

## Tasks and concepts of AI: Generation

Introduction: AI Today and in the Past. Probability and Bayes Rule

## Brief History of AI

## The first thoughts about artificial intelligence

- ▶ Hephaestus created for himself Android robots, such as a giant human-like robot of Talos.
- ▶ Pygmalion revived Galatea.
- ▶ Jehovah and Allah - pieces of clay
- ▶ Particularly pious and learned rabbis could create golems.
- ▶ Albert the Great made an artificial speaking head (which very upset Thomas Aquinas).

# Mechanical machines

Robots and Automatic Machines Were Generally Very Inventive:  
Al-Jazari (XII Century)



# Mechanical machines

Jaquet-Droz automata (XVIII century):



## Mechanical machines

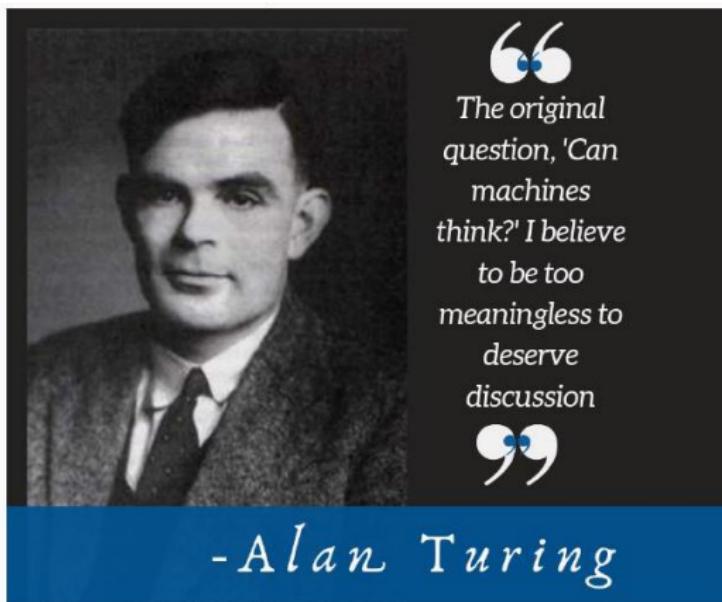
- ▶ But this is in mechanics, in mathematics/logic AI it was quite rudimentary for a long time



Figure 1: Logic machine of Ramon Llull (XIII-XIV centuries)

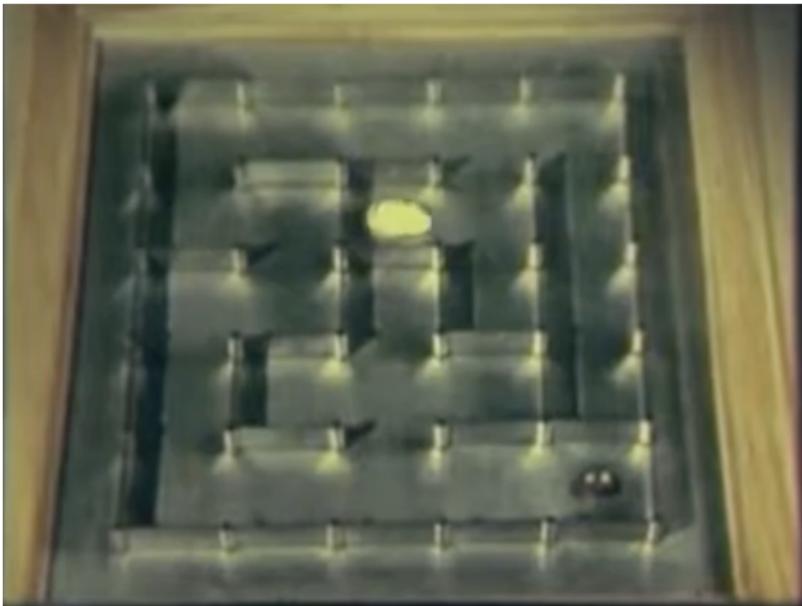
## Turing Test

- ▶ AI as a science begins with a Turing test (1950).
- ▶ The idea of the Turing test is to check if a machine can imitate a human in a conversation.
- ▶ The original formulation was more nuanced.



## Shannon's Theseus

- ▶ YouTube Video
- ▶ Early 1950s, Claude Shannon (The father of Information Theory) demonstrates Theseus
- ▶ A life-sized magnetic mouse controlled by relay circuits, learns its way around a maze.



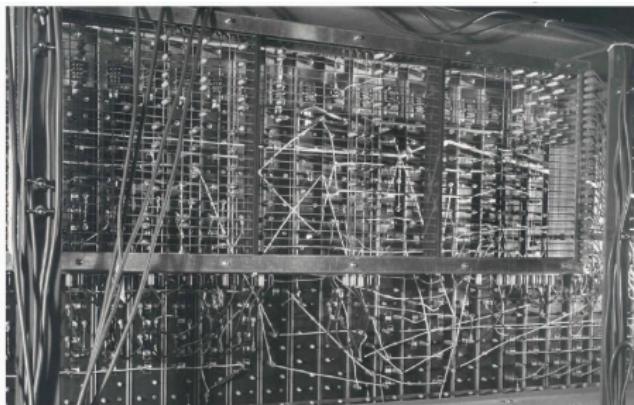
## Stanford Cart

- ▶ YouTube Video
- ▶ Takes 2.6-second for signal to travel from earth to the moon
- ▶ Latest iterations is automated with 3D vision capabilities
- ▶ Pause after each meter of movement and take 10-15 minutes to reassess its surroundings and reevaluate its decided path.
- ▶ In 1979, this cautious version of the cart successfully made its way 20 meters through a chair-strewn room in five hours without human intervention.

# Turing Test

It takes a lot to create an AI system:

- ▶ Processing of a natural language
- ▶ Sensors and actuators
- ▶ Representation of knowledge
- ▶ Inference from the existing knowledge
- ▶ Training on experience (Machine Learning).



## Dartmouth workshop

- ▶ AI as a science appeared in 1956 at the Dartmouth workshop.
- ▶ It was organized by John McCarthy, Marvin Minsky, Claude Shannon and Nathaniel Rochester.
- ▶ It was probably the most ambitious grant proposal in the history of computer science.

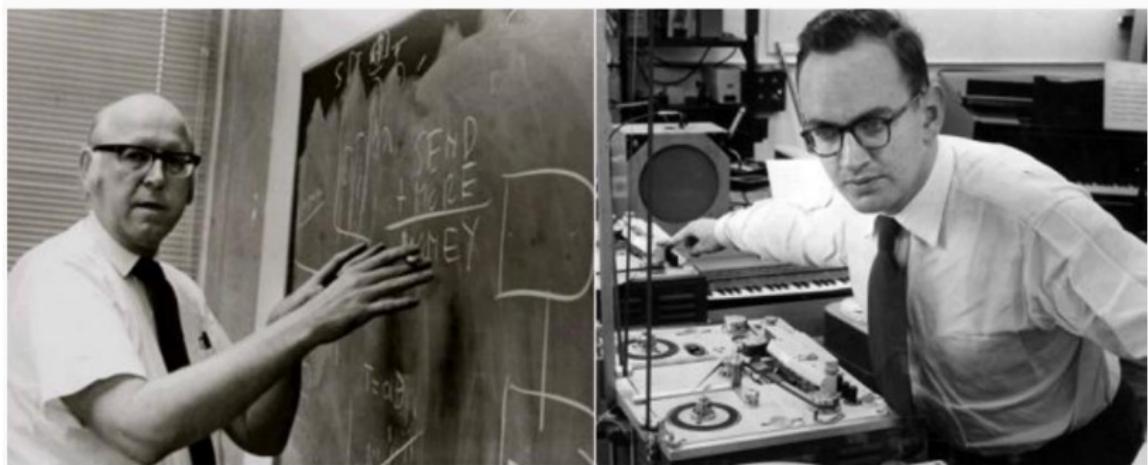


## Dartmouth workshop

We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at Dartmouth College in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

## 1956-1960: Great hopes

- ▶ Optimistic time. It seemed a that we were almost there...
- ▶ Allen Newell, Herbert A. Simon, and Cliff Shaw: *Logic Theorist*.
- ▶ Automated reasoning.
- ▶ It was able to prof most of the Principia Mathematica, in some places even more elegant than Russell and Whitehead.

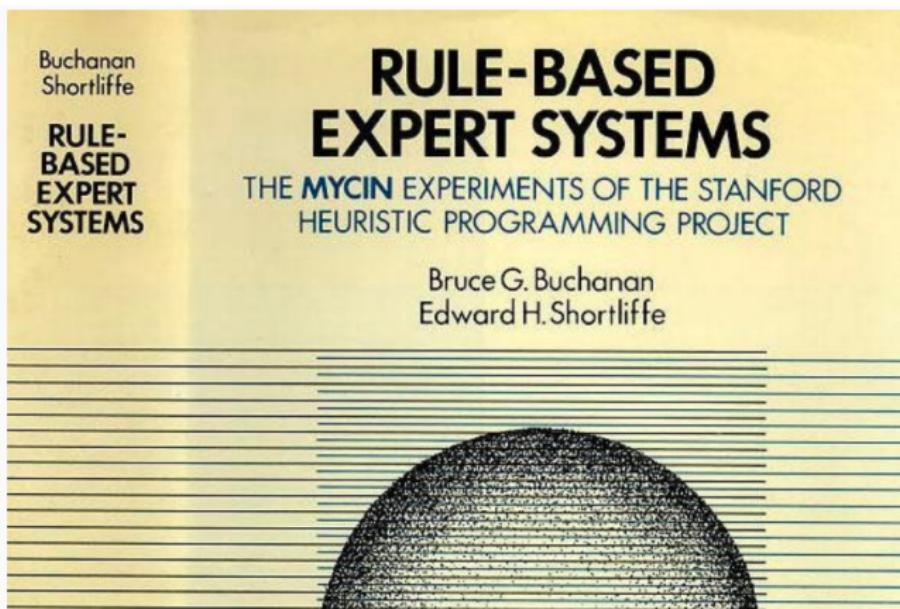


## 1956-1960: Big Hopes

- ▶ General Problem Solver - a program that tried to think as a person
- ▶ A lot of programs that have been able to do some limited things (MicroWorlds):
  - ▶ Analogy (IQ tests with multiple choice questions)
  - ▶ Student (algebraic verbal tasks)
  - ▶ Blocks World (rearranged 3D blocks).

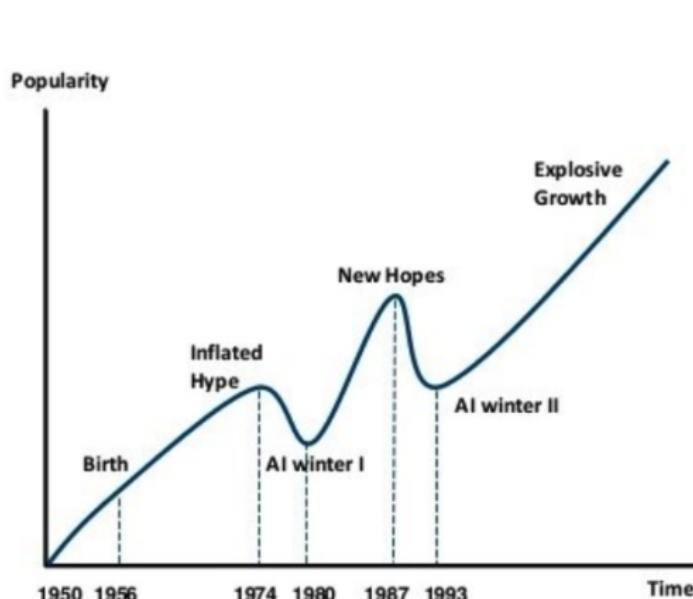
## 1970s: Knowledge Based Systems

- ▶ The bottom line: to accumulate a fairly large set of rules and knowledge about the subject area, then draw conclusions.
- ▶ First success: MYCIN - Diagnosis of blood infections:
  - ▶ about 450 rules
  - ▶ The results are like an experienced doctor and significantly better than beginner doctors.



## 1980-2010: Commercial applications Industry AI

- ▶ The first AI department was at Dec (Digital Equipment Corporation). It is argued that by 1986 he saved the Dec about \$10 million per year.
- ▶ The boom ended by the end of the 80s, when many companies could not live up to high expectations.



Timeline of AI Development

- **1950s-1960s:** First AI boom - the age of reasoning, prototype AI developed
- **1970s:** AI winter I
- **1980s-1990s:** Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s:** AI winter II
- **1997:** Deep Blue beats Gary Kasparov
- **2006:** University of Toronto develops Deep Learning
- **2011:** IBM's Watson won Jeopardy
- **2016:** Go software based on Deep Learning beats world's champions

## 1990-2010: DATA MINING, MACHINE LEARNING

- ▶ In recent decades, the main emphasis has shifted to machine training and search for patterns in the data.
- ▶ Especially - with the development of the Internet.
- ▶ Not too many people remember the original AI ideas, but Machine Learning is now everywhere.
- ▶ But Robotics flourishes and uses Machine Learning at every step.

## Rule-Based System vs Bayes

- ▶ Since 1956, the field of artificial intelligence (AI) has undergone significant transformations
- ▶ *traditional AI* was mostly focused on rule-based systems and boolean logic programming, with limited learning capabilities. It lead to them being brittle in changing environments.
- ▶ On the other hand, *emerging AI* is focused on modeling uncertainties, pattern matching, and deep learning.
- ▶ All of those are data-driven approaches.
- ▶ These approaches are more adaptable and can handle complex and unstructured data. They are also more data-dependent and lack interpretability.

# Rule-Based System vs Bayes

## *Old AI*

*If rain outside, then take umbrella*

*This rule cannot be learned from data. It does not allow inference. Cannot say anything about rain outside if I see an umbrella.*

## *New AI*

*Probability of taking umbrella, given there is rain*

*Conditional probability rule can be learned from data. Allows for inference. We can calculate the probability of rain outside if we see an umbrella.*

- ▶ Bayesian approach is a powerful statistical framework based on the work of Thomas Bayes and later Laplace.
- ▶ It provides a probabilistic approach to reasoning and learning
- ▶ Allowing us to update our beliefs about the world as we gather new data.
- ▶ This makes it a natural fit for artificial intelligence, where we often need to deal with uncertainty and incomplete information.

# DEFINITION

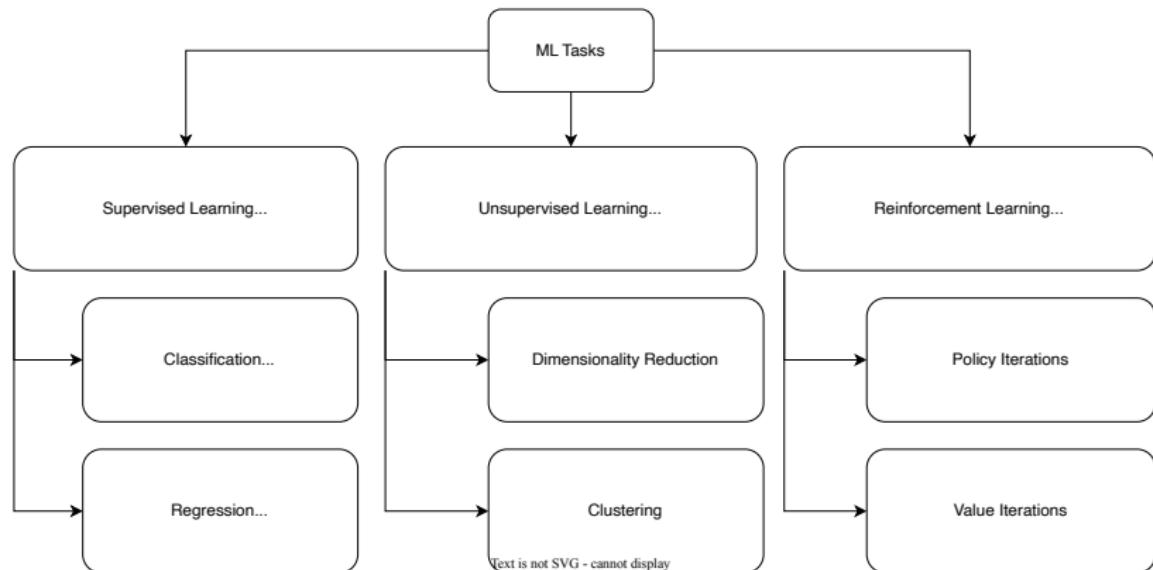
- ▶ How to determine “learning”?

Definition:

The computer program learns as the data is accumulating relative to a certain problem class  $T$  and the target function of  $P$  if the quality of solving these problems (relative to  $P$ ) improves with gaining new experience.

- ▶ The definition is very (too?) General.
- ▶ What specific examples can be given?

# Tasks and concepts of ML



## Tasks and concepts of ML: Supervised Learning

- ▶ training sample – a set of examples, each of which consists of input features (attributes) and the correct “answers” - the response variable
- ▶ Learn a rule that maps input features to the response variable
- ▶ Then this rule is applied to new examples (deployment)
- ▶ The main thing is to train a model that explains not only examples from the training set, but also new examples (generalizes)
- ▶ Otherwise - overfitting

## Tasks and concepts of ML: unsupervised learning

There are no correct answers, only data, e.g. *clustering*:

- ▶ We need to divide the data into pre -unknown classes to some extent similar:
  - ▶ highlight the family of genes from the sequences of nucleotides
  - ▶ cluster users and personalize the application for them
  - ▶ cluster the mass spectrometric image to parts with different composition

## Tasks and concepts of ML: unsupervised learning

- ▶ Dimensionality reduction: data have a high dimension, it is necessary to reduce it, select the most informative features so that all of the above algorithms can work
- ▶ Matrix Completion: There is a sparse matrix, we must predict what is in the missing positions.
- ▶ Anomaly detection: find anomalies in the data, e.g. fraud detection.
- ▶ Often the outputs answers are given for a small part of the data, then we call it semi-supervised Learning.

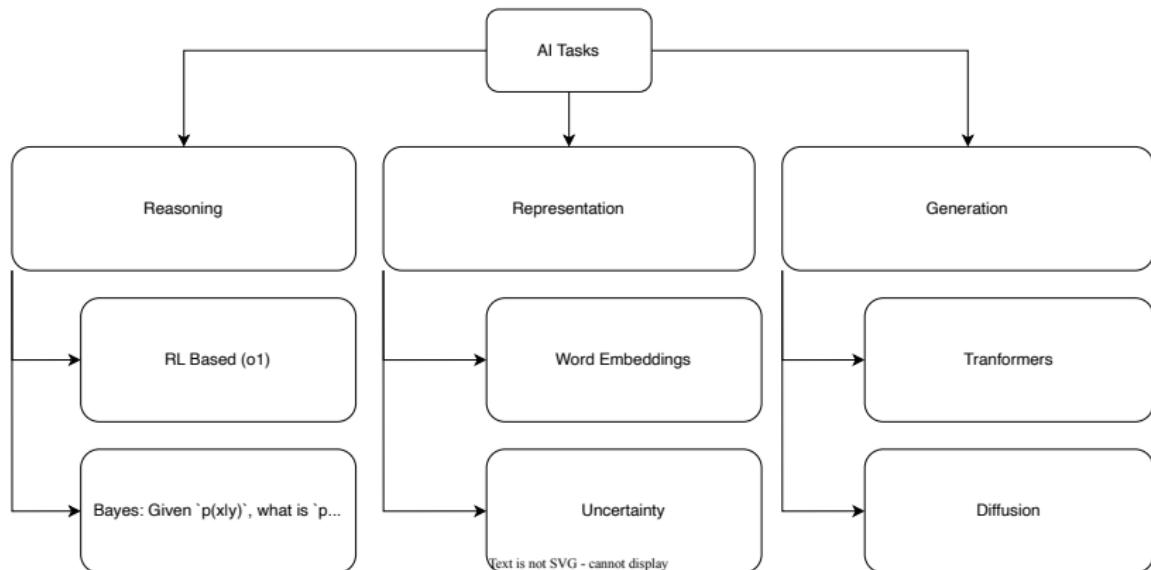
## Tasks and concepts of ML: reinforcement learning

- ▶ Multi-armed bandits: there is a certain set of actions, each of which leads to random results, you need to get as much rewards possible
- ▶ Exploration vs. Exploitation: how and when to proceed from the study of the new to use what has already studied
- ▶ Credit Assignment: You get rewarded at the very end (won the game), and we must somehow distribute this reward on all the moves that led to victory.

## Tasks and concepts of ML: active learning

- ▶ Active Learning - how to choose the following (relatively expensive) test
- ▶ Boosting - how to combine several weak classifiers so that it turns out good
- ▶ Model Selection - where to draw a line between models with many parameters and with a few.
- ▶ Ranking: response list is ordered (internet search)

# Tasks and concepts of AI



## Tasks and concepts of AI: Reasoning

- ▶ Bayesian networks: given conditional probabilities, calculate the probability of the event
- ▶ o1 by OpenAI: a family of AI models that are designed to perform complex reasoning tasks, such as math, coding, and science. o1 models placed among the top 500 students in the US in a qualifier for the USA Math Olympiad (AIME)
- ▶ Gemini 2.0: model for the agentic era

## Tasks and concepts of AI: Representation

- ▶ Knowledge Graphs: a graph database that uses semantic relationships to represent knowledge
- ▶ Embeddings: a way to represent data in a lower-dimensional space
- ▶ Transformers: a deep learning model that uses self-attention to process sequential data

## Tasks and concepts of AI: Generation

In shadows of data, uncertainty reigns,  
Bayesian whispers, where knowledge remains.

With prior beliefs, we start our quest,  
Updating with evidence, we strive for the best.

A dance of the models, predictions unfold,  
Inferences drawn, from the new and the old.

Through probabilities, we find our way,  
In the world of AI, it's the Bayesian sway.

So gather your data, let prior thoughts flow,  
In the realm of the unknown, let your insights grow.  
For in this approach, with each little clue,  
We weave understanding, both rich and true.

Music

## Tasks and concepts of AI: Generation

```
from openai import OpenAI
client = OpenAI(api_key="your-api-key")
response = client.images.generate(
    model="dall-e-3",
    prompt="a hockey player trying to understand the Bayes
size="1024x1024",
    quality="standard",
    n=1,
)
print(response.data[0].url)
```

A humorous and illustrative scene of a hockey player sitting on a bench in full gear, holding a hockey stick in one hand and a whiteboard marker in the other, looking at a whiteboard with a Bayes' Rule calculation.



# Chess and AI

Old AI: Deep Blue (1997) vs. Garry Kasparov

May 11th, 1997

**Computer won world champion of chess**

(Deep Blue)

(Garry Kasparov)



## AlphaGo Zero

- ▶ Remove all human knowledge from training process - only uses self play,
- ▶ Takes raw board as input and neural network predicts the next move.
- ▶ Uses Monte Carlo tree search to evaluate the position.
- ▶ The algorithm was able to beat AlphaGo 100-0. The algorithm was then used to play chess and shogi and was able to beat the best human players in those games as well.

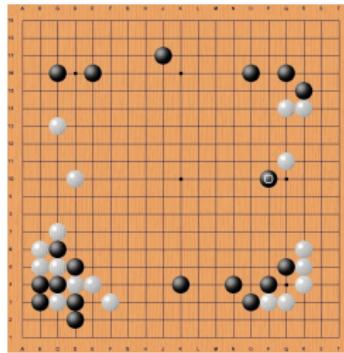


Figure 3: Alpha GO vs Lee Sedol: Move 37 by AlphaGo in Game Two

## Probability in machine learning

- ▶ In all methods and approaches, it is useful not only generate an answer, but also evaluate how confident in this answer, how well the model describes the data, how these values will change in further experiments, etc.
- ▶ Therefore, the central role in machine learning is played by the theory of probability - and we will also actively use it.

## Bayes Approach

## Review of Basic Probability Concepts

Probability lets us talk efficiently about things that we are uncertain about.

- ▶ What will Amazon's sales be next quarter?
- ▶ What will the return be on my stocks next year?
- ▶ How often will users click on a particular Google ad?

*All these involve estimating or predicting unknowns!!*

## Random Variables

**Random Variables** are numbers that we are not sure about. There's a list of potential outcomes. We assign probabilities to each outcome.

**Example:** Suppose that we are about to toss two coins. Let  $X$  denote the number of heads. We call  $X$  the random variable that stands for the potential outcome.

# Probability

**Probability** is a language designed to help us communicate about uncertainty. We assign a number between 0 and 1 measuring how likely that event is to occur. It's immensely useful, and there's only a few basic rules.

1. If an event  $A$  is certain to occur, it has probability 1, denoted  $P(A) = 1$
2. Either an event  $A$  occurs or it does not.

$$P(A) = 1 - P(\text{not } A)$$

3. If two events are mutually exclusive (both cannot occur simultaneously) then

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

4. Joint probability, when events are independent

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

## Probability Distribution

We describe the behavior of random variables with a **Probability Distribution**

**Example:** Suppose we are about to toss two coins. Let  $X$  denote the number of heads.

$$X = \begin{cases} 0 & \text{with prob. } 1/4 \\ 1 & \text{with prob. } 1/2 \\ 2 & \text{with prob. } 1/4 \end{cases}$$

$X$  is called a **Discrete Random Variable**

**Question:** What is  $P(X = 0)$ ? How about  $P(X \geq 1)$ ?

## Pete Rose Hitting Streak

**Pete Rose** of the Cincinnati Reds set a National League record of hitting safely in 44 consecutive games ...

- ▶ Rose was a .300 hitter. Assume he comes to bat 4 times each game.
- ▶ Each at bat is assumed to be independent, i.e., the current at bat doesn't affect the outcome of the next.

*What probability might reasonably be associated with that hitting streak?*

Joe DiMaggio's record is 56! His batting average was .325

## Pete Rose Hitting Streak (Solution)

Let  $A_i$  denote the event that “**Rose hits safely in the  $i$ th game**”

$P(\text{Rose Hits Safely in 44 consecutive games}) =$

$$P(A_1 \text{ and } A_2 \dots \text{ and } A_{44}) = P(A_1)P(A_2) \dots P(A_{44})$$

We now need to find  $P(A_i) \dots$  where  $P(A_i) = 1 - P(\text{not } A_i)$

$$\begin{aligned} P(A_1) &= 1 - P(\text{not } A_1) \\ &= 1 - P(\text{Rose makes 4 outs}) \\ &= 1 - (0.7)^4 = 0.76 \end{aligned}$$

So for the **winning streak** we have  $(0.76)^{44} = 0.0000057!!!$

## Pete Rose Hitting Streak (Inference)

There are **three** basic inferences:

- ▶ This means that the odds for a particular player as good as Pete Rose starting a hitting streak today are **175,470 to 1**
- ▶ Doesn't mean that the run of 44 won't be beaten by some player at some time: the **Law of Very Large Numbers**
- ▶ Joe DiMaggio's record is 56!!!! It's going to be hard to beat. We have  $(0.792)^{56} = 2.13 \times 10^{-6}$  or **455,962 to 1**

## Example: Happiness Index

“happiness index” as a function of salary.

Salary ( $X$ )	Happiness ( $Y$ ): 0 (low)	1 (medium)	2 (high)
low 0	0.03	0.12	0.07
medium 1	0.02	0.13	0.11
high 2	0.01	0.13	0.14
very high 3	0.01	0.09	0.14

Is  $P(Y = 2 | X = 3) > P(Y = 2)$ ?

## Bayes Rule

The computation of  $P(x | y)$  from  $P(x)$  and  $P(y | x)$  is called Bayes theorem ...

$$P(x | y) = \frac{P(y, x)}{P(y)} = \frac{P(y, x)}{\sum_x P(y, x)} = \frac{P(y | x)P(x)}{\sum_x P(y | x)P(x)}$$

This shows now the conditional distribution is related to the joint and marginal distributions.

You'll be given all the quantities on the r.h.s.

## Bayes Rule

Key fact:  $P(x | y)$  is generally different from  $P(y | x)$ !

Example: Most people would agree

$$Pr(Practice\ hard \mid Play\ in\ NBA) \approx 1$$

$$Pr(Play\ in\ NBA \mid Practice\ hard) \approx 0$$

The main reason for the difference is that  $P(Play\ in\ NBA) \approx 0.$

## Independence

Two random variable  $X$  and  $Y$  are independent if

$$P(Y = y \mid X = x) = P(Y = y)$$

for all possible  $x$  and  $y$  values. Knowing  $X = x$  tells you nothing about  $Y$ !

Example: Tossing a coin twice. What's the probability of getting  $H$  in the second toss given we saw a  $T$  in the first one?

## Sally Clark Case: Independence or Bayes?

Sally Clark was accused and convicted of killing her two children  
They could have both died of SIDS.

- ▶ The chance of a family which are non-smokers and over 25 having a SIDS death is around 1 in 8,500.
- ▶ The chance of a family which has already had a SIDS death having a second is around 1 in 100.
- ▶ The chance of a mother killing her two children is around 1 in 1,000,000.

# Bayes or Independence

## 1. Under Bayes

$$\begin{aligned} P(\text{both SIDS}) &= P(\text{first SIDS}) P(\text{Second SIDS}|\text{first SIDS}) \\ &= \frac{1}{8500} \cdot \frac{1}{100} = \frac{1}{850,000} \end{aligned}$$

The  $\frac{1}{100}$  comes from taking into account genetics.

## 2. Independence, as the court did, gets you

$$P(\text{both SIDS}) = (1/8500)(1/8500) = (1/73,000,000)$$

## 3. By Bayes rule

$$\frac{p(I|E)}{p(G|E)} = \frac{P(E \cap I)}{P(E \cap G)}$$

$P(E \cap I) = P(E|I)P(I)$  needs discussion of  $p(I)$ .

## Comparison

- ▶ Hence putting these two together gives the odds of guilt as

$$\frac{p(I|E)}{p(G|E)} = \frac{1/850,000}{1/1,000,000} = 1.15$$

In terms of posterior probabilities

$$p(G|E) = \frac{1}{1 + O(G|E)} = 0.465$$

- ▶ If you use independence

$$\frac{p(I|E)}{p(G|E)} = \frac{1}{73} \text{ and } p(G|E) \approx 0.99$$

The suspect looks guilty.

## Random Variables: Expectation $E(X)$

The **expected value** of a random variable is simply a weighted average of the possible values  $X$  can assume.

The weights are the probabilities of occurrence of those values.

$$E(X) = \sum_x xP(X = x)$$

With  $n$  equally likely outcomes with values  $x_1, \dots, x_n$ ,  
 $P(X = x_i) = 1/n$

$$E(X) = \frac{x_1 + x_2 + \dots + x_n}{n}$$

## Roulette Expectation

- ▶ European Odds: 36 numbers (red/black) + zero
- ▶ You bet \$1 on 11 Black (pays 35 to 1)
- ▶  $X$  is the return on this bet

$$E(X) = \frac{1}{37} \times 36 + \frac{36}{37} \times 0 = 0.97$$

- ▶ If you bet \$1 on Black (pays 1 to 1)

$$E(X) = \frac{18}{37} \times 2 + \frac{19}{37} \times 0 = 0.97$$

Casino is guaranteed to make money in the long run!

## Standard Deviation $sd(X)$ and Variance $Var(X)$

The **variance** is calculated as

$$Var(X) = E((X - E(X))^2)$$

A simpler calculation is  $Var(X) = E(X^2) - E(X)^2$ .

The **standard deviation** is the square-root of variance.

$$sd(X) = \sqrt{Var(X)}$$

## Roulette Variance

- ▶ European Odds: 36 numbers (red/black) + zero
- ▶ You bet \$1 on 11 Black (pays 35 to 1)
- ▶  $X$  is the return on this bet

$$Var(X) = \frac{1}{37} \times (36 - 0.97)^2 + \frac{36}{37} \times (0 - 0.97)^2 = 34$$

- ▶ If you bet \$1 on Black (pays 1 to 1)

$$Var(X) = \frac{18}{37} \times (2 - 0.97)^2 + \frac{19}{37} \times (0 - 0.97)^2 = 1$$

If your goal is to spend as much time as possible in the casino (free drinks): place small bets on black/red

## Example: $E(X)$ and $Var(X)$

**Tortoise and Hare are selling cars.** Probability distributions, means and variances for  $X$ , the number of cars sold

	0	1	2	3	Mean	Variance	sd
cars sold		$X$			$E(X)$	$Var(X)$	$\sqrt{Var(X)}$
Tortoise	0	0.5	0.5	0	1.5	0.25	0.5
Hare	0.5	0	0	0.5	1.5	2.25	1.5

# Expectation and Variance Calculations

Let's do Tortoise **expectations** and **variances**

- ▶ The Tortoise

$$E(T) = (1/2)(1) + (1/2)(2) = 1.5$$

$$\begin{aligned}Var(T) &= E(T^2) - E(T)^2 \\&= (1/2)(1)^2 + (1/2)(2)^2 - (1.5)^2 = 0.25\end{aligned}$$

- ▶ Now the Hare's

$$E(H) = (1/2)(0) + (1/2)(3) = 1.5$$

$$Var(H) = (1/2)(0)^2 + (1/2)(3)^2 - (1.5)^2 = 2.25$$

## Expectation and Variance Interpretation

What do these tell us above the **long run behavior?**

- ▶ Tortoise and Hare have the same expected number of cars sold.
- ▶ Tortoise is more predictable than Hare. He has a smaller variance. The standard deviations  $\sqrt{Var(X)}$  are 0.5 and 1.5, respectively
- ▶ Given two equal means, you always want to pick the lower variance.

# Linear Combinations of Random Variables

Two key properties:

Let  $a, b$  be given constants

- ▶ Expectations and Variances

$$E(aX + bY) = aE(X) + bE(Y)$$

$$Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$$

where  $Cov(X, Y)$  is the covariance between random variables.

## Tortoise and Hare Portfolio

What about Tortoise and Hare? We need to know  $Cov(\text{Tortoise}, \text{Hare})$ . Let's take  $Cov(T, H) = -1$  and see what happens

Suppose  $a = \frac{1}{2}$ ,  $b = \frac{1}{2}$  Expectation and Variance

$$E\left(\frac{1}{2}T + \frac{1}{2}H\right) = \frac{1}{2}E(T) + \frac{1}{2}E(H) = \frac{1}{2} \times 1.5 + \frac{1}{2} \times 1.5 = 1.5$$

$$Var\left(\frac{1}{2}T + \frac{1}{2}H\right) = \frac{1}{4}0.25 + \frac{1}{4}2.25 - 2\frac{1}{2}\frac{1}{2} = 0.625 - 0.5 = 0.125$$

Much lower!

# Bayesian Updating

## “Personalization” = “Conditional Probability”

- ▶ Conditional probability is how AI systems express judgments in a way that reflects their partial knowledge.
- ▶ Personalization runs on conditional probabilities, all of which must be estimated from massive data sets in which you are the conditioning event.

Many Business Applications!! Suggestions vs Search....

## Bayes's Rule in Medical Diagnostics

Alice is a 40-year-old women, what is the chance that she really has breast cancer when she gets positive mammogram result, given the conditions:

1. The prevalence of breast cancer among people like Alice is 1%.
2. The test has an 80% detection rate.
3. The test has a 10% false-positive rate.

The posterior probability  $P(\text{cancer} \mid \text{positive mammogram})$ ?

# Medical Diagnostics - Visualization

Of 1000 cases:

- ▶ 108 positive mammograms.  
8 are true positives. The remaining 100 are false positives.
- ▶ 892 negative mammograms.  
2 are false negatives. The other 890 are true negatives.

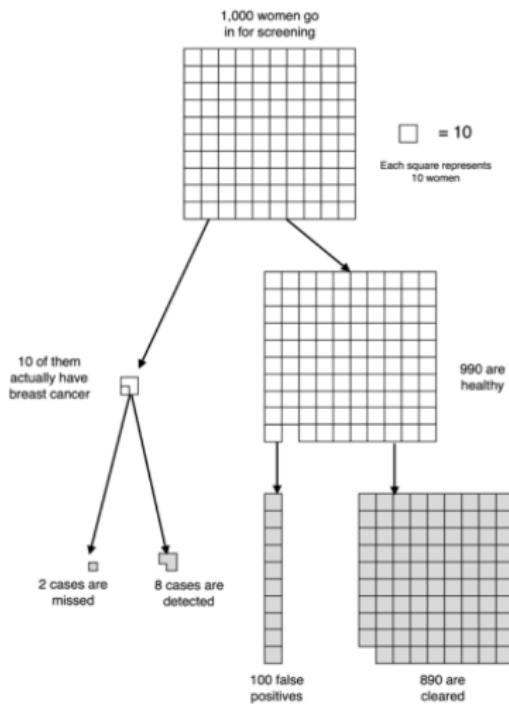


Figure 4: Medical Screening

# Test Marketing a New Product

## **Basic Problem:**

- ▶ Your company is developing a new product and will be test marketing to better gauge the sales of the new product.
- ▶ Based on positive, neutral or negative reactions, *what are the probability of high and low sales?*

Netflix Bayes and AI

## Test Marketing - Setup

Suppose **you** are given the following information

- ▶ New products introduced in the marketplace have high sales 8% of the time and low sales 92% of the time.
- ▶ A marketing test has the following accuracies: If sales are high, then consumer test reaction is positive 70%, neutral 25% and negative 5%. If sales are low, then consumer test reaction is positive 15%, neutral 35% and negative 50%.

## Test Marketing - Solution Steps

**Step 1:** Set-up your notation. Let

$$H = \text{high sales} \quad L = \text{low sales}$$

Pos = positive   Neu = Neutral   Neg = Negative

**Step 2:** List the known conditional probabilities For the marketing test we have

$$P(\text{Pos} | H) = 0.70, P(\text{Neu} | H) = 0.25, P(\text{Neg} | H) = 0.05$$

$$P(\text{Pos} | L) = 0.15, P(\text{Neu} | L) = 0.35, P(\text{Neg} | L) = 0.50$$

Finally, the base rates are  $P(H) = 0.08$  and  $P(L) = 0.92$

## Test Marketing - Calculation

**Step 3:** Describe the posterior probabilities that are required:

$$P(H \mid \text{Pos})$$

The probability of high sales given a positive marketing test.  
Compute the probability of a positive test

$$\begin{aligned} P(\text{Pos}) &= P(\text{Pos} \mid H)P(H) + P(\text{Pos} \mid L)P(L) \\ &= 0.70 \times 0.08 + 0.15 \times 0.92 = 0.194 \end{aligned}$$

## Test Marketing - Final Answer

Now use **Bayes Rule**

$$\begin{aligned} P(H \mid \text{Pos}) &= \frac{P(\text{Pos} \mid H)P(H)}{P(\text{Pos})} \\ &= \frac{0.70 \times 0.08}{0.194} = 0.288 \end{aligned}$$

Hence 28.8% you'll have high sales in the market.

We should interpret this **relative** to our initial probability of only 8%.

## “Personalization” = “Conditional Probability”

Conditional probability is how AI systems express judgments in a way that reflects their partial knowledge.

Personalization runs on conditional probabilities, all of which must be estimated from massive data sets in which you are the conditioning event.

Many Business Applications!! Suggestions vs Search, ....

## How does Netflix Give Recommendations?

Will a subscriber like Saving Private Ryan, given that he or she liked the HBO series Band of Brothers?

Both are epic dramas about the Normandy invasion and its aftermath.

100 people in your database, and every one of them has seen both films.

Their viewing histories come in the form of a big “ratings matrix”.

	Liked Band of Brothers	Didn't like it
Liked Saving Private Ryan	56 subscribers	6 subscribers
Didn't like it	14 subscribers	24 subscribers

$$P(\text{likes Saving Private Ryan} \mid \text{likes Band of Brothers}) = \frac{56}{56 + 14} = 80\%$$

## How does Netflix Give Recommendations? - Complexity

But real problem is much more complicated:

1. **Scale.** It has 100 million subscribers and ratings data on more than 10,000 shows. The ratings matrix has more than a trillion possible entries.
2. **“Missingness”.** Most subscribers haven’t watched most films. Moreover, missingness pattern is informative.
3. **Combinatorial explosion.** In a database with 10,000 films, no one else’s history is exactly the same as yours.

The solution to all three issues is careful modeling.

## How does Netflix Give Recommendations? - Fundamental Equation

The fundamental equation is:

Predicted Rating = Overall Average + Film Offset + User Offset + User-Film

These three terms provide a baseline for a given user/film pair:

- ▶ The overall average rating across all films is 3.7.
- ▶ Every film has its own offset. Popular movies have positive offsets.
- ▶ Every user has an offset. Some users are more or less critical than average.

## Netflix - Latent Features

The User-Film Interaction is calculated based on a person's ratings of similar films exhibit patterns because those ratings are all associated with a latent feature of that person.

There's not just one latent feature to describe Netflix subscribers, but dozens or even hundreds. There's a "British murder mystery" feature, a "gritty character-driven crime drama" feature, a "cooking show" feature, a "hipster comedy films" feature, ...

## The Hidden Features Tell the Story

These latent features are the magic elixir of the digital economy—a special brew of data, algorithms, and human insight that represents the most perfect tool ever conceived for targeted marketing.

Your precise combination of latent features—your tiny little corner of a giant multidimensional Euclidean space—makes you a demographic of one.

Netflix spent \$130 million for 10 episodes on *The Crown*. Other network television: \$400 million commissioning 113 pilots, of which 13 shows made it to a second season.