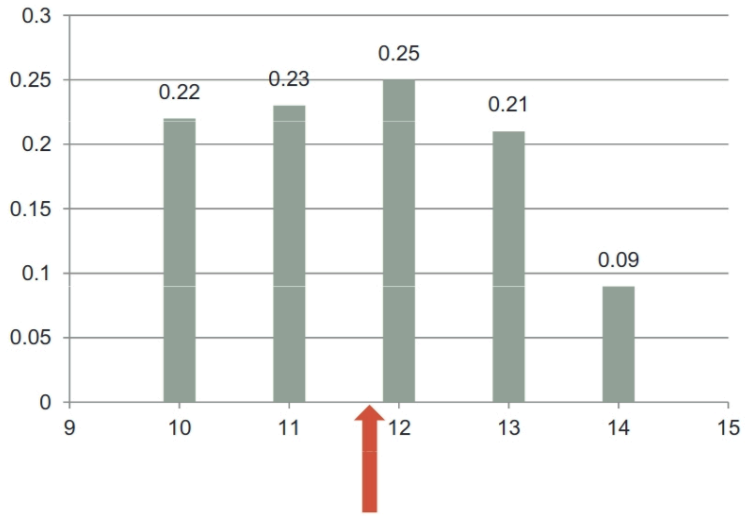
$$\begin{aligned} Var(X) &= \sigma^2 \\ &= E[(X-\mu)^2] \\ &= \sum (x-\mu)^2 Prob(X=x) \end{aligned}$$

Expected value doesn't tell the whole story!!

- Let consider the returns from investment
- Expected value is like center of mass

Return (% per year)	Probability
10	0.22
11	0.23
12	0.25
13	0.21
14	0.09

$$E(X) = 11.72$$



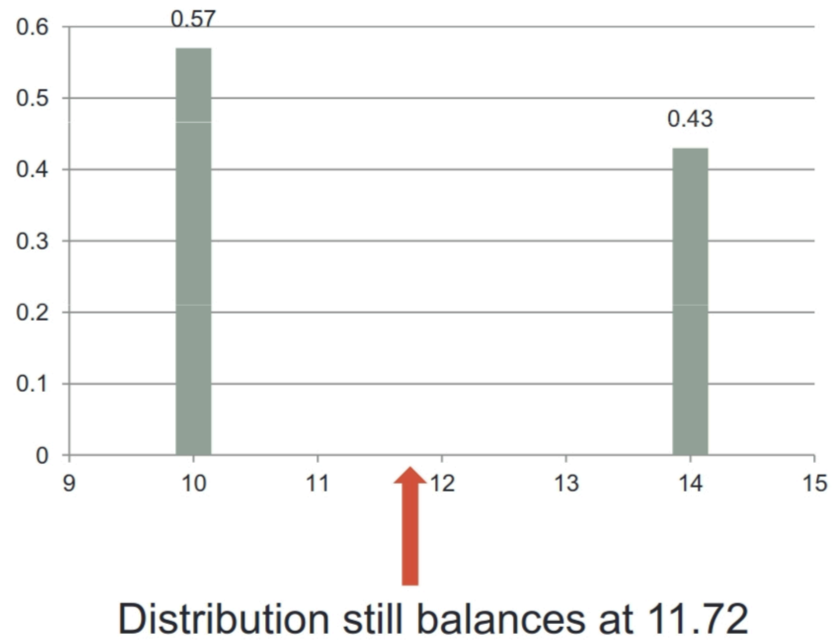
Distribution can be "balanced" at the mean

Expected value doesn't tell the whole story!!

- The two portfolios have the same average return

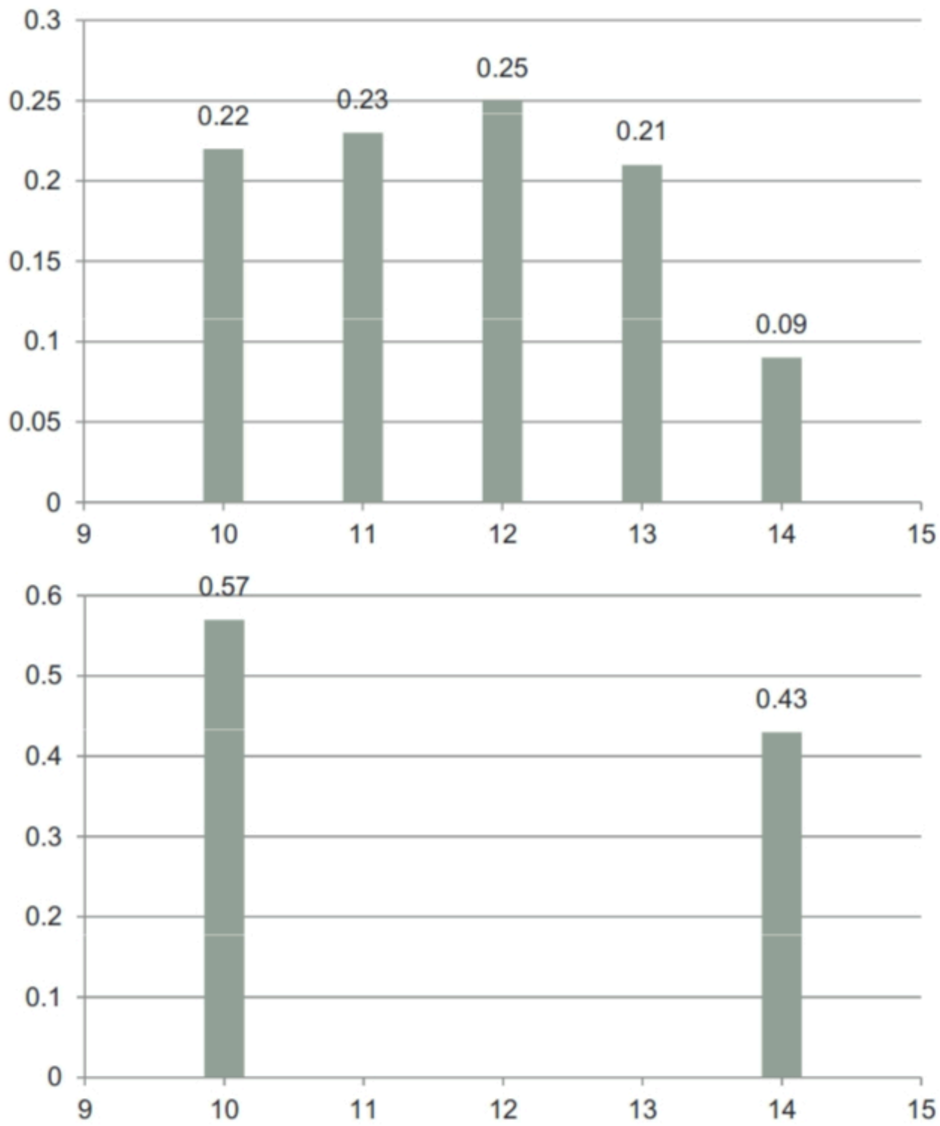
Return (% per year)	Probability
10	0.57
11	0
12	0
13	0
14	0.43

$$E(X) = 10 * 0.57 + 14 * 0.43 = 11.72$$



Expected value doe story!!

- How are they different?



Application in decision making

- Example: Concert project
  - We can calculate the profit (in million bath) for each scenario

Choice	Sale of ticket		
	100%	50%	10%
Project S	8	4	3
Project M	15	12	-1
Project L	25	10	-10

Application in decision making

- If we exactly know how much ticket can be sold, we can choose the project easily. For example, if 50% of tickets are sold, we will choose project M

Choice	Sale of ticket		
	100%	50%	10%
Project S	8	4	3
Project M	15	12	-1
Project L	25	10	-10

Application in decision making

- Unfortunately we do not know how much ticket will be sold.
- To make decision under uncertainty, we have to know the chance of each scenario

Choice	Sale of ticket		
	100%	50%	10%
Project S	8	4	3
Project M	15	12	-1
Project L	25	10	-10
Chance	0.3	0.4	0.5

Application in decision making

Two commonly used criteria for decision making under uncertainty are:

- Expected Monetary Value (EMV)
- Expected Opportunity Loss (EOL)

Both criteria can be computed in the same as we compute expected value for a random variable.

Application in decision making

Expected Monetary Value (EMV)

- P_j is the probability of scenario j where $\sum P_j = 1$
- C_{ij} is the profit of choice i when scenario j happens

EMV for choice i is then:

$$EMV(i) = \sum_j C_{ij} P_j$$

Application in decision making

- Expected Monetary Value (EMV)

Choice	Sale of ticket		
	100%	50%	10%
Project S	8	4	3
Project M	15	12	-1
Project L	25	10	-10
Chance	0.3	0.4	0.5

Select the choice having maximum EMV

Choice	Calculation	EMV
Project S	8(0.3)+4(0.4)+3(0.3)	4.9
Project M	15(0.3)+12(0.4)+(-1)(0.3)	9.0
Project L	25(0.3)+10(0.4)+(-10)(0.3)	8.5

Application in decision making

Expected Opportunity Loss (EOL)

- To compute EOL, we have to convert profit to opportunity loss by subtract the profit of that choice from maximum profit of that scenario

Choice	Sale of ticket		
	100% (Prob = 0.3)	50% (Prob = 0.4)	10% (Prob = 0.3)
Project S	25 – 8 = 17	12 – 4 = 8	3 – 3 = 0
Project M	25 – 15 = 10	12 – 12 = 0	3 – (-1) = 4
Project L	25 – 25 = 10	12 – 10 = 2	3 – (-10) = 13

Application in decision making

Expected Opportunity Loss (EOL)

- The computation of EOL is similar to EMV.

Select the choice having minimum EOL

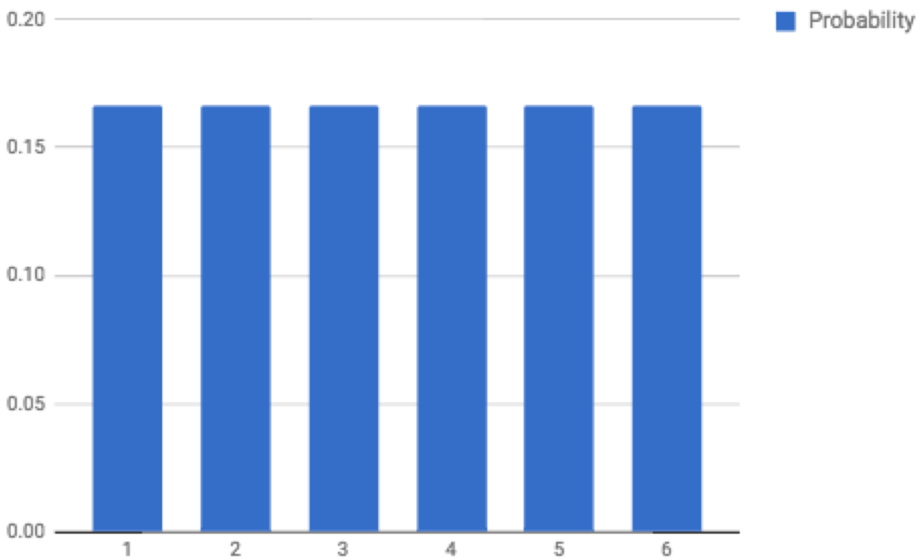
Choice	Calculation	EOL
Project S	17(0.3)+8(0.4)+0(0.3)	8.3
Project M	10(0.3)+0(0.4)+4(0.3)	4.2
Project L	0(0.3)+2(0.4)+13(0.3)	4.7

Commonly used discrete PDFs

- Uniform distribution
- Binomial distribution
- Geometric distribution
- Hyper-geometric distribution
- Negative binomial
- Poisson distribution

Uniform Distributions

- The easiest type of probability to deal with is the probability where all outcomes are the same.
- For example: Rolling a dice



Binomial Distribution

Example: There is a 70% chance that an allergy drug will be effective on a person suffering from the allergy symptoms that the drug treats.

What is the probability that, out of 10 random people, it is effective on 7 people.

Binomial Distribution

Bernoulli Trials

- It is a random experiment with exactly two possible outcomes, "success" and "failure", in which the probability of success is the same every time the experiment is conducted.
- The experiment is a random sample and the responses are independent.

Binomial Distribution

Since a Bernoulli trial has only two possible outcomes, it can be framed as some "yes or no" question:

- Is the product defected or not?
- Was the newborn child a girl? (human sex ratio)
- Does a voter select party A or party B?
- Does a respondent select to drink Coke or Pepsi?

Binomial Distribution

- Let $Prob(sucess) = p$
- $X =$ total number of successes out of n trials
- Probability distribution

$$Prob(X = r) = \binom{n}{r} p^r (1-p)^{n-r} \quad , \qquad r = 0,1, \ldots, n$$

- $X \sim Bin(n, p)$→Random variable X is distributed by Binomial distribution
- Mean and variance are

$$E(X) = np, \qquad Var(X) = np(1-p)$$

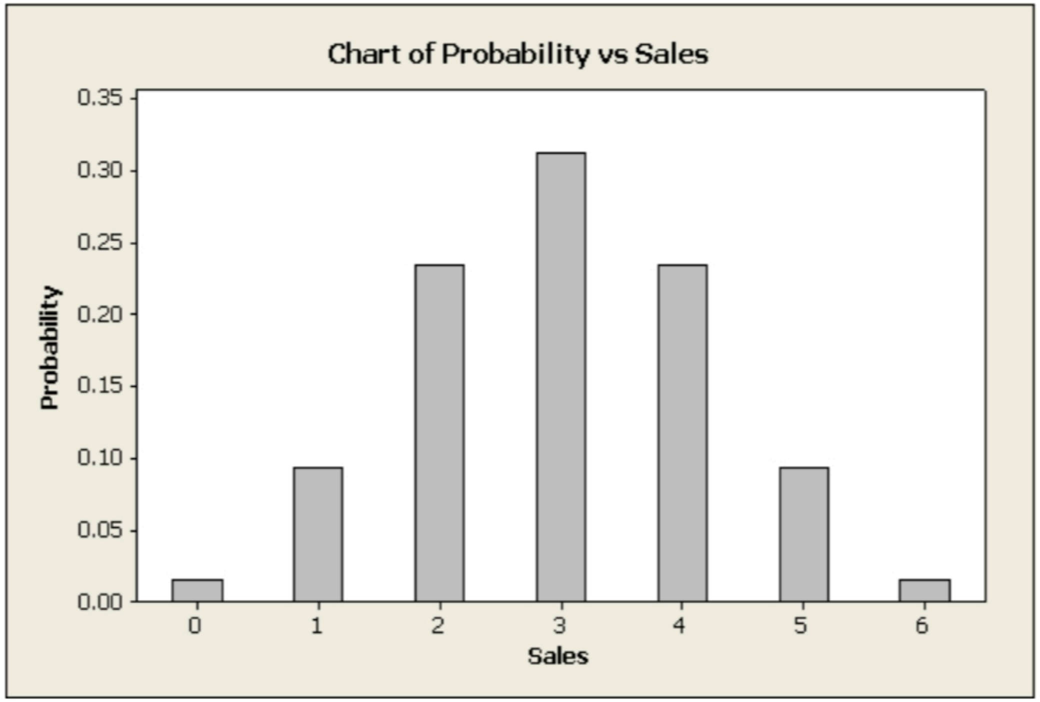
Binomial Distribution

Example: Number of sales

- A salesperson has a 50% chance of making a sale on a customer visit and she arranges 6 visits in a day. What are the probabilities of her making 0,1,2,3,4,5 and 6 sales?
- Let X =number of sales. Assuming the visits result in sales independently, $X \sim Bin(6,0.5)$

Binomial Distribution

Example: Number of sales



Geometric distribution

- Suppose we make an unlimited number of independent Bernoulli trial trials.
- The Geometric distribution tells us the probability of a certain number of failures to get the first success in k
- Let X be the number of trials up to and including the first success.

$$Prob(X = k) = (1 - p)^{k-1} p$$

Geometric distribution

- The probability of having the success on the n^th trial is $q^{n-1}p$. If you list these probabilities you'll have the sequence

$$p, qp, q^2p, q^3p, \ldots$$

This is a geometric series and it is the reason why we call it geometric distribution.

Geometric distribution

- Comparison with binomial distribution

Geometric Distribution	Binomial Distribution
The random variable, X, counts the number of trials required to obtain that first success.	The random variable, X, counts the number of successes in those trials.
The probability distribution function is given by $$(1-p)^{k-1}p$$	The probability distribution function is given by $$\binom{n}{k}(1-p)^{n-k}p^k$$

Geometric distribution

Example: Voting

- A polling employee wants to find out why people vote party A.
- They randomly select a person exiting a polling station, where the probability is p = .25 that a person voted party A.
- The distribution tells them how many people they will need to ask before finding a person who actually voted Independent.

Geometric distribution

Example: Safety engineering

- A safety engineer studies industrial accidents in a plant, suspecting that 40% are caused by employees not following instructions.
- To test the theory, accident reports are randomly selected until one is found (a "Success") that is caused by an employee failing to follow safety procedures .

The Hyper-geometric Distribution

- Like the Binomial Distribution, the Hyper-geometric Distribution is used when you are conducting multiple trials. We are also counting the number of "successes" and "failures."
- The main difference is, the trials are dependent on each other. The current trials affects the next trials.
- This is usually the case when the size of population is limited.

The Hypergeometric Distribution

Example: Driving license

- In a class of 25 students, 15 of them have a driver's license. If 5 students are randomly selected, how could we determine the probability that one of these students has their license? Two of them? What might the formula for the probability distribution look like?

The Hypergeometric Distribution

- $X =$ total number of successes out of r trials
- The probability of x successes out of r trials is then

$$Prob(X = x) = \frac{\binom{k}{x}\binom{N-k}{r-x}}{\binom{N}{r}}$$

Where

- N represents the total number of objects
- r represents the number of objects we are selecting and
- *K* represents the number successful objects
- x is less then or equal K

The Hypergeometric Distribution

From driving license example:

- N (total number of objects) = 25
- r (the number of objects we are selecting) = 1,2,…25
- *K* (the number successful objects) = 15

Negative binomial distribution

- Negative binomial distribution is a probability distribution of number of occurences of successes and failures in a sequence of independent trails before a specific number of success occurs.

- Let the fixed number of success be r. Experiment should be carried out until r successes are observed

Negative binomial distribution

- Let X be a random variable represent the total number of trials and we want r success trials.
- The probability that we will have r^{th} success at x^{th} trial is

$$Prob(X = x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}$$

Negative binomial distribution

Example: Oil survey

- An oil company conducts a geological study that indicates that an exploratory oil well should have a 20% chance of striking oil. What is the probability that the first strike comes on the third well drilled?
- So, x = 3, r = 1, and p = 0.2

$$Prob(X = 3) = \binom{3-1}{1-1}(1-p)^{3-1}p^{1} = 0.128$$

Negative binomial distribution

Example: Oil survey

- What is the probability that the third strike comes on the seventh well drilled?
- So, x = 7, r = 1, and p = 0.2

$$Prob(X = 7) = \binom{7-1}{3-1}(1-p)^{7-3}p^3 = 0.049$$

Poisson Distribution

- A discrete probability distribution for number of occurrence in a specified interval of time or space.
- It is applicable if
  - The probability of occurrence is the same for any two interval of equal length
  - The occurrence or non-occurrence from different interval is independent.

Poisson Distribution

Not only with time, we can also use Poisson distribution with space to find the probability for:

- the number of typos on a printed page. (an interval of space — the space being the printed page.)
- the number of cars passing through an intersection in one minute. (an interval of time — the time being one minute.)
- the number of customers at an ATM in 10-minute intervals.
- the number of students arriving during office hours.

Poisson Distribution

- Let X be a random variable representing the number of occurrence in the interval.
  - Ex. X = number of calls to an ISP in an hour
- Let λ be the mean number of occurrence in an interval.
- Probability of r occurrence is then

$$Prob(X = r) = \frac{\lambda^r e^{-\lambda}}{r!} \qquad , \qquad r = 0,1,2,\ldots$$

- $X \sim Po(\lambda)$
- Mean and variance are

$$E(X) = \lambda \qquad\qquad Var(X) = \lambda$$

Poisson Distribution

Example: ISP

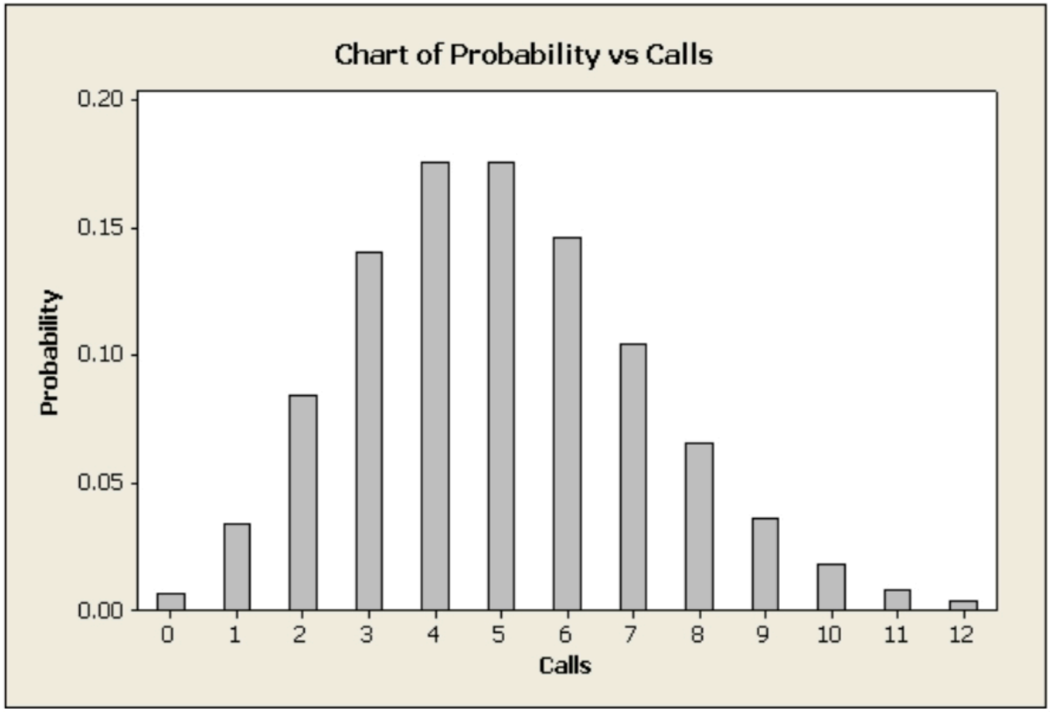
- An Internet service provider (ISP) has thousands of subscribers, but each one will call with a very small probability. The ISP knows that on average 5 calls will be made in one minute.
- Let X =number of calls made in a minute.
- Then $X \sim Po(5)$ and

$$Prob(X = r) = \frac{5^r e^{-5}}{r!} \quad , \qquad r = 0,1,2, \dots .$$

- For example

$$Prob(X = 4) = \frac{5^4 e^{-5}}{4!} = 0.1755$$

Poisson Distribution



Cumulative distribution function (CDF)

- Let X be a random variable with probability distribution f .
- Cumulative distribution function (CDF) is probability to get the value below a specified value

$$Prob(X \leq a) = F(a) = \sum_{x \leq a} f(x)$$

$$Prob(X > a) = 1 - Prob(X \leq a)$$

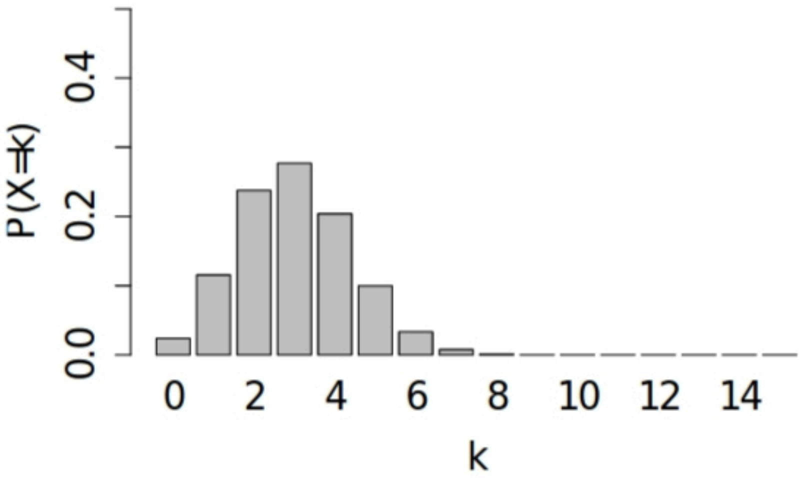
Cumulative distribution function (CDF)

Example: ISP

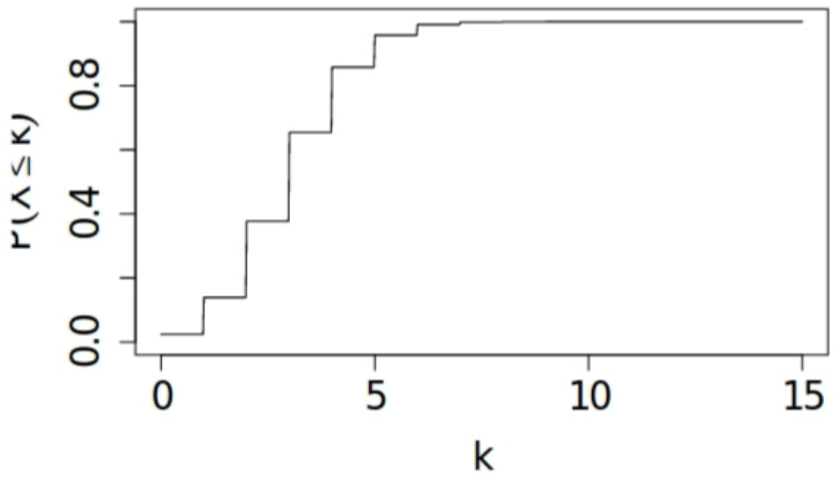
- From ISP example, we want to find the probability that number of customer call is less then or equal to 4

$$
\begin{aligned}
Prob(X \leq 4) = \quad & Prob(X = 0) + Prob(X = 1) + \\
& Prob(X = 2) + Prob(X = 3) + \\
& \qquad\qquad Prob(X = 4)
\end{aligned}
$$

Cumulative distribution function (CDF)



PDF



CDF

Cumulative distribution function (CDF)

- The probability that the value of random variable is in a range can be computed by

$$Prob(a \le X \le b) = F(b) - F(a)$$

Continuous density function

Probability Distribution/Density Functions (PDF)

Continuous functions:

- when there are an infinite number of inputs.
- **Continuous variables** are obtained by measuring, so there is no limitation on the input value.
- For example: When looking at baby weights, the inputs could be the weight and the output could be the probability that a baby is that weight.

Probability Density Functions (pdfs)

The key features of pdfs are

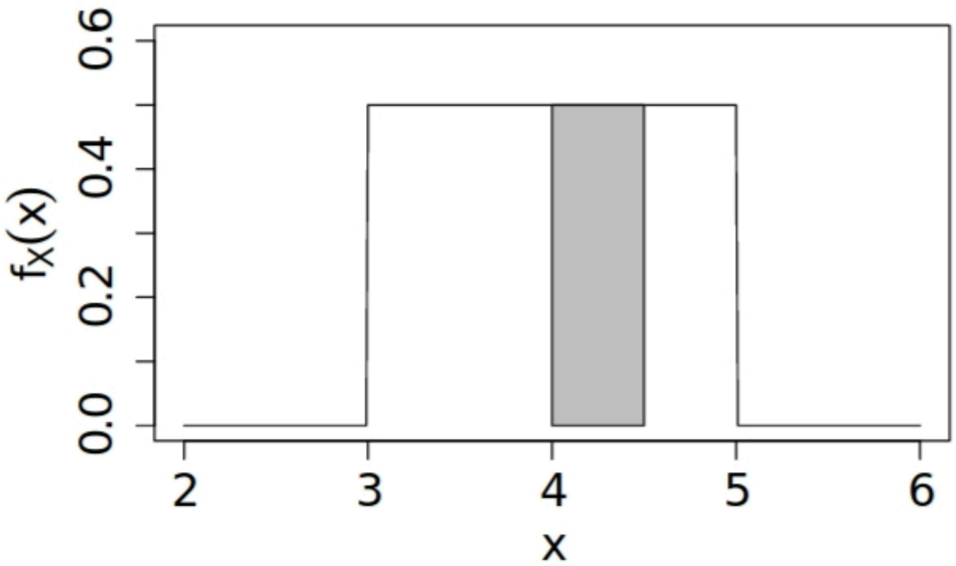
- pdfs never have negative values
- The area under a pdf is one
- Areas under the curve correspond to probabilities

$$Prob(a \leq X \leq b) = \int_a^b f(x)dx$$

- $Prob(X \leq x) = Prob(X < x)$ since $Prob(X = x) = 0$

Probability Density Functions (pdfs)

- Uniform pdf of bus waiting times. The shaded area indicates the probability $Prob(4 < X < 4.5)$



Expectation and variance

- Expectation

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

- $Variance$

$$Var(X) = \int_{-\infty}^{\infty} \left(x - E(X)\right)^2 f(x) dx$$

The Uniform Distribution

- Outcomes measured on a continuous scale.
- All outcomes equally likely.

The Exponential Distribution

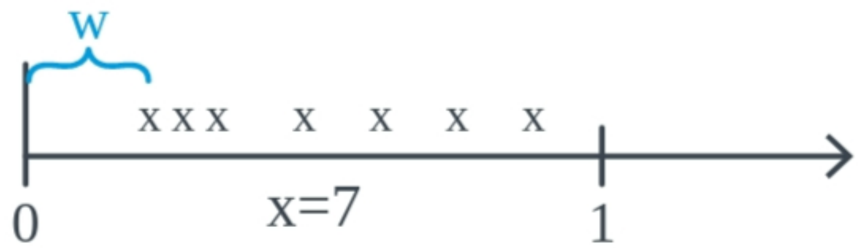
- Exponential distribution is used to describe the time between events in a Poisson process.
- In Poisson process events occur continuously and independently at a constant average rate.

The Exponential Distribution

- It can be used to model
  - lifetimes of products
  - times between “random” events
  - arrivals of orders
  - customers in a queueing system
- It has one (positive) parameter θ (rate parameter)

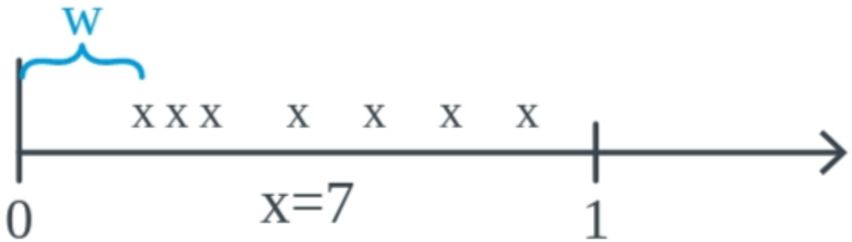
The Exponential Distribution

- Suppose X, following an (approximate) Poisson process, equals the number of customers arriving at a bank in an interval of length 1.
- If λ, the mean number of customers arriving in an interval of length 1, is 6, say, then we might observe something like this:



The Exponential Distribution

- we could alternatively be interested in the continuous random variable W, the waiting time until customer arrives.



- If λ equals the mean number of events in an interval, and θ equals the mean waiting time until the next customer arrives, then:

$$\theta = \frac{1}{\lambda}$$

The Exponential Distribution

General form of pdf

$$f(x) = \begin{cases} \dfrac{1}{\theta} e^{-x/\theta} & x \geq 0 \\ 0 & otherwise \end{cases}$$

and probabilities can be calculated using

$$Prob(X \leq x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x/\theta} & x > 0 \end{cases}$$

The Exponential Distribution

Example: Restaurant customer

- Customer arrive at a restaurant according to an approximate Poisson process at a mean rate of 30 customers per hour.
- What is the probability that the arrival time of the new customer is mote then 3 minutes?

The Exponential Distribution

- There are 30 customer in 60 minutes. So, $\theta = 2$
- Probability of the gap between customer being more than 1 minutes is

$$\begin{aligned} Prob(X \geq 3) &= 1 - Prob(X < 3) \\ &= 1 - \left(1 - e^{-3/2}\right) = e^{-3/2} \end{aligned}$$

Normal or Gaussian distribution

- The most commonly used A continuous probability distribution
- It is symmetric about the mean
- Data near the mean are more frequent in occurrence than data far from the mean.

Normal or Gaussian distribution

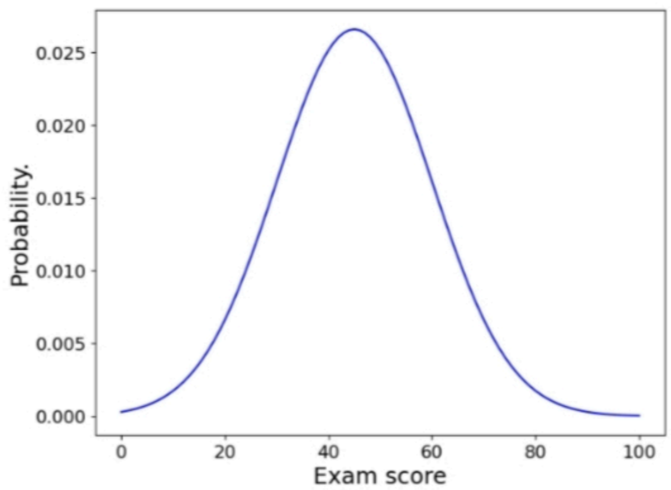
- Two parameters for the normal distribution are mean μ and standard deviation σ

$$f(x) = \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

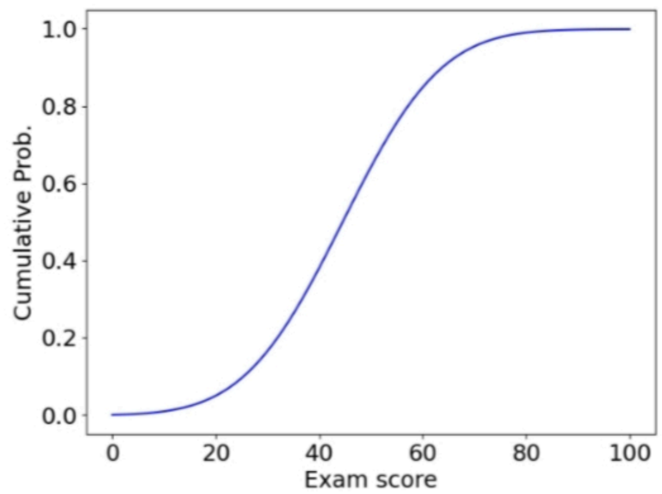
Normal or Gaussian distribution

Example: Examination score

- Examination score is distributed by normal distribution



Probability distribution

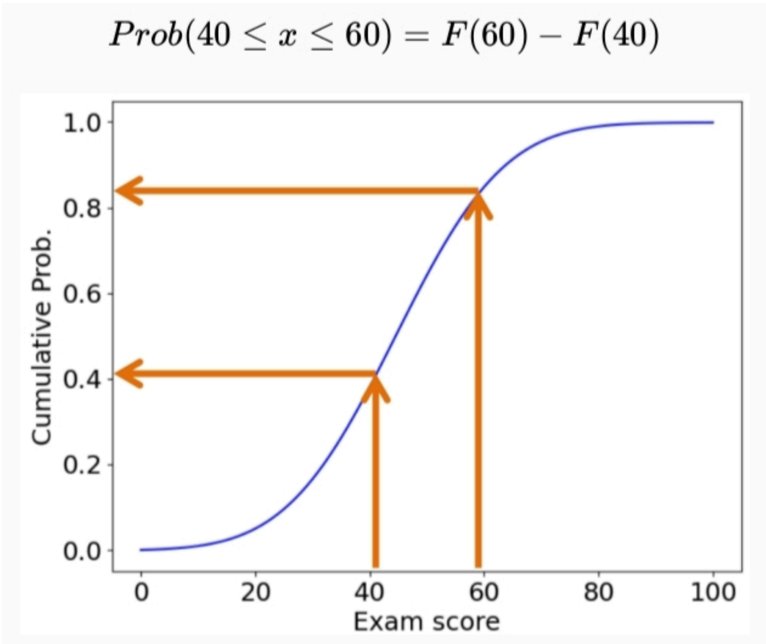


Cumulative probability distribution

Cumulative distribution function (CDF)

Example: Examination score

- Probability that score is between 40 and 60.



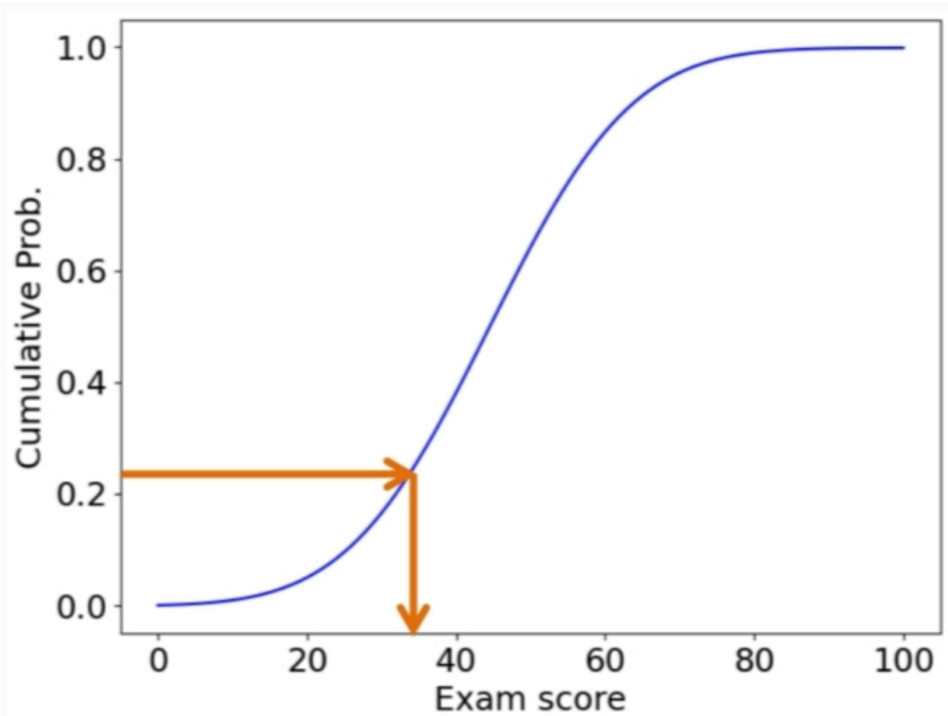
Percentile

- The r^{th} percentile is the value of the variable X such that its cumulative probability is r/100.
- In other words, r percent of samples are less than the r^{th} percentile
- Ex. Median is 50 percentile

$$Prob(X \le median) = F(median) = 0.5$$

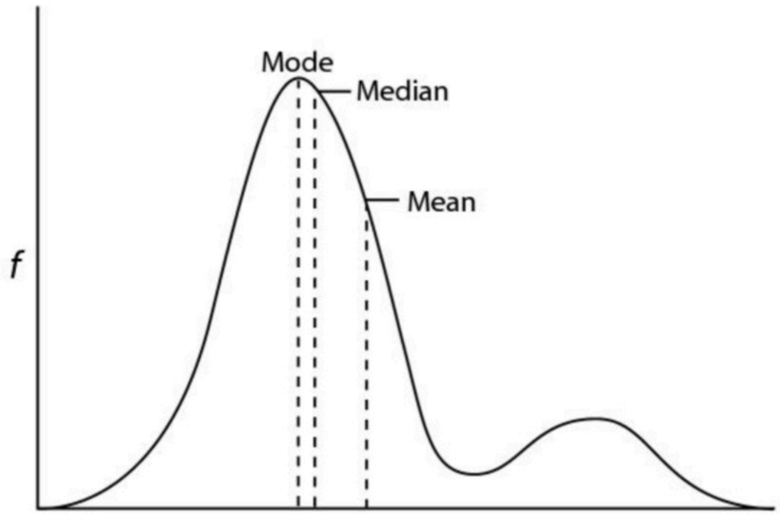
Percentile

- Example: Finding 25 percentile



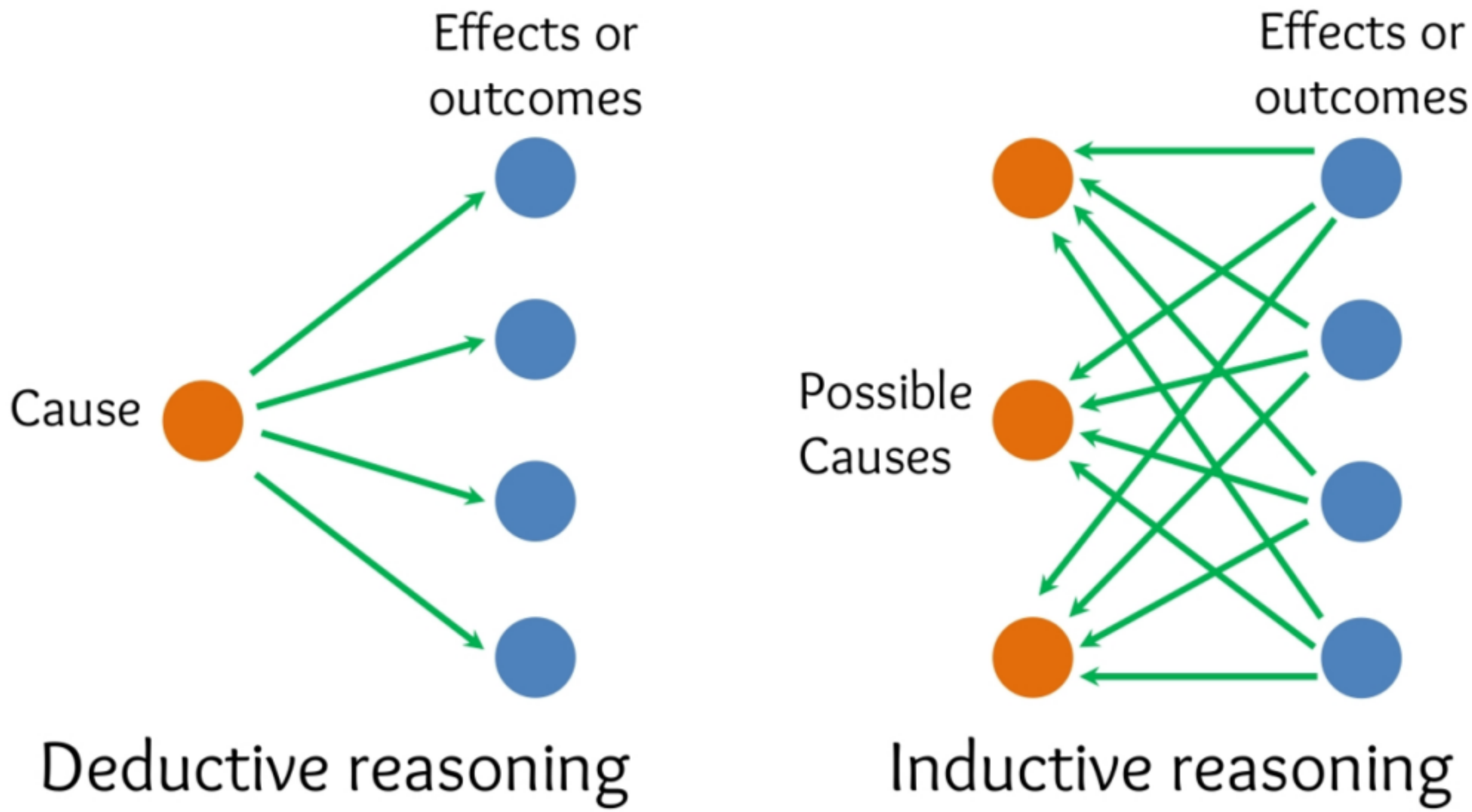
Mean Mode Median

- Mean is from expectation calculation
- Median is 50 percentile
- Mode is where PDF has maximum value



Introduction to Bayesian statistics and its applications

Inference in sciences



Bayes theorem

- The Bayes' theorem is surprisingly very simple:

$$Prob(H|E) = \frac{Prob(E|H)Prob(H)}{Prob(E)}$$

$$\begin{aligned} Prob(E) &= \sum_{i=1}^{N} Prob(E, H_i) \\ &= \sum_{i=1}^{N} Prob(E|H_i)Prob(H_i) \end{aligned}$$

Bayes theorem

$$Prob(H|E) = \frac{Prob(E|H)Prob(H)}{Prob(E)}$$

Let H be hypothesis, and E evidence

- Prob(H) <- **prior**
- Prob(E) <- **evidence**
- Prob(H|E) <- **posterior**
- Prob(E|H) <- **likelihood**

The origin of Bayes' theorem

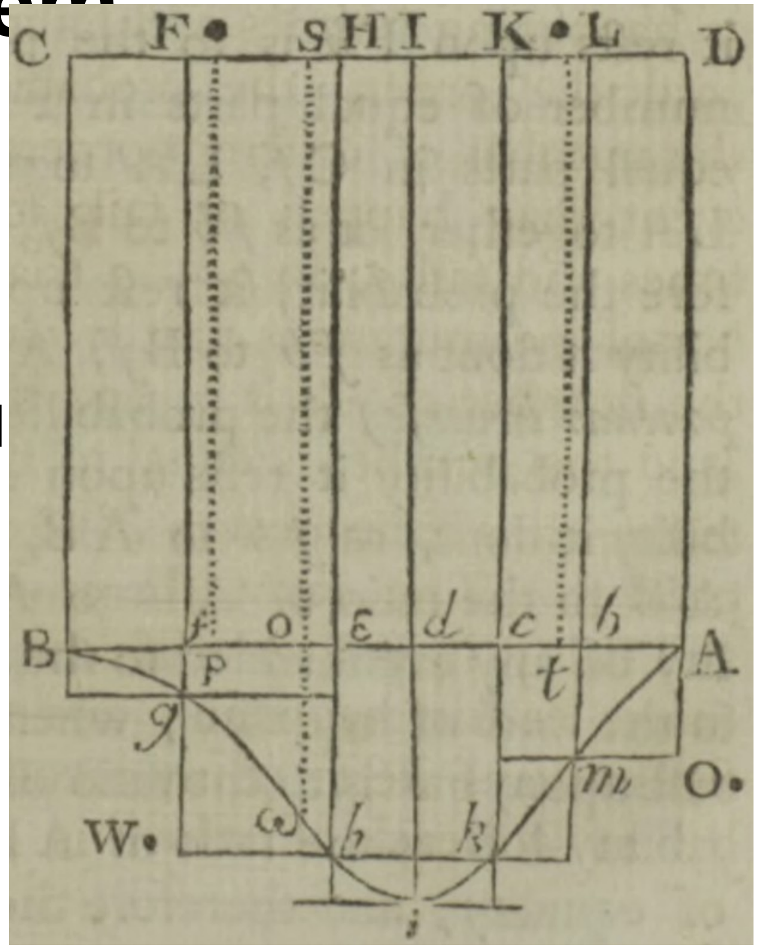
LII. *An Eſſay towards ſolving a Problem in the Doctrine of Chances. By the late Rev. Mr.* Bayes, *F. R. S. communicated by Mr.* Price, *in a Letter to* John Canton, *A. M. F. R. S.*

Dear Sir,

Read Dec. 23, 1763.

I Now ſend you an eſſay which I have found among the papers of our deceaſed friend Mr. Bayes, and which, in my opinion,

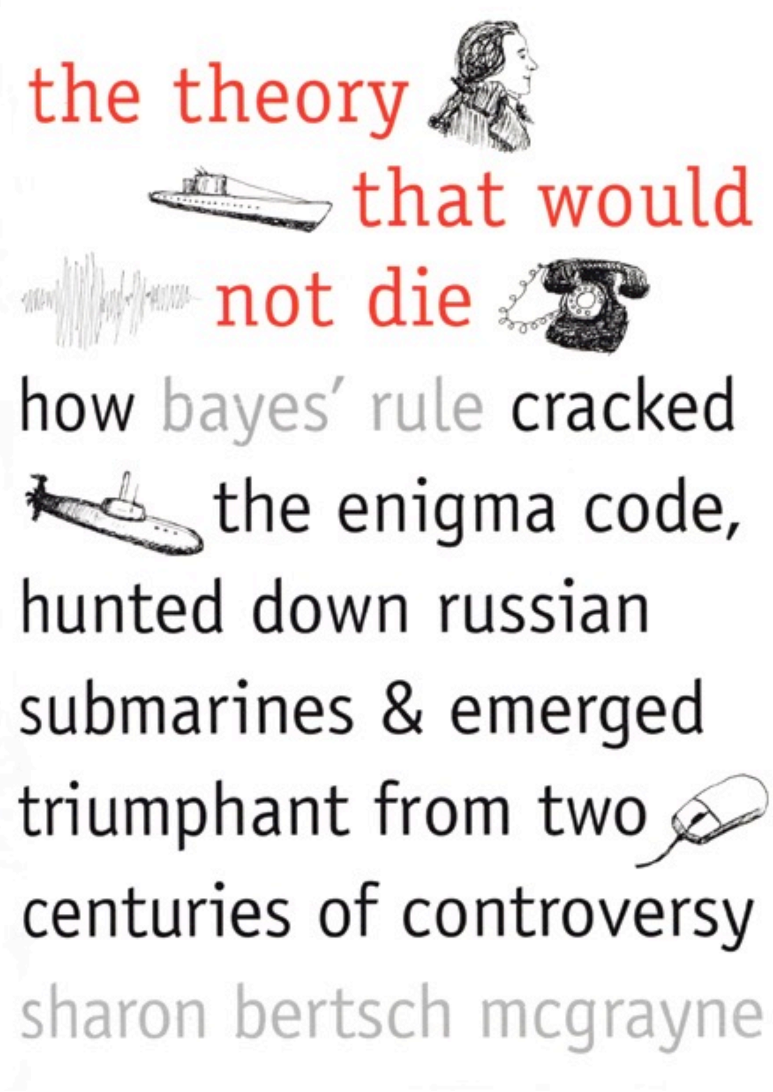
The origin of Bayes' theorem



Bayes' pool table experiment

- The pool table has corners ABCD and unit width and height.
- A white ball is throwed and falled at line os and the distance between lines os and AD is θ
- Then, n red balls are throwed and there are k red balls fall at the right side of os
- What is the probable value of θ ?
- That is, we want to find the probability that θ is in a range given ball throwing information

History


the theory
that would
not die
how bayes' rule cracked
the enigma code,
hunted down russian
submarines & emerged
triumphant from two
centuries of controversy
sharon bertsch mcgrayne

We all have prior

We all have prior

The case of Sally Clark

- Mother Sally Clark charged with double murder because of two cot-deaths (SIDS).
- Hypotheses
  - H_1 = Two children died of SIDS.
  - H_2 = Sally Clark committed double murder.
- Evidences
  - D_1 = First cot-death.
  - D_2 = Second cot death.

The case of Sally Clark

Comment from pediatrician

- Chance of cot-death from non-smoking mother is 1:8,500.

$$Prob(D_1 \mid H_1) = 1 : 8,500$$

$$\begin{aligned} Prob(D_1, D_2 \mid H_1) &= Prob(D_2 \mid H_1) Prob(D_1 \mid H_1) \\ &= (1 : 8,500) * (1 : 8,500) \\ &= 1 : 72M \end{aligned}$$

The case of Sally Clark

- She was charged with double murder, tried and convicted in 1999 because of very rare event because $Prob(D_1, D_2 | H_1)$ is extremely small.

What was wrong with this conclusion?

The case of Sally Clark

- Error 1: Evidence are not independent

The chance of the second cot-death given the first cot-death is 1:100.

$$\begin{aligned} Prob(D_1, D_2 \mid H_1) &= Prob(D_2 \mid D_1, H_1) * Prob(D_1 \mid H_1) \\ &= (1:8,500) * (1:100) \\ &= 1:850,000 \end{aligned}$$

The case of Sally Clark

- Error 2: Prosecutor's fallacy

1 in 850,000 is very rare event; but does not mean unlikely to be inn

$$Prob(D_1, D_2 \mid H_1) \text{ IS NOT EQUAL } Prob(H_1 \mid D_1, D_2)$$

The case of Sally Clark

- Error 2: Prosecutor's fallacy
  - Conditional probability is **not** symmetric.

$$Prob(A \mid B) \neq Prob(B \mid A)$$

  - Examples:

$$Prob(\text{innocent} \mid \text{DNA match}) \quad \neq \quad Prob(\text{DNA match} \mid \text{innocent})$$

$$Prob(\text{has beard} \mid \text{male}) \quad \neq \quad Prob(\text{male} \mid \text{has beard})$$

The case of Sally Clark

- Error 3: Ignoring alternative hypothesis

Double murder is also very **rare!**

$$
\begin{aligned}
Prob(1^{st}\ murder) &= 1 : 21,700 \\
Prob(2^{nd}\ murder \mid 1^{st}\ murder) &= 200 \cdot Prob(1^{st}\ murder) \\
Prob(H_2) &= (1 : 21,700) * (200 : 21,700) \\
&= 1 : 2.4M
\end{aligned}
$$

The case of Sally Clark

$$Prob(H_2) = 1 : 2.4M \quad \Rightarrow \quad Prob(H_1) = 1 - Prob(H_2)$$

$$\begin{aligned} Prob(D_1, D_2 \mid H_1) &= 1 : 850,000 \\ Prob(D_1, D_2 \mid H_2) &= 1 \end{aligned}$$

The case of Sally Clark

$$
\begin{aligned}
Prob(H_1 \mid D_1, D_2) &= \frac{Prob(D_1, D_2 \mid H_1)Prob(H_1)}{Prob(D_1, D_2 \mid H_1)Prob(H_1) + Prob(D_1, D_2 \mid H_2)Prob(H_2} \\
&= 0.74 \\
Prob(H_2 \mid D_1, D_2) &= 1 - Prob(H_1 \mid D_1, D_2) \\
&= 0.26
\end{aligned}
$$

The case of Sally Clark

$$\begin{aligned} Prob(H_1 \mid D_1, D_2) &= \frac{Prob(D_1, D_2 \mid H_1) Prob(H_1)}{Prob(D_1, D_2 \mid H_1) Prob(H_1) + Prob(D_1, D_2 \mid H_2) Prob(H_2} \\ &= 0.74 \\ Prob(H_2 \mid D_1, D_2) &= 1 - Prob(H_1 \mid D_1, D_2) \\ &= 0.26 \end{aligned}$$

The case of Sally Clark

$$\frac{Prob(H_1 \mid D_1, D_2)}{Prob(H_2 \mid D_1, D_2)} = 2.8$$

The chance of innocence is greater than guilty!!!!!

Sally Clark was then release after three years in prison.

Harvard Medical School Survey

- A high accuracy disease detector was developed using machine learning with a 5% false negative rate and a 1% false negative rate.
- Having positive test, what is the probability that a patient has the disease?

Harvard Medical School Survey

- Let's see what we think the answer should be. First, some background and terminology:
  - *Prevalence* in the population is the percent of people who have the disease. It is also called the *base rate* of the disease.
  - A *positive* result means that according to the test the person has the disease.
  - A *negative* result means that according to the test the person doesn't have the disease.
  - The test can give a wrong result. The *false positive rate* is the proportion of positive results among people who don't have the disease.

Harvard Medical School Survey

- Lets define some variables:
  - T+ : "Positive test"
  - T− : "Negative test"
  - D+ : "Patient has disease"
  - D− : "Patient does not have disease"

- False positive rate: $Prob(T+|D-) = 0.01$
- False negative rate: $Prob(T-|D+) = 0.05$

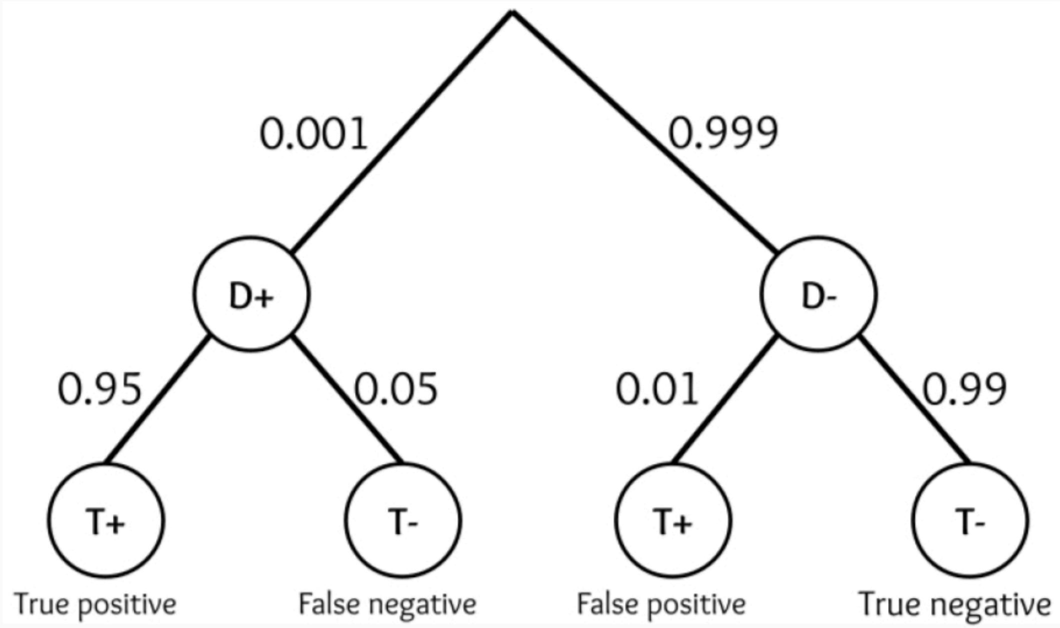
Harvard Medical School Survey

- We have a test result and want to know the chance of having a disease. That is:

$$Prob(D+|T+) = \frac{Prob(T+|D+)Prob(D+)}{Prob(T+)}$$

$$\begin{aligned} Prob(T+) &= Prob(T+, D+) + Prob(T+, D-) \\ &= Prob(T+|D+)Prob(D+) + Prob(T+|D-)Prob(D-) \end{aligned}$$

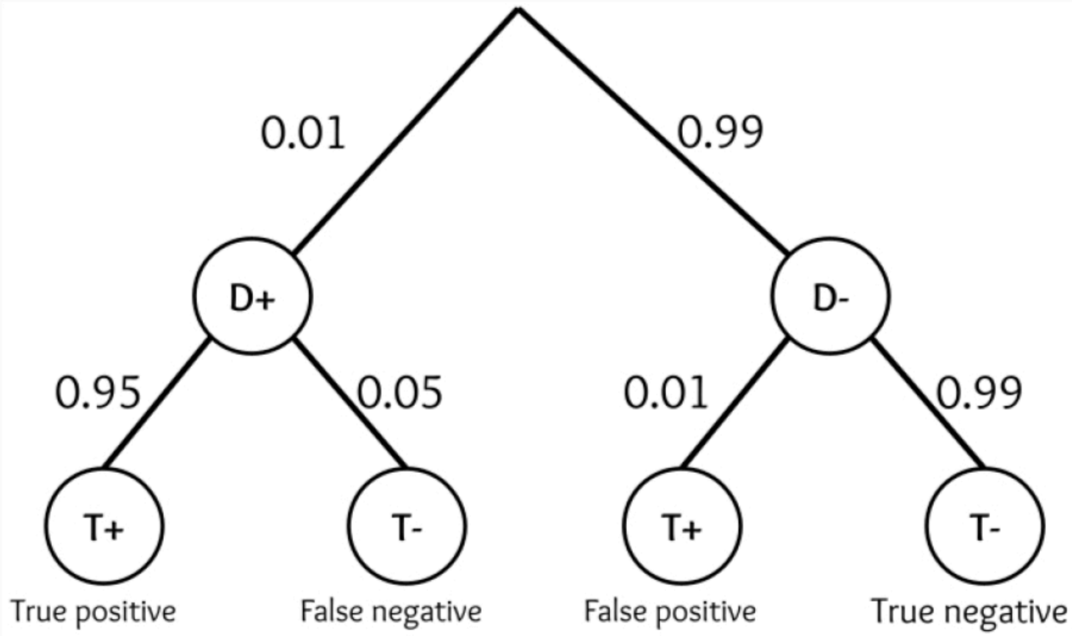
Harvard Medical School Survey



- **Low risk patient**: $Prob(D+) = 0.001$, $Prob(D-) = 0.999$

$$Prob(D+|T+) = \frac{0.95 * 0.001}{0.95 * 0.001 + 0.01 * 0.999} = 0.09$$

Harvard Medical School Survey



- **High risk patient**: $Prob(D+) = 0.01$, $Prob(D-) = 0.99$

$$Prob(D + |T+) = \frac{0.95 * 0.01}{0.95 * 0.01 + 0.01 * 0.99} = 0.49$$

Detect rare event is hard

- If we want to be strongly confident you have detected a very rare event, we need an extremely accurate detector.

Example:

- You want to design a blood test for a rare disease that occurs by chance in 1 person in 100,000.
- If you have the disease, the test will report that you do with probability p (and that you do not with probability $(1-p)$).
- If you do not have the disease, the test will report a false positive with probability q.
- You want to choose the value of p so that if the test says you have the disease, there is at least a 50% probability that you do.

Detect rare event is hard

- Using the same notations from the previous example
- $Prop(D+) = 10^{-5}, Prob(D-) = 1 - 10^{-5}$
- $Prob(T+|D+) = p \;, \; Prob(T+|D-) = q$

$$Prob(D+|T+) = \frac{Prob(T+|D+)Prob(D+)}{Prob(T+|D+)Prob(D+) + Prob(T+|D-)Prob(D-)}$$

We want to have the probability of having disease greater than 50%

$$0.5 \leq \frac{p \times 10^{-5}}{p \times 10^{-5} + q(1 - 10^{-5})}$$

Detect rare event is hard

- Some algebraic manipulation, we obtain

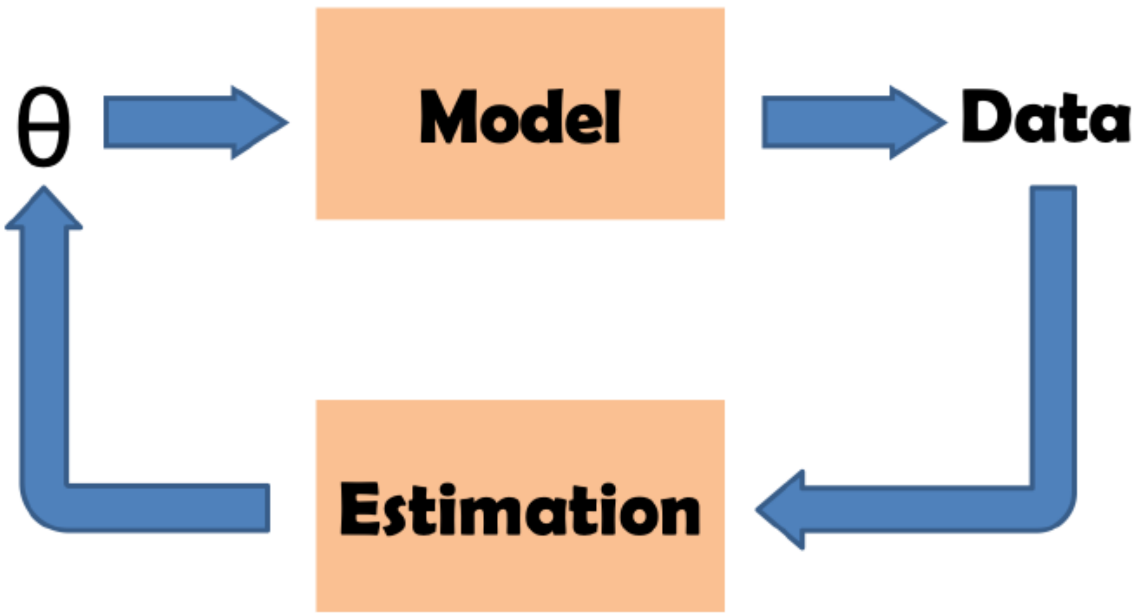
$$p \geq 99999q$$

- One plausible pair of values is

$$\begin{aligned} p &= 1 - 10^{-5} \\ q &= 10^{-5} \end{aligned}$$

Bayesian parameter estimation

- Sequential estimation



Bayesian classification

- Example: Voice classification

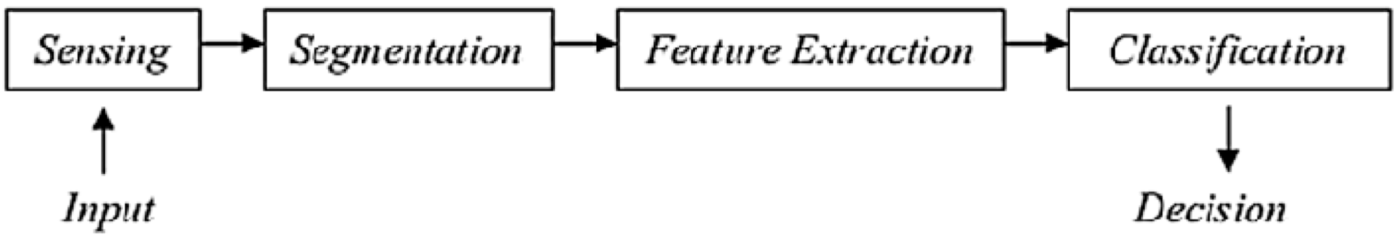
$$P(\text{“Hello dear”}|\text{Signal}) = \frac{P(\text{Signal}|\text{“Hello dear”})P(\text{“Hello dear”})}{P(\text{Signal})}$$

$$\begin{aligned} P(\text{Signal}) = \; & P(\text{Signal}|\text{“Hello dear”})P(\text{“Hello dear”}) \\ & + P(\text{Signal}|\text{“Hello deer”})P(\text{“Hello deer”}) \end{aligned}$$

Bayesian classification

Feature based classification

- Features are the object characteristics.
- Properties of good features:
  - Very similar for the object in the same category or class
  - Very different for objects in different categories or classes



Bayesian classification

Probability of a class C given features

$$F = \{F_1, F_2, \ldots, F_N\}$$

$$Prob(C \mid F) \quad = \quad \frac{Prob(F \mid C)Prob(C)}{Prob(F)} \quad \propto Prob(F \mid C)Prob(C)$$

Class determination

Class C_1 is more probable than class C_2, if

$$\frac{Prob(F \mid C_1)Prob(C_1)}{Prob(F \mid C_2)Prob(C_2)} > 1$$

$$Prob(F \mid C_1)Prob(C_1) > Prob(F \mid C_2)Prob(C_2)$$

Bayesian classification

Multiclass classification

$$
\begin{aligned}
\text{Plausibale class} &= \underset{C}{\operatorname{argmax}} Prob(C \mid F) \\
&= \underset{C}{\operatorname{argmax}} Prob(F \mid C) Prob(C)
\end{aligned}
$$

Independent features

Order does not matter. That is why we call it “Naive” Bayes classifier.

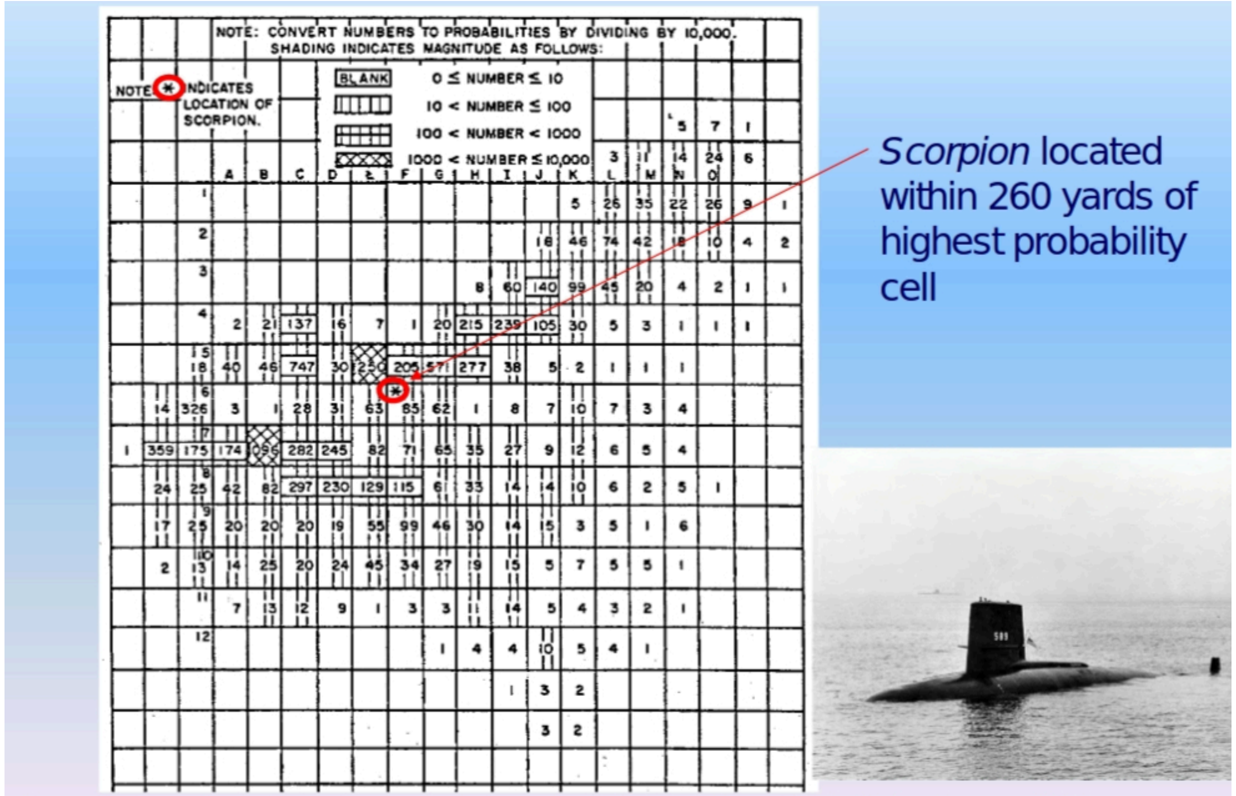
$$
\begin{aligned}
Prob(F \mid C) &= Prob(F_1 \mid C) Prob(F_2 \mid C) \ldots Prob(F_N \mid C) \\
&= \prod_{i=1}^{N} Prob(F_i \mid C)
\end{aligned}
$$

Bayesian classification

- Which area has more chance to detect man-made object?

Bayesian Search theory

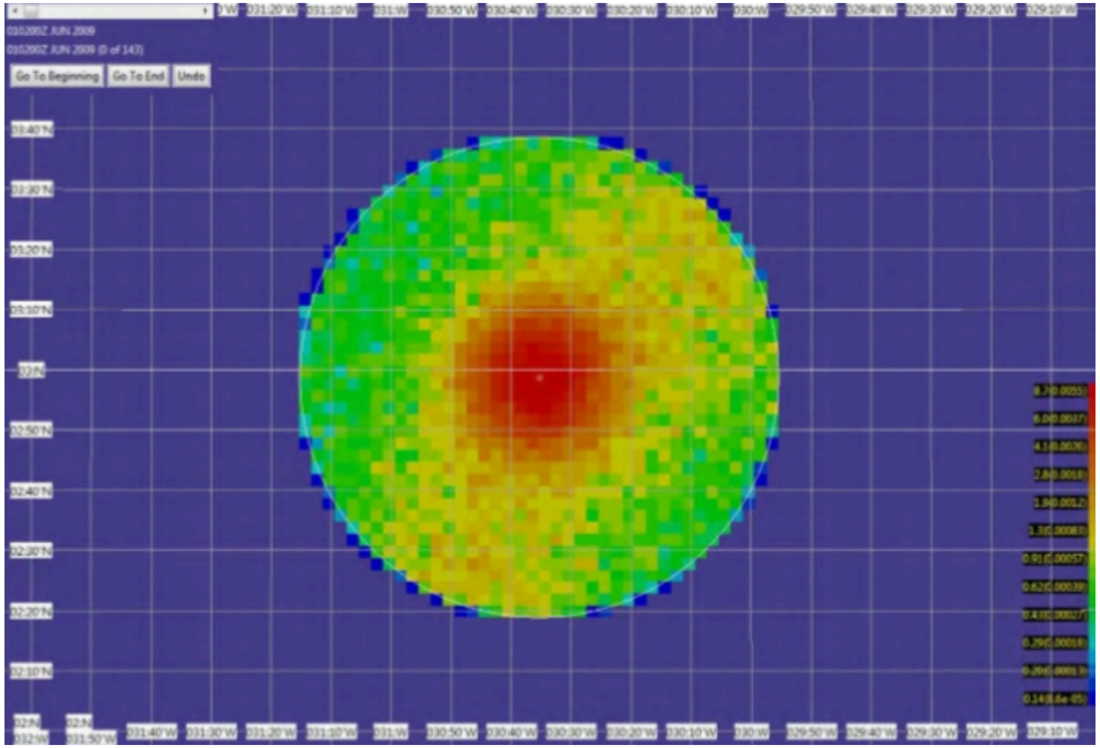
- Bayesian Search Theory applies Bayesian statistics to a search problem in a more efficient way than just randomly searching all of the possibilities.



Operations analysis during the underwater search for *Scorpion*, by H R Richardson and L D Stone, *Naval Research Logistics*, 1971, 18:141-157

Bayesian Search theory

- Start from making prior distribution



Bayesian Search theory

- Let Y_i be a random variable with two states:
  - $Y_i = 0$ means “The submarine is **not** present in sector *i*”
  - $Y_i = 1$ means “The submarine is present in sector *i*”
- Let X_i be a random variable for the outcome of searching a sector i.
  - $X_i = 0$ if the submarine is **not** found in the sector i
  - $X_i = 1$ if the submarine is found in the sector I
- Clearly X_i is dependent on Y_i

Bayesian Search theory

- If the submarine is not present in the i^{th} sector, it cannot possibly be located in that sector, so:

$$Prob(X_i = 1|Y_i = 0) = 0$$

- On the other hand, if it is indeed present in the i^{th} sector, the probability that it will be found is p:

$$Prob(X_i = 1|Y_i = 1) = p$$

Bayesian Search theory

- A search is not guaranteed to be successful, so $p < 1$
- That is, there is a nonzero probability that we fail to detect the submarine in a search of the i^{th} sector, even though it is really there.
- Assume that the a priori probability of the submarine being in the i^{th} sector is π_i

Bayesian Search theory

- If we search at the sector i^th and the submarine is not found, the probability a posteriori that the ship is there is

$$Prob(Y_i = 1 | X_i = 0) = \frac{Prob(X_i = 0 | Y_i = 1) Prob(Y_i = 1)}{Prob(X_i = 0)}$$

Bayesian Search theory

$$Prob(Y_i = 1 | X_i = 0) = \frac{Prob(X_i = 0 | Y_i = 1) Prob(Y_i = 1)}{Prob(X_i = 0)}$$

$$\begin{aligned}
Prob(X_i = 0) &= Prob(X_i = 0, Y_i = 0) + Prob(X_i = 0, Y_i = 1) \\
&= Prob(X_i = 0 \mid Y_i = 0) Prob(Y_i = 0) \\
&\quad + Prob(X_i = 0 \mid Y_i = 1) Prob(Y_i = 1) \\
&= 1 \times (1 - \pi_i) + (1 - p) \times \pi_i \\
&= 1 - p\pi_i
\end{aligned}$$

Bayesian Search theory

- The posterior is then

$$Prob(Y_i = 1 | X_i = 0) = \left[\frac{1-p}{1-p\pi_i}\right]\pi_i$$

- Failure to locate the submarine in a search of sector i^{th} does not preclude the possibility that it is there, but makes the posterior probability smaller than the prior probability: the ratio

$$\frac{1-p}{1-p\pi_i}$$

Bayesian Search theory

- The probability p refers to the efficiency of the search.
- If p is large, the prior π_i will be reduce drastically.
- How does the search at sector ith affect the possibility at other location?

$$Prob\left(Y_j = 1 \middle| X_i = 0\right) = \left[\frac{1}{1 - p\pi_j}\right]\pi_j$$

- The fact that the submarine is not located by a search of sector ith enhances our belief that it is in any of the sectors $i \neq j$, for the factor that multiplies the prior probability π_j is greater than one.