CS 6316 ~ ML

Probability Review

Review This Material On Your Own

We will review the mainly the last three sections

(Review questions included with this material)

-- Turn them in; Goes to: Participation Grade --

Outline

- Introduction
- Some Basic Concepts
- Probability Definition and Rules
- Conditional Probability & Bayes Theorem
- Random Variables
- Normal Distribution

1 - Introduction

Probability - Part of Life

- An Essential Part of Living – and ***Learning!***
- From the day we are born, we try to make sense of uncertain events around us
- Survival, success and progress depend on our ability to predict and plan for the future
- Extensively used in Statistics, Business, Forecasting, Machine Learning and hundreds of other areas: to make sense of existing information and draw conclusions for the future and as yet unknown situations and events

Examples from everyday life

- Getting Education
- Choosing a place to live
- Keeping your money with a bank or an investor
- Choosing a friend, a job or a pet

In our everyday life choices and decisions, we unconsciously or subconsciously, estimate the chance (probability) of being successful and / or happy…

Making sense of uncertain situations

- Most of the things and events of our lives are uncertain, particularly the future
- We try to estimate the chances of successful outcomes in business, medicine, weather, law, or success on our everyday projects

An interesting aside: *Abductive Reasoning*

- Abductive reasoning usually starts with an incomplete set of observations and proceeds to the likeliest possible explanation for the group of observations
- It is based on making and testing hypotheses using the best information available
- Often entails making an *educated guess* after observing a phenomenon for which there is no clear explanation

An interesting aside: *Abductive Reasoning*

- Useful for forming hypotheses to be tested

- Abductive reasoning is often used by
  - Doctors who make a **diagnosis** based on test results
  - Jurors who make **decisions** based on the evidence presented to them

Abduction: examples

- Example 1:
  - The library has many books
  - I have a book in my hand
  - Therefore, the book was taken from the library
  - *((is this logic sound?))*
- Example 2:
  - When it rains, the grass gets wet
  - The grass is wet
  - Therefore, it must have rained
  - *((is this logic sound?))*

Abduction: Sherlock Holmes

- But by far the best example of abduction comes from Sherlock Holmes. Contrary to popular opinion – Sherlock Holmes rarely, if ever, deduces anything! He may occasionally induce something, but most of the time he infers the best explanation from his observations.

2 – Some Basic Concepts

Fundamental terminology and concepts

Experiment- Defined

- An **experiment** is the process by which observations (or measurements) are taken
- In probability theory, an experiment is any procedure that can be repeated and has a set of possible outcomes, known as the **sample space**

Trial, Outcome and Sample Space

- For example, if one were to toss the same coin one hundred times, each toss would be considered a **trial** within the experiment composed of all hundred tosses
- The result of each trial is an **outcome**. From the above experiment of 100 trials, we might get 30 outcomes as Heads and rest of the 70 outcomes as Tails
- The 100 outcomes of all the 100 trials is the **Sample Space (S)**

An Event

- In probability theory, an **event** is a set of outcomes of an experiment (a subset of the sample space) to which a probability is assigned
- Often, the term **event** is the name given to the collection of possible outcomes that are the focus of the experiment
- When an experiment is performed, a particular event either happens, or it doesn't!

Simple Event

- Any particular single outcome of an experiment is known as a **simple event** (sometimes called as outcome).
- Relation between these terms are presented on the next slide

Relationship between 4 concepts

Relation between an experiment, sample space, an event and a simple event:

- **Experiment**: we toss a coin three times
- Sample Space: All possible simple events: total of all possible outcomes
- **Sample space** = 8 = 8 **simple events**:

  HHH, HHT, HTH, THH, HTT, THT, TTH, TTT
- **Event (E)**: getting exactly 2 heads
- P(E) = number of *favorable* outcomes (2 heads)

  ÷ sample space = 3/8 = 0.375

Random and Deterministic Event

- A **random event** is something unpredictable. As it is unpredictable, you can never give it an exact value. For example, you couldn’t predict whether you will get a Head in tossing of a coin.
- The opposite of random is **deterministic**, which means it can be calculated exactly. For example, your height can be calculated to the nearest inch, so height is deterministic.

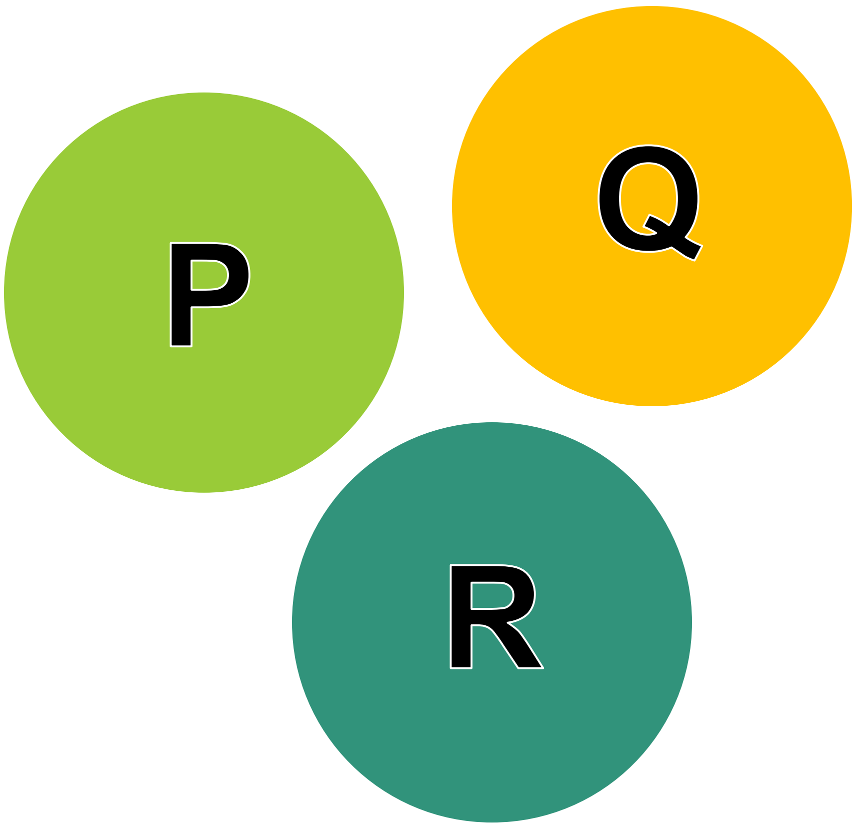
Mutually Exclusive Events

- Two events are **mutually exclusive** if, when one event occurs, the other cannot, and vice versa. Here are some examples:
- H and T in Tossing of a coin
- Numbers 1 to 6 in throwing a die
- Raining at UVA, or Not raining at UVA
- Possible final grade of a student in a particular course. If a student has final grade “A”, s/he cannot also have a “B” for the same course

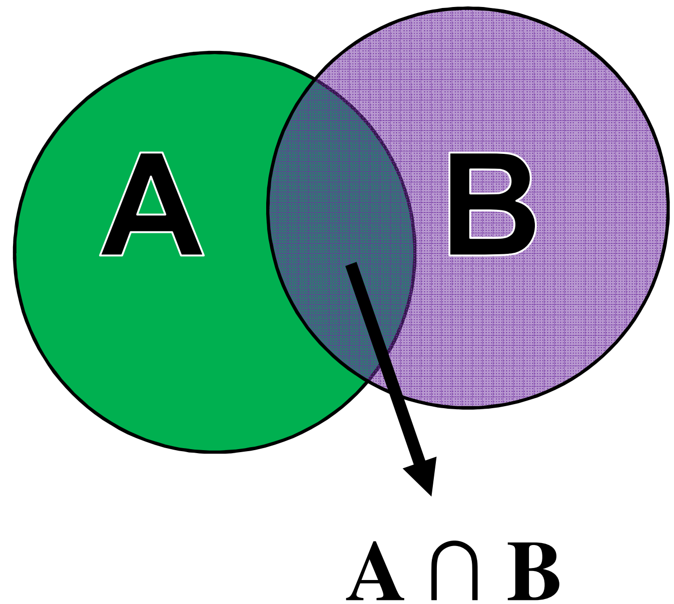
Inclusive Events

- **Mutually inclusive** events are the ones in which there are some common outcomes between the given events
- Inclusive events are the events that can happen at the same time (not mutually exclusive)
  - [A] We randomly pick an animal from a shelter and it is black in color
  - [B] We randomly pick an animal from a shelter and it is a cat
- A and B are inclusive because we may pick a black cat from the shelter

Graphic Illustrations



Three mutually exclusive events



Two inclusive events

Independent Events

- Two events, A and B, are **independent** if the fact that A occurs does not affect the probability of B occurring
- Some examples:
  - <u>Tossing a coin</u>: We toss a coin and get "Head." This outcome has no effects on what we get for the second tossing of the coin
  - <u>Rolling a Die</u>: The number we get on the first rolling of a die has no impact on the number we might get on the second rolling

Non-Independent Events

- **Non-independent** events are the events that may occur together or occurrence of an event may increase or decrease the chance of occurrence of the other event
  - Example 1: If we pull a marble at random from a bag containing 8 black and 2 red marbles, and it came red, the chance of getting a 2nd red marble has reduced
  - Example 2: Two events, severe snowstorm today (A) and university will be closed tomorrow (B) are not independent events

Null Event

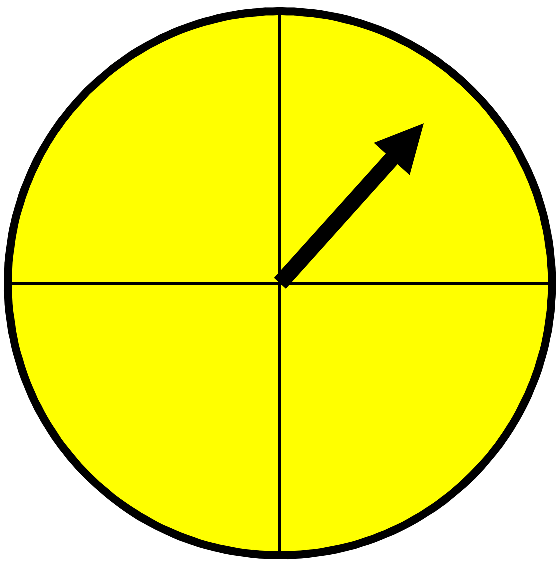
- A **null event** is an event that is impossible. Or more precisely, since an event is a subset of a sample space, the null event is the **empty set**
- **P(A) = 0** is not always a null event
  "*My best friend is not going to eat an Apple tomorrow- he never does*"
  (Highly unlikely but not impossible)

Certain Events

- A **certain event** is an event that will always occur (probability = 1) Examples:
  - If there is a bag full of coins (S = all coins in the bag), and you pull one coin then P(It's a coin) = 1. It is because the event we have selected has the same possibilities as possibilities of sample space (S)
  - If it is Thursday, the probability that tomorrow is Friday is **certain**, and therefore the probability is 1

Certain Event

- **Certain event**: the spinner will land on a yellow region (Probability is 1)



Special Notes on Sample Space

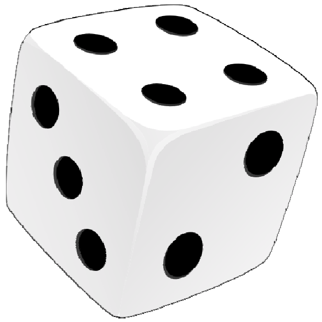
- It is important to clarify a few points about the sample space:
- Sometimes **sample space (S)** is also called "all possibilities". "total possibilities", "All possible outcomes" or "*universe*"
- Even for the same subject, "S" can mean different things, depending on what are the "total possibilities" that are part of the experiment
- The sample space is the full set of possibilities for any particular study / experiment

Special Notes on Sample Space

- Let us assume that a zoo has 1000 animals, including lions, tigers, jaguars, and 100 reptiles.
  Reptiles includes: crocodiles (5), lizards (7), and pythons (10)
  Out of 10 Pythons: 2 are Burmese Pythons
- Note (below) how the concept of **S** changes, with the nature of the question or problem under consideration:
- Q1: If an **animal** escapes from the zoo, P(animal is a reptile)?
  S = **1000** = all animals [probability of x=P(x)]
- Q2: If a **reptile** leaves the zoo,
  P(reptile is a Python)? S = **100** = all reptiles
- Q3: If a **python** leaves the zoo,
  P(python is a Burmese python)? S = **10** = all pythons

Examples of Sample Space

- One throw of 1 die:
  - S = 6 = 6 outcomes
  - S = (1, 2, 3, 4, 5, 6)



- One throw of 2 dice:
  - S = 6 * 6 = 36 outcomes
- One throw of 3 dice:
  - S = 6 * 6 * 6 = 1296 outcomes
- One throw of 2 coins:
  - S = 4 (**HH, HT, TH, TT**)

3 – Probability Definition and Rules

Definition of Probability

- **Probability** is the likelihood / chance of something happening or being the case. It is expressed by the **ratio** of the number of actual occurrences to the total number of possible occurrences
- Probability of an event **(E)** happening is always from **zero to one** (both inclusive): **$0 \leq P(E) \leq 1$**

Probability is Often an Estimate

- The probability of an event A measures “how often” A will occur. We write **P(A)**, compared to total possibilities
- Suppose that an experiment is performed *n* times. The relative frequency for an event A (compared to total frequency) is probability of A occurring, if we let *n* get infinitely large However approximate estimates are made with *less than infinite occurrences*

Simple Formula of Probability

$$\color{red}{Probabilty\ of\ an\ Event} = \frac{Number\ of\ favorable\ outcomes}{Total\ number\ of\ possible\ outcomes}$$

- Example: A shelter has a number of cats, if we pick a cat random, what is the probability that the cat will be ginger?

$$\frac{\#\ ginger\ cats}{\#\ of\ all\ the\ cats}$$

- **Favorable outcomes** is a shorter way of saying "outcomes of the event under study" (Number of ginger cats in above example)

A Simple Example – 1

- A bag contains 2 Artificial Intelligence (AI) books and 6 Machine Learning (ML) books. One book is picked at random. What is the probability that it is a ML book?
  - Total # of books = sample space (s) = 8
  - ML books = Event (E) = # of outcomes = 6
  - Probability = # of outcomes in E / # in S
  - Answer: E / S = 6 / 8 = 3 / 4 = 0.75
    (Using set theory: Elements of set E / Elements of set S)

    E is always a subset of S that is E <= S

A Simple Example – 2

- A die is rolled, what is the probability that we get 5 or 6?
  - Total # of possibilities of roll of a die = sample space (s) = elements of set (S) = 6
  - Event (E) = # of outcomes (5 OR 6) = elements of subset E = 2
  - Probability = # of elements in E / # of elements in set S
  - Answer: E / S = 2 / 6 = $^{1}/_{3}$

Discrete and Continuous Probability

- **Discrete Probability**: finite number of outcomes – like outcomes of tossing a coin or rolling a die
- **Continuous Probability**: outcomes vary along continuous scale – like weight of a person, length of a table or water depth of a river
  - Note: *"infinite" level of precision on a continuous scale*

Some Important Rules

- **$P(A^C)$ = P(~A)** = 1-P(A) = P(**compliment** of A) = P(not A)
  P(A or ~A) = P(A) + P(~A) = 1 ["not" sometimes written **¬**]
- If A and B are **Mutually exclusive**: (never happen together)
  - **P (A or B)** = P(A) + P(B) [Addition Rule]
  - **P (A and B)** = **0** *(can <u>never</u> roll a 1 AND 6 with a single die)*
- If A and B are **independent**: (may happen together)
  - **P (A and B)** = P(A ∩ B) = P(A) * P(B) [Multiplication Rule]
  - **P (A or B)** = P(A) + P(B) - P(A ∩ B)
    (subtract intersection so items aren't counted twice)
- If two events A and B are independent then **Conditional Probability**:
  - Probability of event A, when event B has occurred: **P(A|B)**
  - **P(A|B)** = *P(A, given B)* = **P(A ∩ B) / P(B)** (for P(B) > 0)
  - Remember: P(A|B) ≠ P(B|A)!

Example: Mutually Exclusive Events

- RULE: P(A or B) = P(A) + P(B)
- A pet shop has 100 cats: 85 are Tabby, 10 Siamese and 5 Scottish Fold. One cat is adopted. What is the probability that the cat is not Tabby?

  Can derive this in more than one way:
  - P(not Tabby) = 1 – P(Tabby)

    = 1 – (85/100) = **0.15**
  - P(not Tabby) = P(Siamese) + P(Scottish Fold)

    = 0.1 + 0.05 = **0.15**

Example: Independent Events

- P (A and B) = P(A ∩ B) = P(A) * P(B)
  [Multiplication Rule]
- Assume two independent events A and B
- Two events are independent when the outcome of the first event does not influence the outcome of the second event
- Probability of getting 6 <u>and</u> 6 by throwing 2 dice
- **P(A and B) = P(A) * P(B)** = P(get a 6) * P(get a 6)
- = 1/6 * 1/6 = **1/36**

Example: Independent Events

- $P(A \text{ or } B) = P(A) + P(B) - P(A \cap B)$
- What is the probability of drawing a **black** card <u>or</u> a **ten** in a deck of cards?
  - 26 black cards P(black) = 26/52
  - 4 tens in a deck of cards P(10) = 4/52
  - 2 black tens P(black and 10) = 2/52
- P(black or ten) = P(black)+P(10) – P(black 10s) = 26/52 + 4/52 – 2/52 = 30/52 – 2/52 = 28/52 = **7/13**

4 – Conditional Probability, Total Probability, and Bayes Theorem

***** Important section! *****

Conditional Probability

- Conditional probability concerns the odds of one event occurring, given that another event has occurred

- Expressed as:

$P(A|B)$ = probability of A, given B

$= \frac{P(A \cap B)}{P(B)}$

Conditional probability

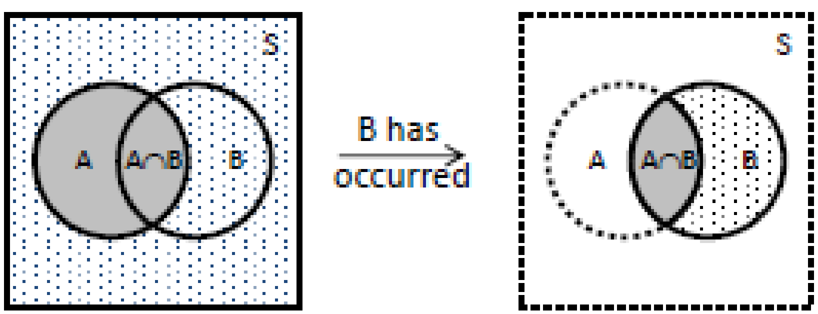
– If A and B are two events, the probability of event A when we already know that event B has occurred is

$$P[A|B] = \frac{P[A \cap B]}{P[B]} \quad if \;\; P[B] > 0$$

- This conditional probability P[A|B] is read:
  – the "conditional probability of A conditioned on B", or simply
  – the "probability of A given B"

– Interpretation

- The new evidence *"B has occurred"* has the following effects
  - The original sample space S (the square) becomes B (the rightmost circle)
  - The event A becomes $A \cap B$
- P[B] simply re-normalizes the probability of events that occur jointly with B



Conditional Probability

- Example 1: In a class of 100 students, 40 bought Desktops, 30 bought Notebooks, and 20 bought a Desktop and a Notebook. If a student chosen at random bought a Desktop, what is the probability that s/he also bought a Notebook?
  - P(bought Notebook | bought Desktop)?
  - P(bought Desktop) = P(D) = 40/100 = 0.4
  - P(bought both) = $P(D \cap N)$ = 20/100 = 0.2
  - So… $P(N|D) = P(D \cap N) / P(D) = 0.2 / 0.4 =$ **0.5**

Conditional Probability

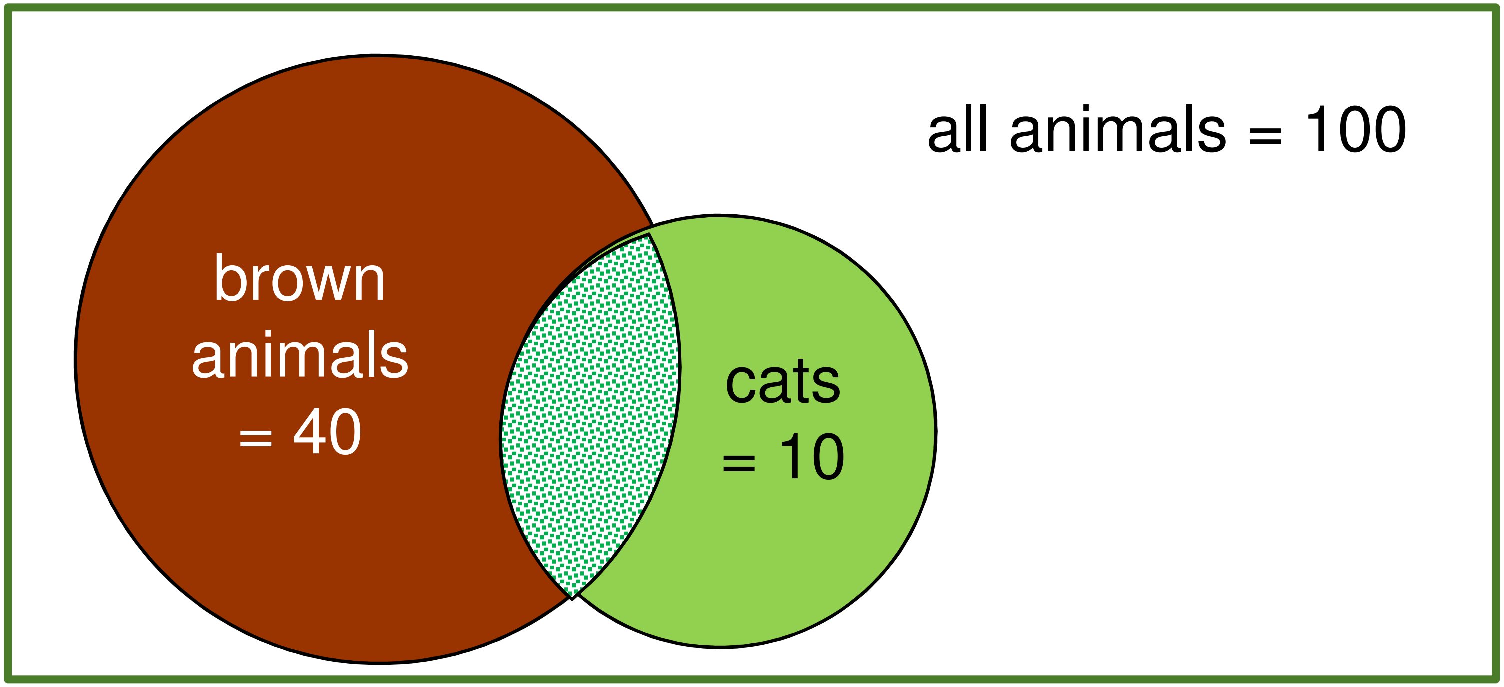
- Example 2: Let us assume that an animal shelter has 100 animals with 40 of them being brown in color
- From 100 animals, 10 are cats and from these 10, 8 cats are brown in color. An animal ran away from the shelter
- What is the probability that the animal was a cat?
- If it is known that the animal was <u>brown</u>, what is the probability that the animal was a cat?

Conditional Probability

- All animals =100, Brown animals = 40
  Cats = 10: Out of 10, Brown cats = 8
- P(Animal that ran away was a cat)?
- P(Animal that ran away was a cat|Animal was Brown)?
- P(A) = Animal was a cat: 10 / 100 = 1/10 = 0.1
- P(B) = Animal was brown: 40/100 = 4/10 = 0.4
- P(A∩B) = Brown cat = 8 / 100 = 0.08
- Conditional Probability: P(A|B) = P(A∩B) / P(B)
  = 0.08 / 0.4 = **0.2** (8 brown cats / 40 Brown animals)
  (Please Notice the change from 0.1 to 0.2)

Conditional Probability: Find Needle in a Haystack

- Find a needle (dotted area) in brown circle (in a smaller haystack) rather than from a rectangle (in a bigger haystack)
  Dotted area (intersection): Area representing **brown cats**
  Left (brown) circle: Area representing **brown animals**



Total Probability (TP)

- **Total probability** expresses the total probability of an outcome which can be realized via several distinct events—hence the name
- Total Probability breaks up probability calculations into distinct parts. It's used to find the probability of an event, A, when you don't know enough about A's probabilities to calculate it directly. Instead, you take a related event, B, and use that to calculate the probability for A

Total Probability

Some simple expressions of TP:

- ~A = anything other than A:
- P(~A) = 1- P(A)
- P(A or ~A) = P(A) + P(~A) = 1
- $P(A) = P(A \cap B_1) + \ldots + (A \cap B_n) = 1$
- $P(A) = P(A \cap B) + P(A \cap \sim B) = 1$

- If a pet shop has 80 white cats and 20 black cats, then the probability that **a cat** picked at random is black or not black = 1

Total Probability

- A pet shop has 100 animals = Sample space = (S):
  Dogs 40 (B_1), Cats 30 (B_2), Rabbits 20 (B_3), Birds 10 (B_4)
- Let us call an event A, "brown" colored pet
- $A = A \cap S = A \cap (B_1 \text{ or } B_2 \text{ or } B_3 \text{ or } B_4)$
  $= (A \cap B_1) \text{ or } (A \cap B_2) \text{ or } (A \cap B_3) \text{ or } (A \cap B_4)$
- Since we know, B_i are mutually exclusive, then

$P(A) = P(A \cap B_1) + (A \cap B_2) + (A \cap B_3) + (A \cap B_4)$

"BROWN DOGS" + "BROWN CATS" + "BROWN RABBITS" + "BROWN BIRDS"

=> $P(A|B_1)\,P(B_1) + P(A|B_2)\,P(B_2) + P(A|B_3)\,P(B_3) + P(A|B_4)\,P(B_4)$

Dogs 40	**Cats 30**	**Rabbits 20**	**Birds 10**

Theorem of total probability

- Let $B_1, B_2 \dots B_N$ be a partition of S (mutually exclusive that add to S)
- Any event A can be represented as

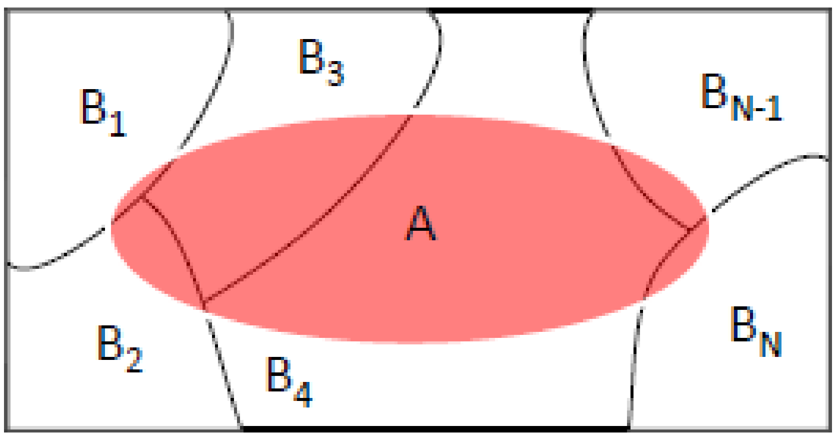
$$A = A \cap S = A \cap (B_1 \cup B_2 \dots B_N) = (A \cap B_1) \cup (A \cap B_2) \dots (A \cap B_N)$$

- Since $B_1, B_2 \dots B_N$ are mutually exclusive, then

$$P[A] = P[A \cap B_1] + P[A \cap B_2] + \cdots + P[A \cap B_N]$$

- and, therefore

$$P[A] = P[A|B_1]P[B_1] + \cdots P[A|B_N]P[B_N] = \sum_{k=1}^{N} P[A|B_k]P[B_k]$$



CSCE 666 Pattern Analysis | Ricardo Gutierrez-Osuna | CSE@TAMU

Bayes Theorem

- Assume {B_1, B_2, … B_N) is a partition of S
- Suppose that event A occurs
- What is the probability of event B_j?
- Using the definition of conditional probability and the theorem of total probability we obtain:

$$P\left[B_j \middle| A\right] = \frac{P[A \cap B_j]}{P[A]} = \frac{P\left[A \middle| B_j\right] P[B_j]}{\sum_{k=1}^{N} P[A|B_k] P[B_k]}$$

- This is known as **Bayes Theorem**/Rule, and is *(one of) the most useful relations in probability and statistics!*

Bayes Theorem

Bayes theorem and statistical pattern recognition

- When used for pattern classification, BT is generally expressed as

$$P[\omega_j|x] = \frac{p[x|\omega_j]P[\omega_j]}{\sum_{k=1}^{N} p[x|\omega_k]P[\omega_k]} = \frac{p[x|\omega_j]P[\omega_j]}{p[x]}$$

  - where ω_j is the j-th class (e.g., phoneme) and x is the feature/observation vector (e.g., vector of MFCCs)
- A typical decision rule is to choose class ω_j with highest $\mathrm{P}[\omega_j|x]$
  - Intuitively, we choose the class that is more "likely" given observation x
- Each term in the Bayes Theorem has a special name
  - $P[\omega_j]$ prior probability (of class ω_j)
  - $P[\omega_j|x]$ posterior probability (of class ω_j given the observation x)
  - $p[x|\omega_j]$ likelihood (probability of observation x given class ω_j)
  - $p[x]$ normalization constant (does not affect the decision)

Common Applications of Bayes Theorem

- Sometimes we know the conditional probabilities of a set of events, such as:

  P(defective heat shield| factory 1)
  P(defective heat shield| factory 2)
  P(defective heat shield| factory 3)

  …and we want to calculate from these **known values**, some unknown value that is *opposite* of these known values, for example:
  Given a defective shield (HS), what's the probability that it came from factory 2? = P(factory2 | HS)?

The next few slides illustrate situations where the condition stated above may exist

Bayes Theorem Applications

- Situation 1: three factories produce note book computers and we know the conditional probability of defective computers
- If a defective computer shows up in a repair shop, we may like to know what is the probability that it came from a particular factory
- **Known:** P (Defective| factory1), P (Defective| factory2), P (Defective| factory3)
- **Find:** P(from a particular factory | Defective)

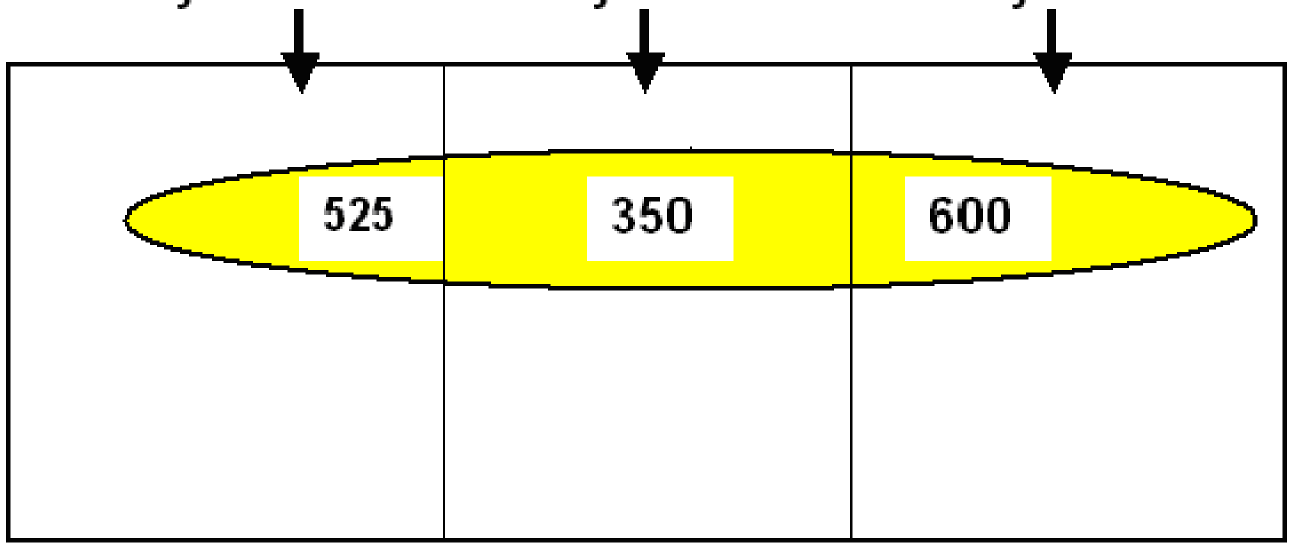
Bayes Theorem Applications

- Situation 2: a country is divided into 4 states. We know the percentage of Diabetic patients (thus the probabilities) in each state
- If a Diabetic patient shows up in the emergency room, we may like to know what is the probability of this patient being from a particular state

Bayes Applications – Example 1 – Defective Chair Example

Production of 3 factories | Factory A = 35K | Factory B = 35K | Factory C = 30k

\# Defective



% of total Prod	35%	35%	30%
Prob (from factory)	0.35	0.35	0.30

Bayes Applications – Example 1 – Defective Chair Example

- A chair produced by ***TopWood*** Company was found to be defective **(D)**. There are **three factories (A, B, C)** where such chairs are produced. An inspector is responsible for investigating the sources of defects. This is what the inspector knows about the company's chair production and the possible sources of defects:

Factory	Ratio of total production	Pr of defective
A	0.35 = P(A)	0.015 = P(D | A)
B	0.35 = P(B)	0.010 = P(D | B)
C	0.30 = P(C)	0.020 = P(D | C)

- If a randomly selected chair is defective, the Inspector wants to know:
  What is the probability that the chair was manufactured in factory C?

Bayes Applications – Example 1 – Defective Chair Example

- **P(defective chair was manufactured in factory C)?**
  Find P(C|D)
  Rule of *Conditional Pr*: P(B|A) = P(A∩B) / P(A)

- **P(C|D)** = P(C∩D) / P(D) = P(D|C)*P(C) / P(D)
- = P(D|C)*P(C) / P(D∩A) or P(D∩B) or P(D∩C)
- **Since being defective is mutually exclusive in 3 factories:**
- = P(D|C)*P(C) / {P(D|A)*P(A) + P(D|B)*P(B) + P(D|C) * P(C)}
- = (0.02) (0.30) / (0.015)*(0.35) + (0.01)*(0.35) + (0.02)*(0.30) = 0.006 / 0.01475 = **0.407**

Bayes Theorem (BT)- Example 2

- Bayes theorem is applied in many professions
- Some interesting application are in the Medical profession
- Following few slides illustrate the application of Bayes Theorem in Healthcare diagnostics

BT Example 2 - Terminology

- **Positive Test Result**: If the result of a medical test shows that the person has the disease, the result is considered positive
- **False positive**: As we know, Science is not perfect, sometimes a person is perfectly healthy but the results come positive, it's called False Positive
- **True Positive**: When a person truly has the disease and lab test comes positive, it's called True Positive

BT Example 2 - Terminology

- **Negative Test Result**: If the result of a medical test shows that the person does not have the disease, the result is considered negative
- **False Negative**: When a person does have the disease but the test came negative (no disease)
- **True Negative**: When a person is healthy and the lab result confirmed that the person is healthy (remember: negative=no disease)

BT Example 2 - Terminology

Specificity of a Test (%)

- Medical tests are rarely 100% accurate
- The *true-negative* **rate P(NEG | ¬COND)** of a test is called its **SPECIFICITY**

- *Rate at which the test correctly detects that the patient* ***does not*** *have the condition (NEG)*

BT Example 2 - Terminology

Sensitivity of a Test (%)

- The *true-positive* **rate P(POS | COND)** of a test is called its **SENSITIVITY**

- *Rate at which the test correctly detects that the patient **does** have the condition (POS)*

BT Example 2 – the problem

- John Doe lives in a small town where **1%** of the people have ***Colon Cancer*** (**prevalence**). We have used a test to test John for colon cancer. This test has:

  90% Sensitivity and **98% Specificity**

- John's test result came back **positive** (**+**) (*showing that he has Colon Cancer*)
- **QUESTION:** What is the probability that John, *really* does have the Colon Cancer?

★BT Example 2– Find the Answer

P(John has the condition "colon cancer" given the test came back positive) $= P(cond \mid +)$

$$= \frac{P(+|cond)P(cond)}{P(+)} =$$

Continue working on this problem. Use BAYES THEOREM.

What is the answer? You may work with a partner.

5 – Random Variables

…including…

Cumulative Distribution Functions (CDF)

Probability Density Functions (PDF)

Random Variables

- When the value of a variable is the outcome of a statistical experiment, that variable is a **random variable**
- Example: if we toss a coin twice, this experiment has four possible outcomes: HH, HT, TH, TT. Let variable X represent the number of heads that result from this experiment.
  Variable X can take on the values 0, 1 or 2:
  0: (TT);   **1**: (HT or TH);   **2**: (HH)
- In this example, X is a random variable; because its value is determined by the outcome of a statistical experiment

Random Variables

- The function that assigns values to each outcome is **deterministic**. This means that each time you provide the function with the same parameter values, the result is the same
- For example if X is the number of times a coin is tossed, every time X = 2, we get the same 4 possible outcomes (HH, HT, TH,TT).

Discrete Random Variables

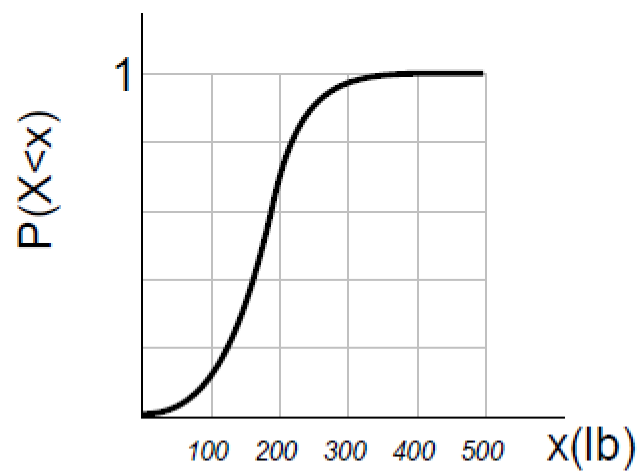
- A **discrete variable** is a variable which can only take a countable number of values. In example of tossing a coin twice, the number of heads can only take 3 values (0, 1, 2) and so the variable is discrete
- Similarly, if we are making plastic buckets and each pound of plastic makes 2 buckets; then buckets produced will always be countable whole numbers: 10 pounds of plastic will always produce 20 buckets

Continuous Random Variables

- A **continuous random variable** differs from a discrete random variable in that it takes on an uncountable infinite number of possible outcomes
- Height of a tree, weight of a human being or an animal, depth of water in a river and amount of rainfall in an area are some of the examples of continuous variables

Cumulative Distribution Function

- In probability theory, the **Cumulative Distribution Function (CDF)**, $F_X(x)$, of a **real-valued** random variable X, evaluated at x, is the probability that X will take a value less than or equal to x (event $\{X \leq x\}$)
- $F_X(x) = P[X \leq x]$
- Note: that x is a fixed value
- Graph *always increasing*
- *(e.g. X is weight; x is 180lb)*



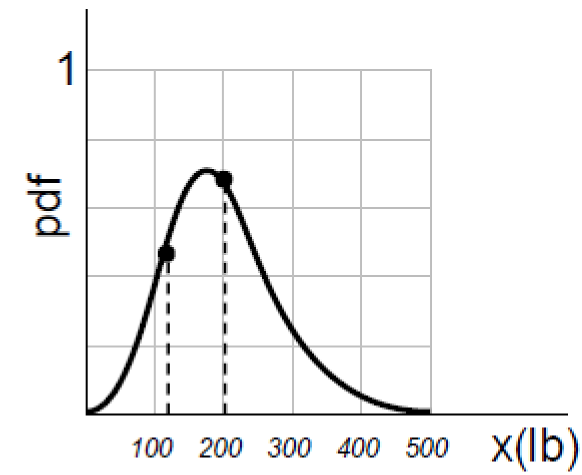
cdf for a person's weight

Probability Density Function

- The **Probability Density Function (PDF)**, $f_X(x)$, of a **continuous** random variable X, if it exists, is defined as the derivative of $F_X(x)$

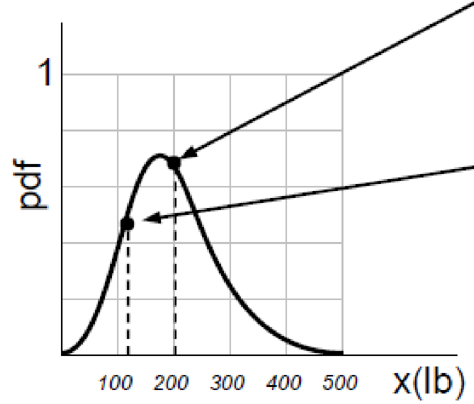
$$f_X(x) = \frac{dF_X(x)}{dx}$$

- Given by the area under the PDF
- $P[a < x < b] = \int_a^b f_X(x)dx$
- $1 = \int_{-\infty}^{\infty} f_x(x)dx$



pdf for a person's weight

Probability Density Function



pdf for a person's weight

- **What is the probability of somebody weighting 200 lb?**
  - According to the pdf, this is about 0.62
  - This number seems reasonable, right?
- **Now, what is the probability of somebody weighting 124.876 lb?**
  - According to the pdf, this is about 0.43
  - But, intuitively, we know that the probability should be zero (or very, very small)
- **How do we explain this paradox?**
  - The pdf DOES NOT define a probability, but a probability DENSITY!
  - To obtain the actual probability we must integrate the pdf in an interval
  - So we should have asked the question: what is the probability of somebody weighting 124.876 lb plus or minus 2 lb?

Calculating PDF (in an interval)

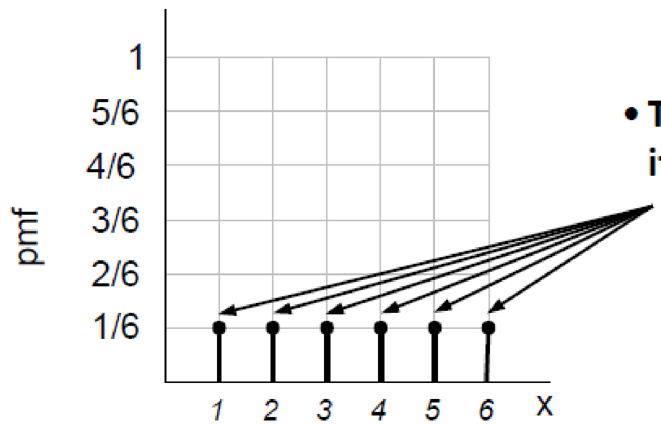
- The Probability Density Function (PDF), $f_X(x)$, of a **continuous** random variable X, such that for any two numbers a and b with $a \leq b$,

$$P[a < x < b] = \int_a^b f_X(x)dx$$

- That is, the probability that x takes on a value in the interval $[a,b]$ is the area above this interval and under the graph of the density function

Probability Mass Function

- However, for **discrete** random variables, the equivalent to the PDF is the **Probability Mass Function (PMF)**.
- A probability mass function (PMF) is a function that gives the probability that a discrete random variable is exactly equal to some value



pmf for rolling a (fair) dice

- **The probability mass function is a 'true' probability (reason why we call it a 'mass' as opposed to a 'density')**
  - The pmf is indicating that the probability of any number when rolling a fair dice is the same for all numbers, and equal to 1/6, a very legitimate answer
  - The pmf DOES NOT need to be integrated to obtain the probability (it cannot be integrated in the first place)

6 – Normal Distribution and Normal Curve

Gaussian Distribution

Normal Curve

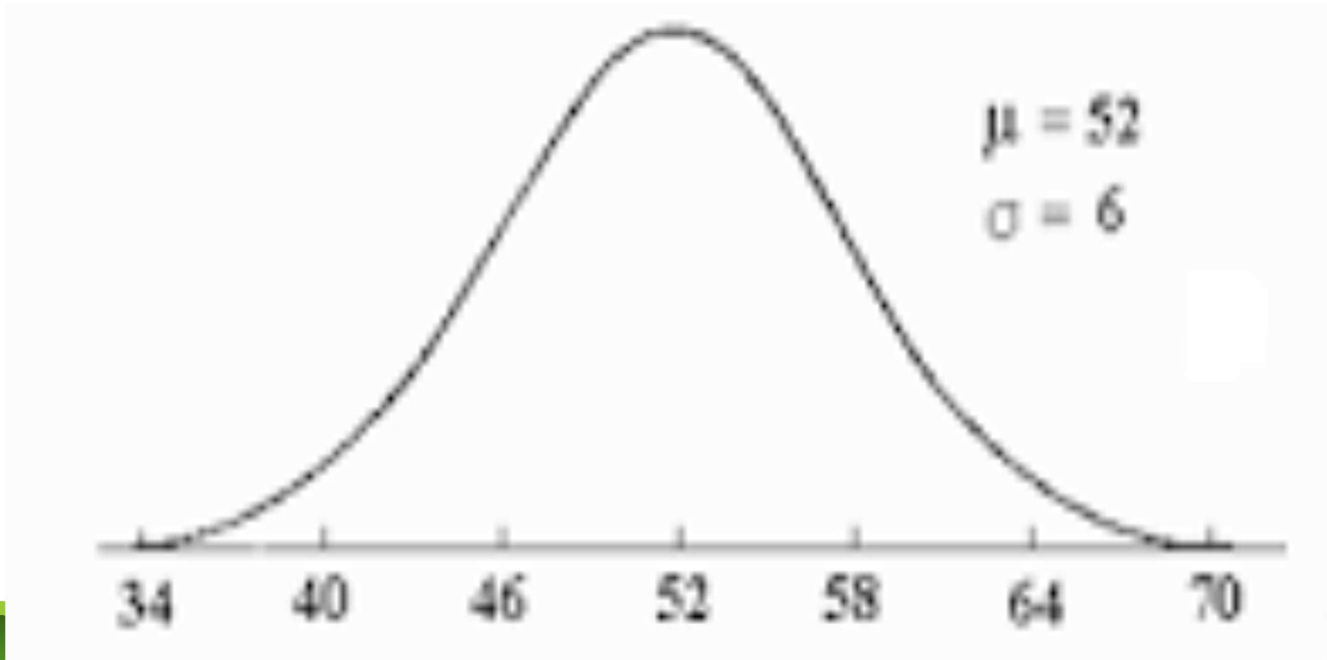
- **Normal Curve** is a bell shaped curve showing a particular distribution of probability over the values of a random variable. Also called Gaussian curve, or probability curve.
- Many measurement variables found in nature follow a predictable pattern with symmetry where much of the distribution of the data is clumped around the center and a few observations are found on the extremes. Data that has this pattern are said to be **bell-shaped or having a normal distribution**

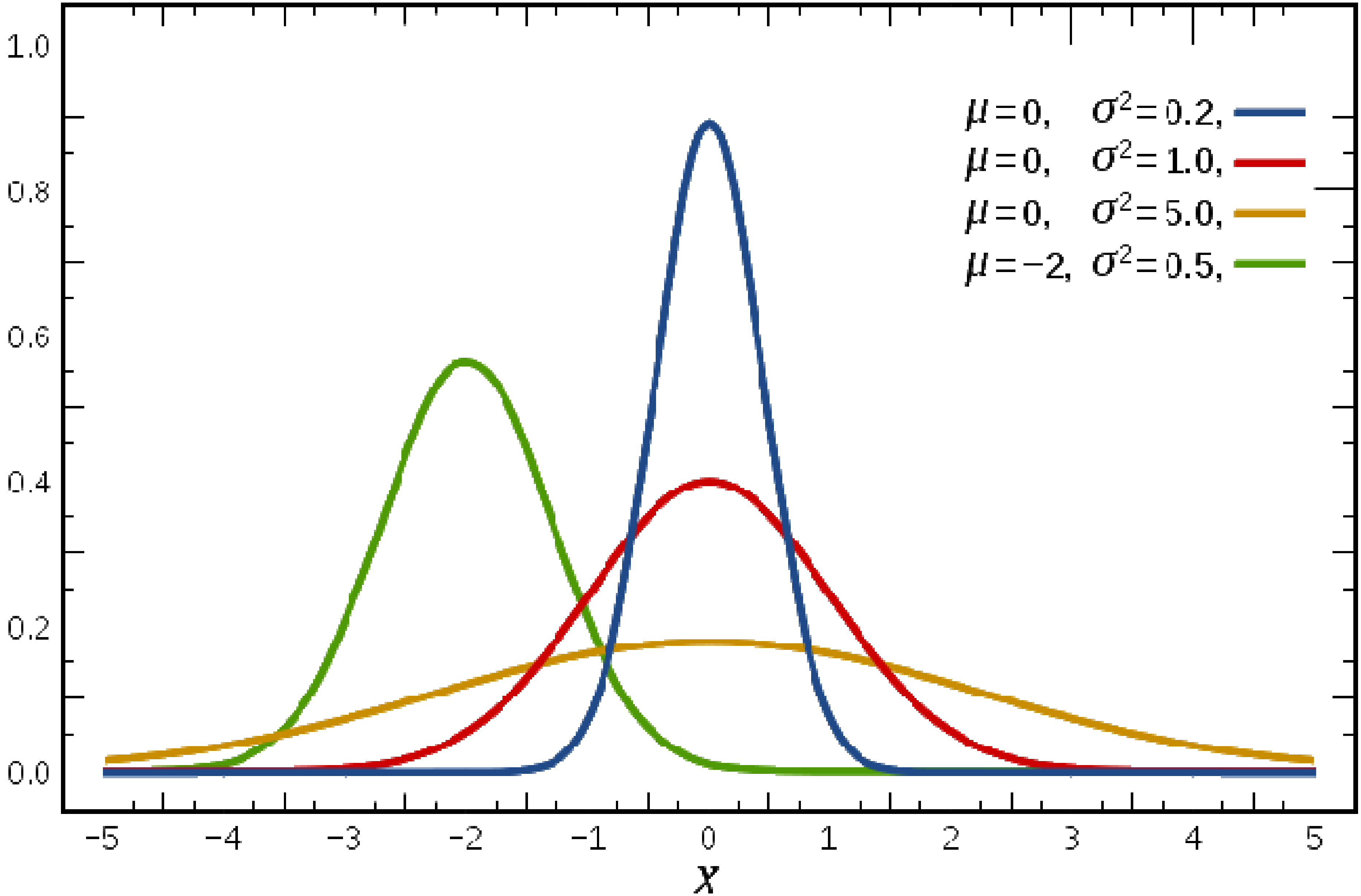
The Gaussian Distribution (GD)

- The **Gaussian distribution** is a continuous function. The Gaussian distribution is also commonly called the "normal distribution" and sometimes as Z distribution

The Gaussian Distribution Curve

- The following is a bell shaped curve with mean 52 and standard deviation 6. A Gaussian Curve can be fully defined by Mean and Standard Deviation of the data represented by the curve





https://en.wikipedia.org/wiki/Gaussian_function

More about Gaussian Distribution

- The normal (Z) distribution is a continuous distribution that arises in many natural processes. "Continuous" means that between any two data values we could (at least in theory) find another data value

- The bell-shaped normal curve has probabilities that are found as the area between any two z values

Example: Heights of Women

- For example, the heights of Women (or Men) in any country vary continuously and are the result of so many tiny random influences that the overall distribution of women's (and Men’s) heights in America are very close to normal

Example: Repeated values of weight

- If 100 people buy 1 kg to tomatoes each from the same grocery store, measured on the same scale there, when measured (in grams) on a very sensitive scale, all 100 values will not be 1000 grams. Most of the figures will be close to 1000 grams, a few will be much lower or much higher than 1000 grams. The frequencies drawn by a graph will be bell shaped

Measurements of Circumference

- Let 1000 individuals measure the circumference of the trunk of a tree in Millimeters.
- When these 1000 figures are plotted, the graph of the frequency distribution will be very close to bell shaped (Normal Curve)

Normal Distribution - Applications

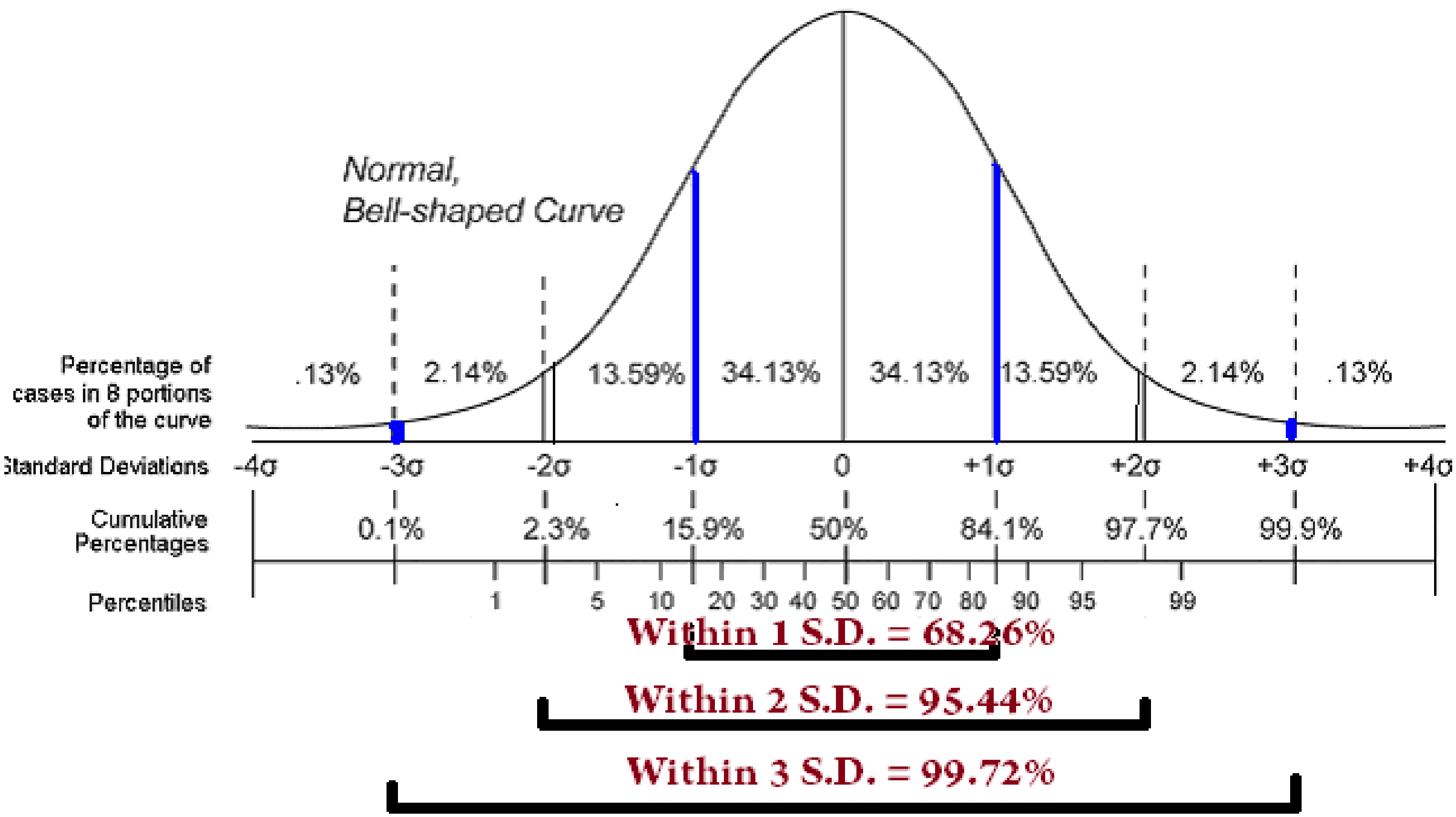
- The normal distribution (ND) is most familiar and useful to teachers in describing the frequency of standardized test scores. Significance of most research analysis techniques and the level of confidence in forecasts make frequent use of normal distribution
- *The things you can say about a ND, can now apply to your data → can better forecast probabilities*

Some characteristics of ND

Normal Distribution is not any distribution, but one with several unique characteristics:

- A normal curve for a normal distribution can be drawn by using just 2 parameters: Mean and Standard Deviation
- It is always **symmetrical**, with equal areas on both sides of the curve
- The highest point on the curve corresponds to the **mean value**
- The area between given standard deviation units includes a determined percent area

Key Properties of ND



Some Properties of ND

1 - The **mean** value is at the **center** of the curve

2 - Half of the values are on the left and half on the right

3 - About **68%** of the values are within one Standard Deviation (SD) (half on the left of the mean and the other half on the right of the mean)

4 - About **95%** of the values are within two SD

5 - About **99%** of the values are within three SD

6 - The concept and properties of Normal Distribution is often used for finding probabilities of future events and finding the level of confidence in a prediction

Three very important points

When (based on available data), estimating probability, making predictions or improving the quality of machine learning applications, the following three point are very important:

- **Quality of data**: No matter how good is the application, the poor quality data will produce unreliable results (GIGO)
- **Size of data**: Other things being the same, increasing the observation (data size) improves the quality of output
- **Dispersion**: Higher is the dispersion of data or the value of standard deviation, lower will be the trust we can put in the findings