# Artificial Intelligence Algorithms and Mathematics

CSCN 8000

# Probability

- Probability
- Conditional Probability
- Bayes Theorem
- Naïve Bayes Model
- Spam Filter using Naïve’s Bayes

# What is Probability?

- Probability is how likely an event (E) is to occur.

$$P(E) = \frac{Number\ of\ needed\ outcomes\ of\ event\ E}{Total\ number\ of\ outcomes\ in\ sample\ space\ S}$$

- Example 1: Action: Flip Coin, Sample Space (S) = {Heads, Tails}
  - Sample Space Size =2
  - What is the probability for the event (E) Head to occur?
    - $P(H) = ½$
- Example 2: Action: Flip 2 Coins, Sample Space (S) = {HH,HT,TH,TT}
  - What is the probability for the event (E) = HH to occur?
    - $P(HH) = \frac{1}{4}$
- Example 3: Action: Roll Dice, Sample Space (S) = {1,2,3,4,5,6}
  - What is the probability for getting 1?
    - $P(1) = \frac{1}{6}$
  - What is the probability for getting an even number?
    - $P(Even) = \frac{3}{6} = \frac{1}{2}$

# Complement of a Probability

- $P(A^{'}) = 1 - P(A)$

$$P(Tails) = 1 - P(Heads) = 1 - \frac{1}{2} = \frac{1}{2}$$

$$P(Odd\ Number\ in\ Dice) = ?$$

# Joint Events

- Joint Events are ones that could share some outcomes
- Example: Suppose, in a college, students can play different sports. Assume the total number of students is 10. Number of students who play soccer (S) is 3, Number of students who play basketball (B) is 4, student who play both S and B is 2.
- $P(S) = 0.3, P(B) = 0.4$
- $P(Soccer\ \boldsymbol{and}\ Basketball) = P(S \cap B) = ?$
  - $P(S \cap B) = \frac{2}{10}$
- $P(Soccer\ \boldsymbol{or}\ Basketball) = P(S \cup B) = P(S) + P(B) - P(S \cap B)$
  - $P(S \cup B) = P(S) + P(B) - P(S \cap B) = \frac{1}{2}$

# Disjoint Events

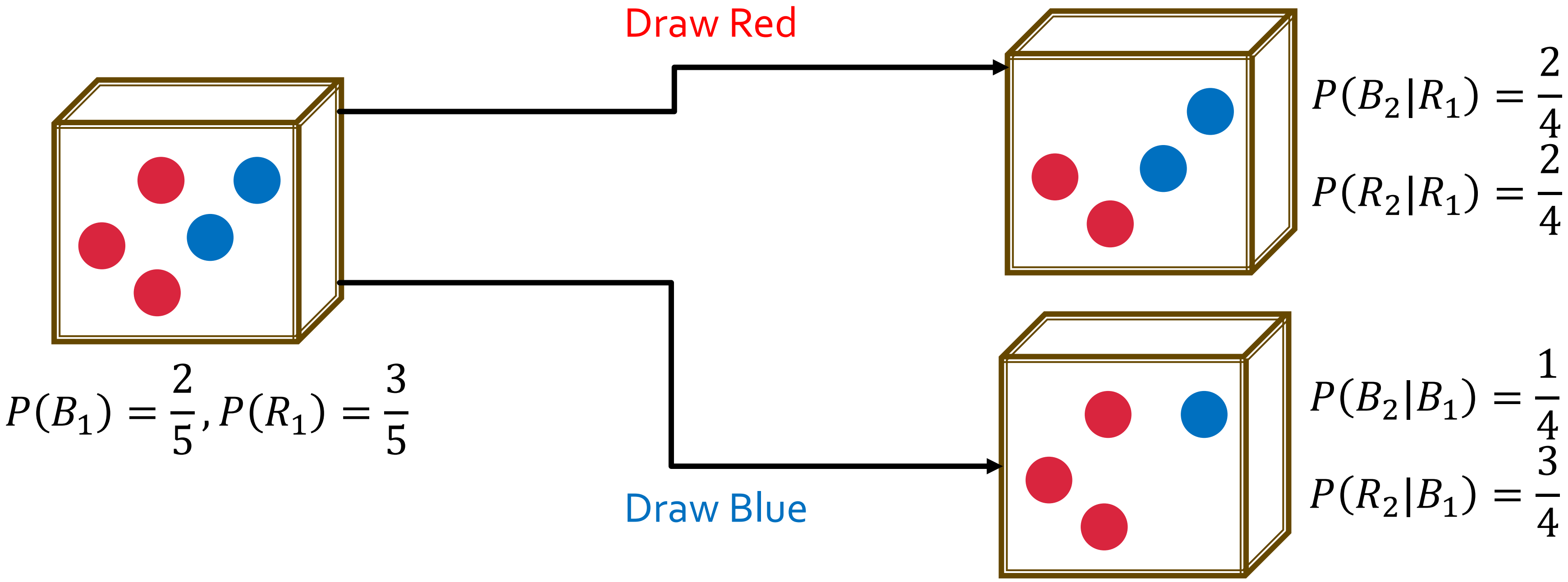
- Disjoint Events are ones that could **not** share any outcome.
- Example: Suppose, in a college, students can play only one sport. Assume the total number of students is 10. Number of students who play soccer (S) is 3, Number of students who play basketball (B) is 4.
- $P(S) = 0.3, P(B) = 0.4$
- $P(Soccer\ \boldsymbol{and}\ Basketball) = P(S \cap B) = ?$
  - $P(S \cap B) = 0$
- $P(Soccer\ \boldsymbol{or}\ Basketball) = P(S \cup B) = ?$
  - $P(S \cup B) = P(S) + P(B) - P(S \cap B) = 0.7$

# Independent Events

- Occurrence of one event does not affect the occurrence of another following event.
- Example: Flipping a coin twice
  - Event A: Getting Heads on First Coin Flip
  - Event B: Getting Heads on Second Coin Flip
  - Are A and B dependent?
  - What is the probability of A **and** B?
    - $P(A \cap B) = P(A) \ * P(B)$

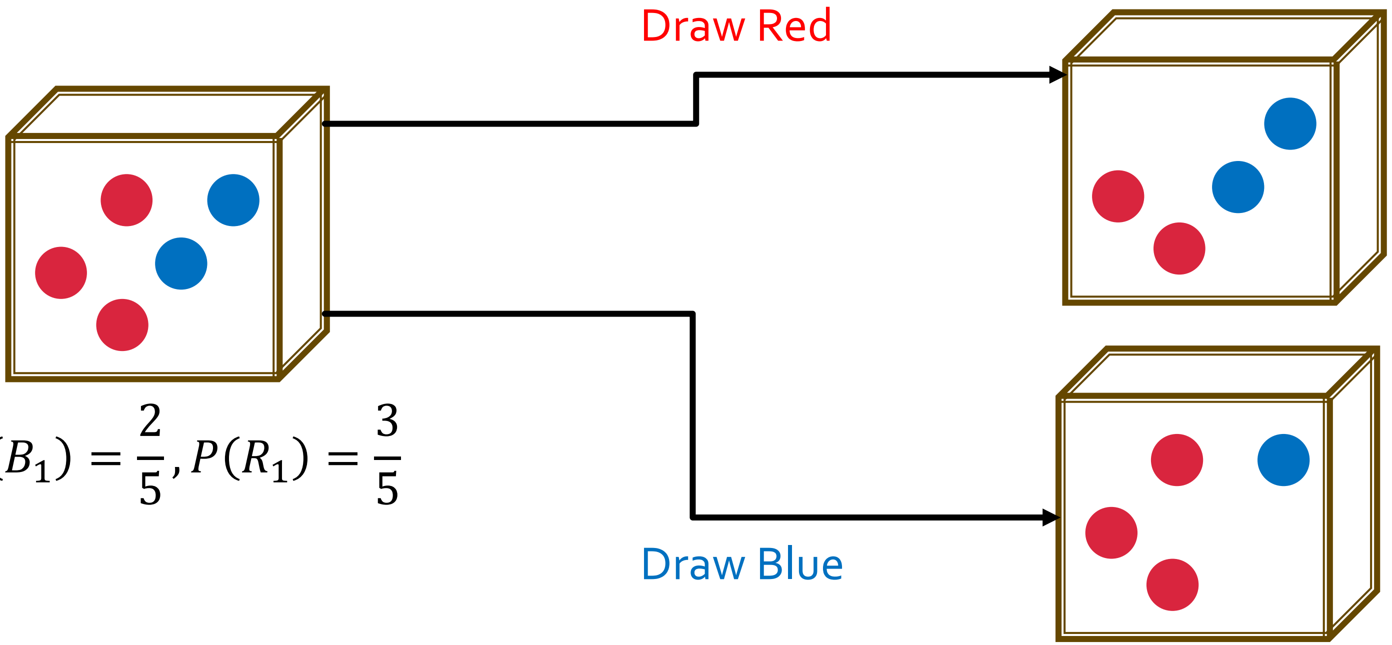
# Conditional Probability

- Probability of event happening given that another event already occurred.
- Example: Suppose we have a box containing 3 red marbles (R) and 2 blue marbles (B). We want to find the probability of drawing a marble on the second draw, given that we drew a marble on the first draw.



# Dependent Events

- Probability of event happening given that another event already occurred.
- Example: Suppose we have a box containing 3 red marbles (R) and 2 blue marbles (B). We want to find the probability of drawing a marble on the second draw, given that we drew a marble on the first draw.




$$P(B_1) = \frac{2}{5}, P(R_1) = \frac{3}{5}$$

$$P(B_2|R_1) = \frac{2}{4}$$

$$\boldsymbol{P(B_2 \cap R_1) = P(R_1) * P(B_2|R_1) = \frac{3}{5} * \frac{2}{4} = \frac{3}{10}}$$

$$P(R_2|B_1) = \frac{3}{4}$$

$$\boldsymbol{P(R_2 \cap B_1) = P(B_1) * P(R_2|B_1) = \frac{2}{5} * \frac{3}{4} = \frac{6}{20}}$$

# Recap of Probability Rules

- Complement : $P(A') = 1 - P(A)$
- Union (OR): $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Intersection (AND):
  - If A and B are independent:
    - $P(A \cap B) = P(A) * P(B)$
    - $P(A \cap B|C) = P(A|C) * P(B|C)$
  - If B depends on A: $P(A \cap B) = P(A) * P(B|A)$
- If A depends on a set of events $\{B_1, B_2, B_3, \dots\}$:
  - $P(A) = P(B_1) * P(A|B_1) + P(B_2) * P(A|B_2) + P(B_3) * P(A|B_3) + \cdots$

# Bayes Theorem

$$\underbrace{P(A|B)}_{\textit{Posterior}} = \frac{\overbrace{P(B|A)}^{\textit{Likelihood}} * \overbrace{P(A)}^{\textit{Prior}}}{P(B)}$$

# Bayes Theorem – Spam Filter

- Suppose we have a dataset for emails being Spam or Not based on the occurrence of the word "Award" in the email.
- Target: What is the probability of Spam given the award feature $P(S|A)$?
- $P(S|A) = \frac{P(A|S) * P(S)}{P(A)}$
- $P(S) = \frac{4}{6}$
- $P(A|S) = \frac{2}{4}$
- $P(A) = P(S) * P(A|S) + P(S') * P(A|S')$

$$= \frac{4}{6} * \frac{2}{4} + \frac{1}{3} * \frac{1}{2} = \frac{1}{2}$$

- $P(S|A) = \frac{\frac{2}{4} * \frac{4}{6}}{\frac{1}{2}} = \frac{2}{3}$

| Spam (S) | AWARD (A) |
|---|---|
| Yes | No |
| Yes | Yes |
| No | No |
| Yes | Yes |
| No | Yes |
| Yes | No |

# Bayes Theorem – Spam Filter

- Suppose we have a dataset for emails being Spam or Not based on the occurrence of the words "Award" and "Gift" in the email.
- Target: What is the probability of Spam given the award and prize features $P(S|A,G)$?
- $P(S|A \cap G) = \frac{P(A \cap G|S) * P(S)}{P(A \cap G)}$
- $P(S|A \cap G) = \frac{[P(A|S) * P(G|S)] * P(S)}{P(A) * P(G)}$ *Assume Independence!*
- $P(S|A \cap G) = \frac{\left[\frac{1}{2} * \frac{1}{4}\right] * \frac{4}{6}}{\frac{1}{2} * \frac{2}{6}} = \frac{1}{2}$

| Spam (S) | AWARD (A) | GIFT (G) |
|---|---|---|
| Yes | No | Yes |
| Yes | Yes | No |
| No | No | No |
| Yes | Yes | No |
| No | Yes | Yes |
| Yes | No | No |

# Naïve Bayes Theorem

- Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of ***conditional independence*** between every pair of features given the value of the class variable.
- General Naiive Bayes Formulation:

$$P(y|x_1, x_2, x_3, \dots) = \frac{P(x_1, x_2, x_3, \dots|y) * P(y)}{P(x_1, x_2, x_3, \dots)} = \frac{[P(x_1|y) * P(x_2|y) * \cdots] * P(y)}{P(x_1) * P(x_2) * \cdots}$$

- What if it's a multi-class classification problem?
  - Compare $P(y_1|x_1, x_2, x_3, \dots), P(y_2|x_1, x_2, x_3, \dots), P(y_3|x_1, x_2, x_3, \dots), \dots$

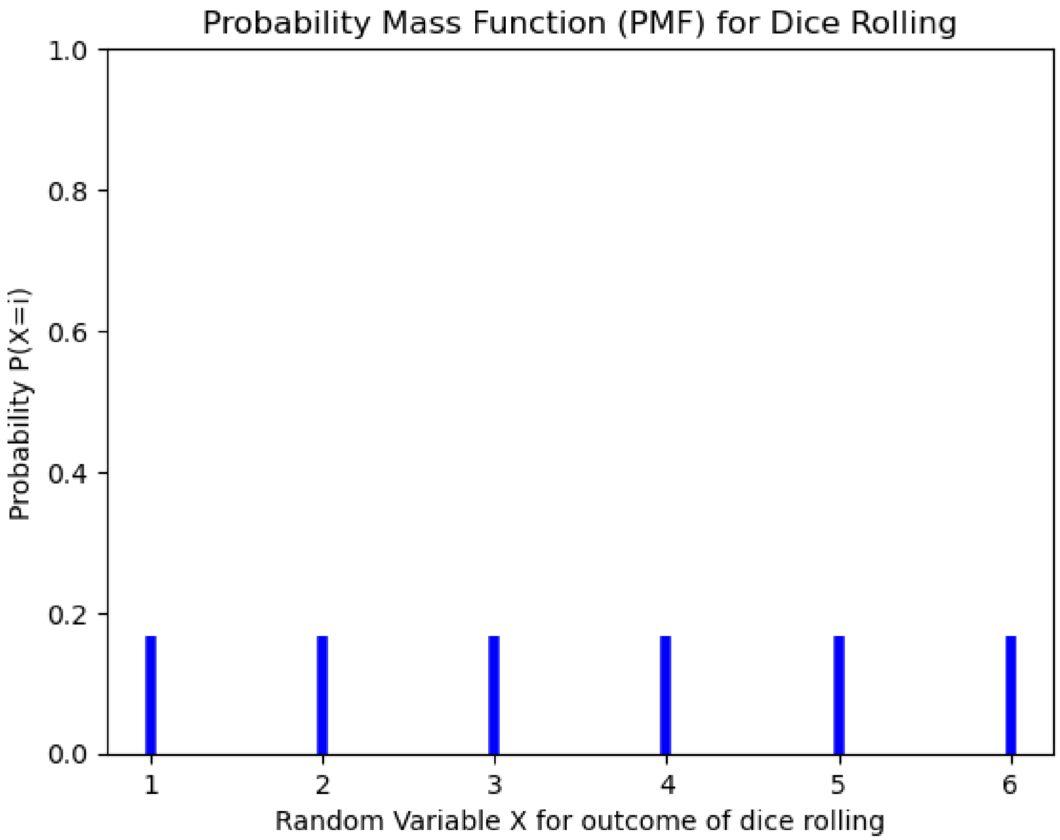
# Random Variables

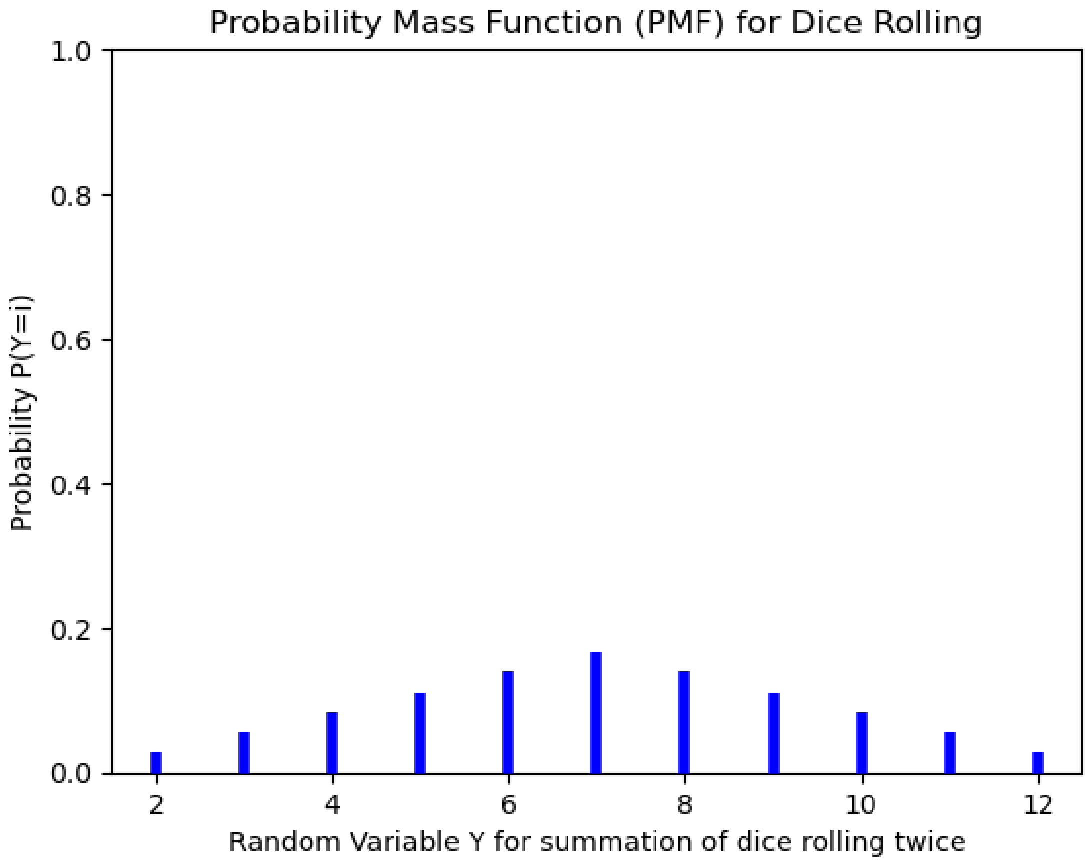
- Random variables are a way to assign numerical values to outcomes of a random process (experiment), making it easier to perform calculations and analyze probabilities.
- Recall the Dice rolling experiment with sample space $\{1,2,3,4,5,6\}$.
  - Case 1: Define a random variable $X$ that represents the value of rolling the dice ***once.***
    - $P(X = 1) = P(X = 2) = P(X = 3) = \ \ldots = \frac{1}{6}$
  - Case 2: Define a random variable $Y$ that represents the summation of dice values after ***two rolls.***
    - $P(Y = 2) = P(X_1 = 1 \ and \ X_2 = 1) = \frac{1}{6} * \frac{1}{6} = \frac{1}{36}$
    - $P(Y = 3) = P(X_1 = 2 \ and \ X_2 = 1) + P(X_1 = 1 \ and \ X_2 = 2) = \frac{1}{18}$
    - $P(Y = 1) = ?$

# Discrete and Continuous Random Variables

- Discrete random variables:
  - Can take on only a countable number of distinct values.
  - Example: random variable $X$ representing number of heads after flipping coin $three$ times.
- Continuous random variables:
  - Can take on any value within a given range (infinite values).
  - Example: random variable $X$ representing the height of a randomly selected person.

# Plotting Discrete Random Variables



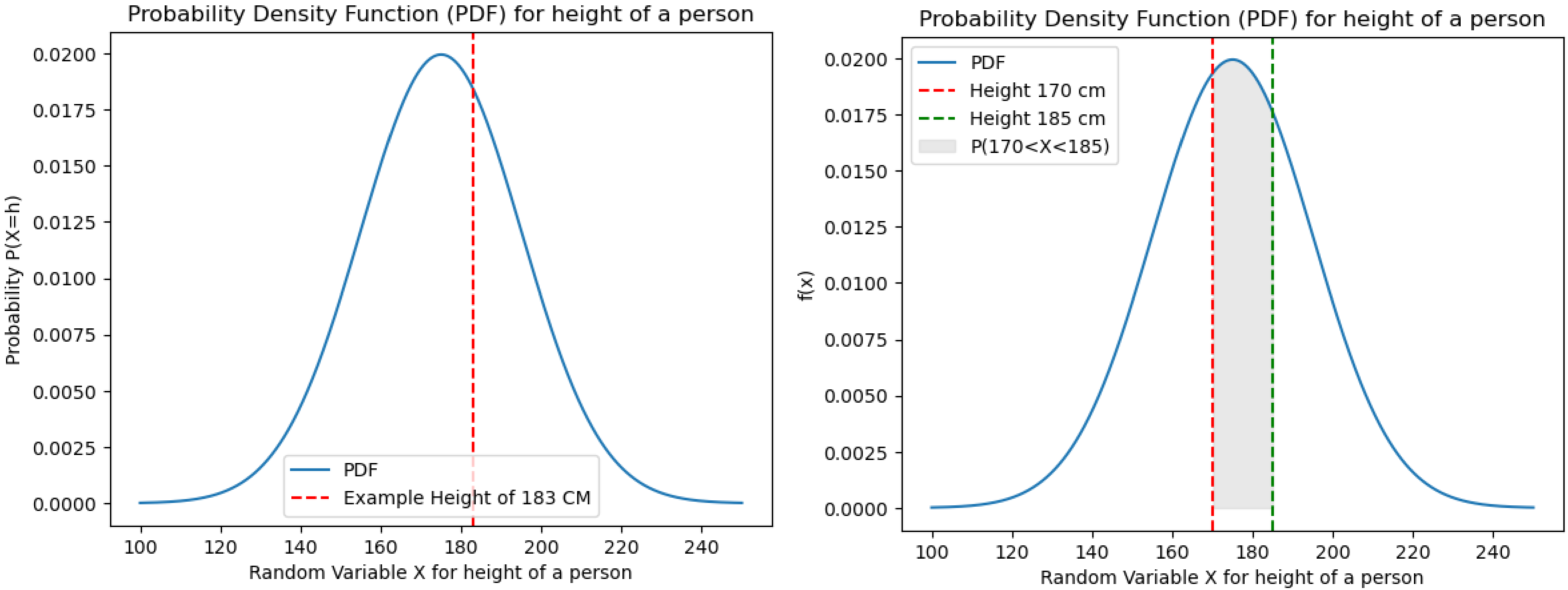

***Probability Mass Function***

# Probability Mass Function

- PMF gives you the probabilities for the discrete random variables, also called discrete density function.
- Ex: Dice Possible Outcomes { 1,2,3,4,5,6}->Sample Space
- For each of them, look at the probability they happen. This forms a *probability distribution*
- All discrete random variables can be modeled by their probability mass function, also called PMF.
- Probability Mass Function :

$$p_X(x) \geq 0$$
$$\sum_x p_X(x) = 1$$

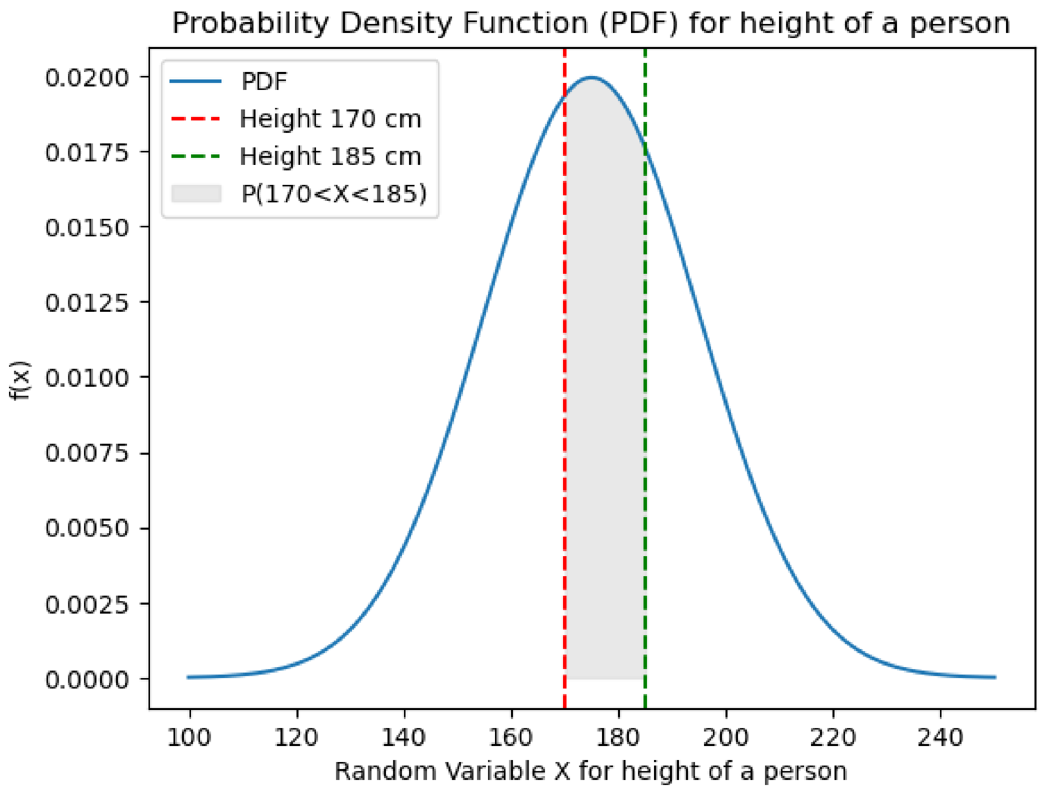
# Plotting Continuous Random Variables




***Probability Density Function***

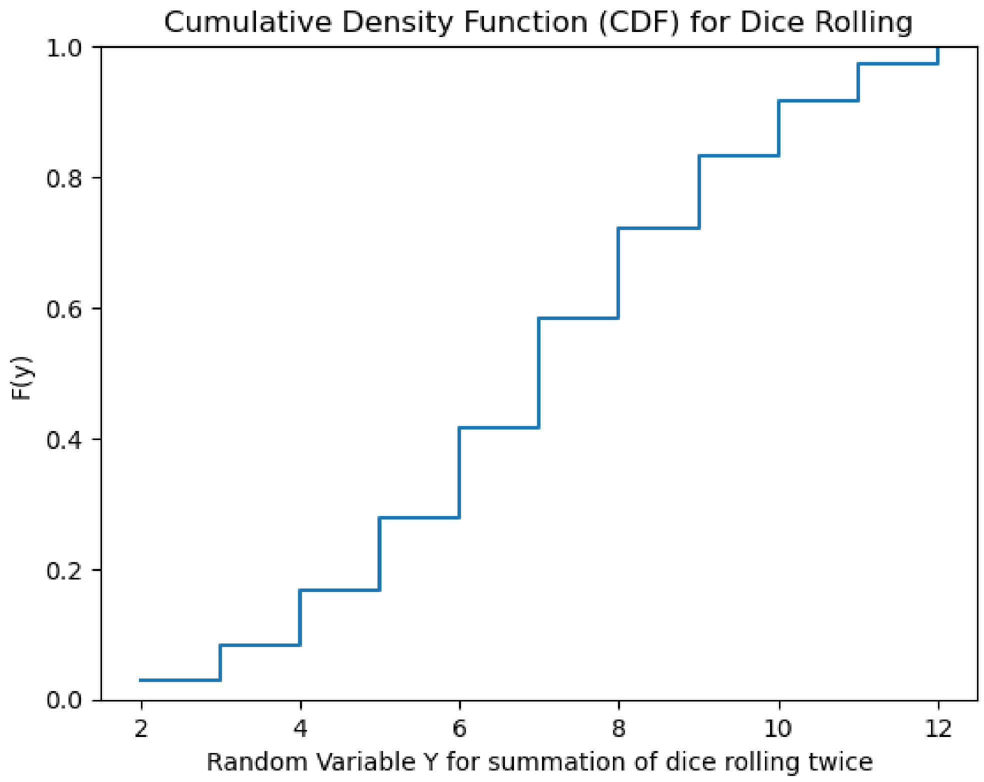
# Probability Density Function

- For a continuous random variable $X$, the Probability Density Function, denoted as $f(x)$, represents the relative likelihood of the variable taking on a particular value $x$.
- Needs to satisfy two properties:
  - Non-Negativity:
    - $f(x) \geq 0, for\ all\ x\ in\ range\ of\ X$
  - Area Under Curve (AUC):
    - $P(a < X < b) = \int_a^b f(x)\, dx$
    - $P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)\, dx = 1$

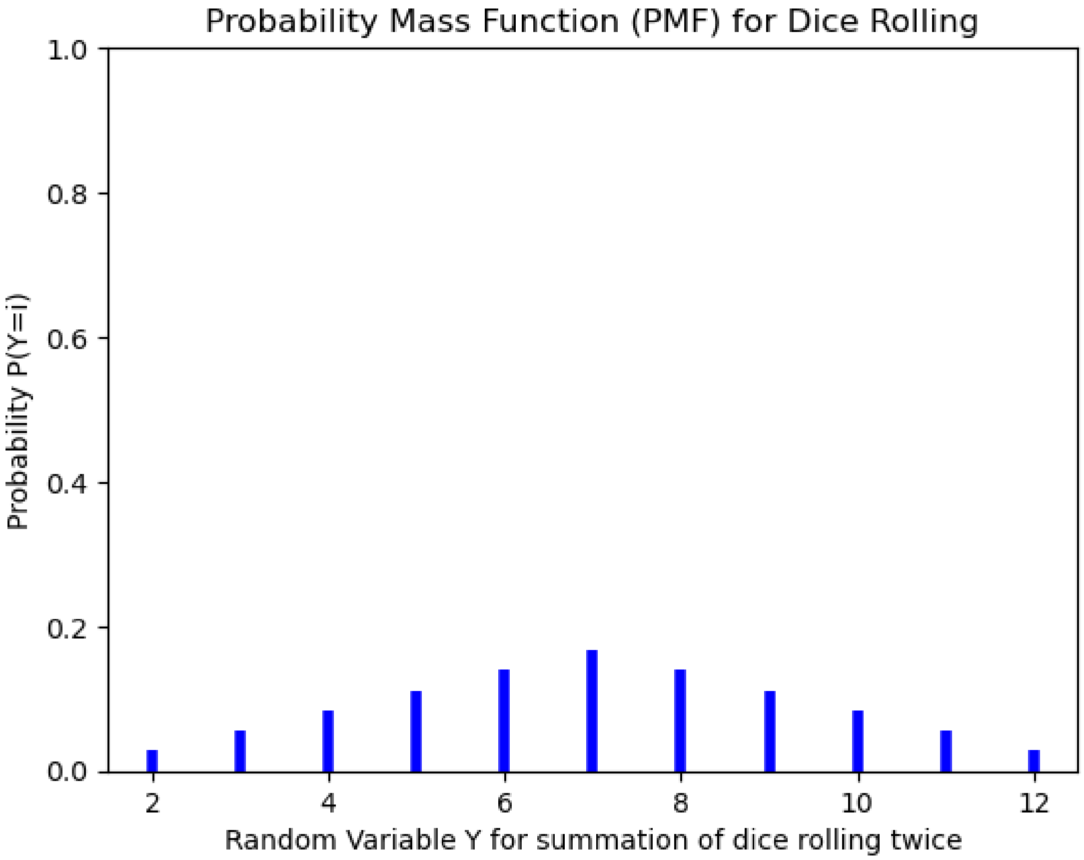


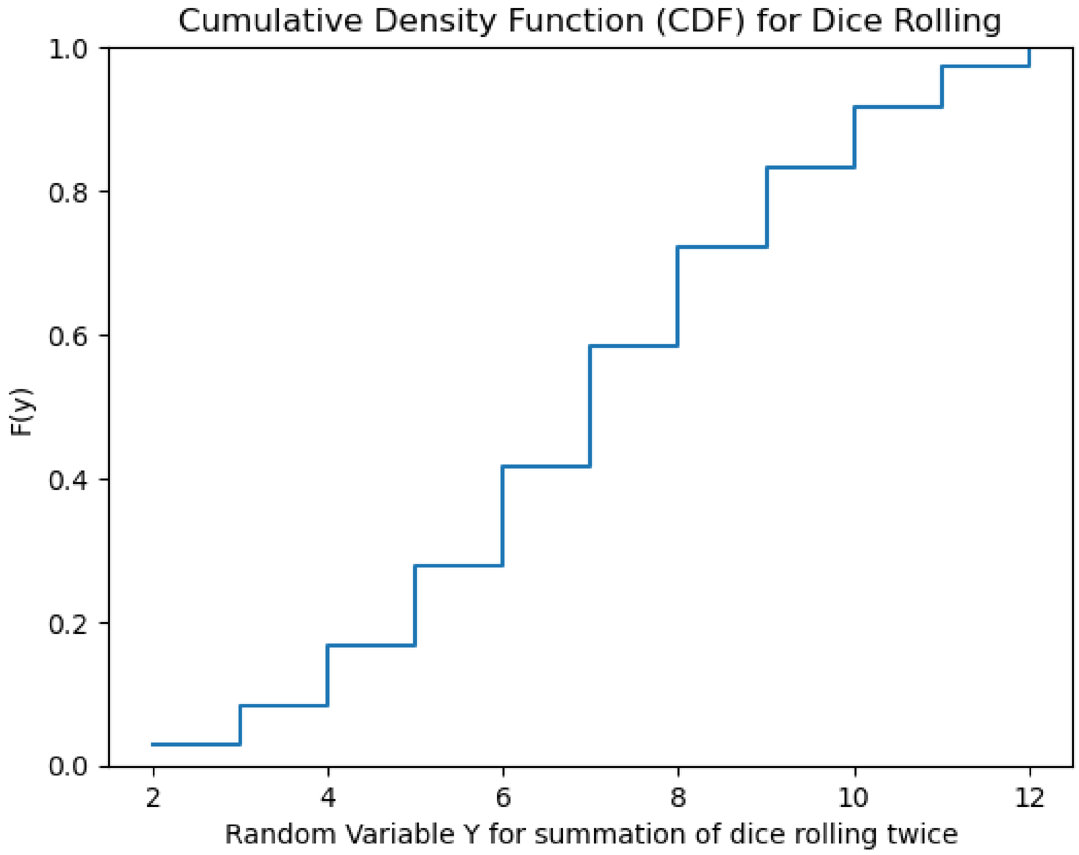
# Cumulative Distribution Function (CDF)

- CDF provides information about the probability that a random variable takes on a value less than or equal to a specified point.
- For Discrete Random Variables:
  - $F(x) = P(X \leq x) = \sum_{x_i \leq x} P(X = x_i)$
- For Continuous Random Variables:
  - $F(X) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt$

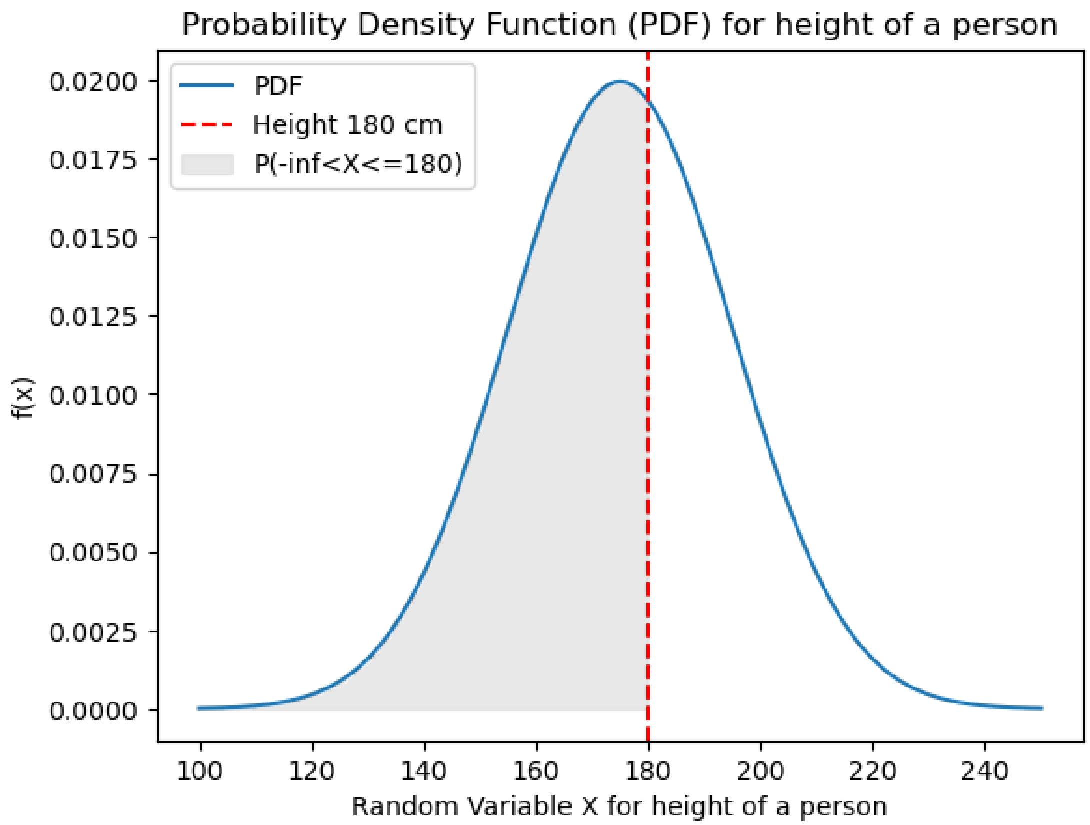


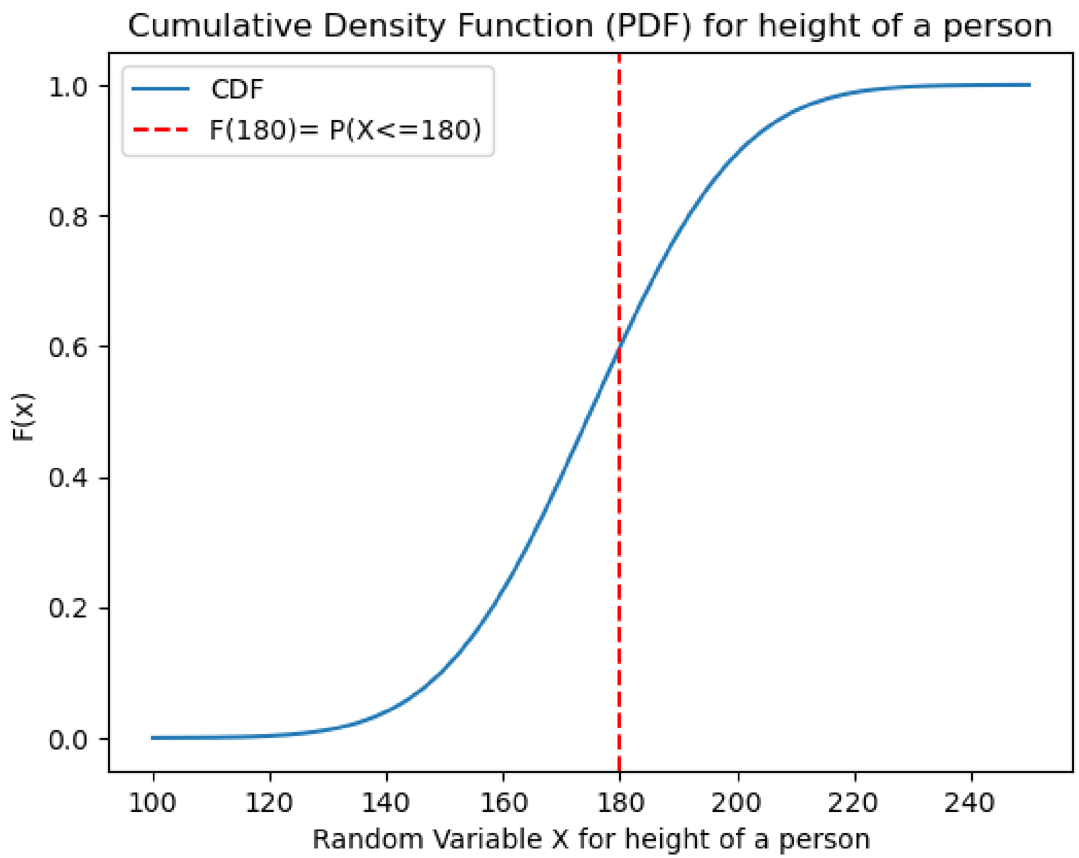
# Cumulative Distribution Function (CDF)



# Cumulative Distribution Function (CDF)



# Recall: Naïve Bayes Theorem

- General Naiive Bayes Formulation:

$$P(y|x_1, x_2, x_3, \ldots) = \frac{P(x_1, x_2, x_3, \ldots|y) * P(y)}{P(x_1, x_2, x_3, \ldots)} = \frac{[P(x_1|y) * P(x_2|y) * \cdots] * P(y)}{P(x_1) * P(x_2) * \cdots}$$

- How did we calculate $P(x_1|y)$ in the Spam Filter?

# Bernoulli Naiive Bayes

- Suited for binary or binarized data where each feature represents the presence or absence of a certain characteristic.
- Primarily used for binary data (Bernoulli Distribution)
- $P(x_1|y)$ is estimated by counting occurrences of True/False values in the feature.

| Spam (S) | AWARD (A) | GIFT (G) |
|---|---|---|
| Yes | No | Yes |
| Yes | Yes | No |
| No | No | No |
| Yes | Yes | No |
| No | Yes | Yes |
| Yes | No | No |

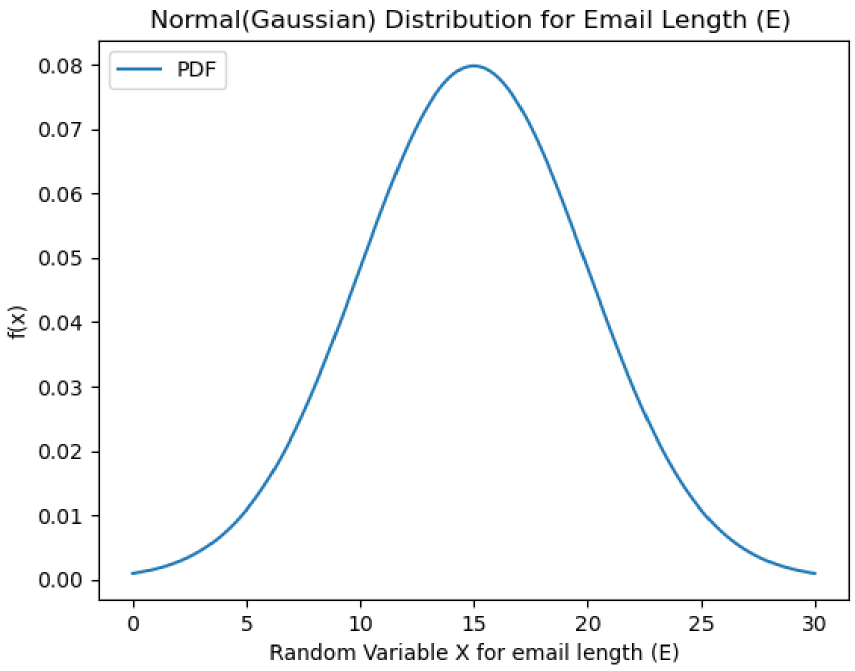
# MultiNomial Naiive Bayes

- It estimates the probability of observing counts among a fixed, discrete set of categories.
- Primarily used for discrete data with multinomial distribution.
- $P(x_1|y)$ is estimated by counting occurrences of each value in the given feature .

| Spam (S) | IMPORTANT WORD (G) |
|---|---|
| Yes | AWARD |
| Yes | GIFT |
| No | GIFT |
| Yes | AWARD |
| No | GIFT |
| Yes | AWARD |

# Gaussian Naive Bayes

- Assumes that the probability of features given the class follows a Gaussian (normal) distribution.
- Assumes that features are continuous and can take any real value.
- $P(x_1|y) = f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(x-\mu)^2}{2\sigma^2})$




| Spam (S) | Email Length (E) |
| --- | --- |
| Yes | 10.5 |
| Yes | 12.4 |
| No | 13.5 |
| Yes | 17.8 |
| No | 20.5 |
| Yes | 3.56 |

# References


- https://learning.oreilly.com/library/view/hands-on-machine-learning/9781098125967/ch01.html#what_is_machine_learning
- https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes

# Thank you!

- Any questions?

**Thank You**

**Youssef Abdelkareem**

**yabdelkareem@conestogac.on.ca**

** Credits for Dr. Anasuya Bhima for assisting in the slides structure*