Dr. Denton Bobeldyk

CIS 365 Artificial Intelligence

Uncertainty in AI

Week in Review

Blackboard Check-in

Uncertainty in AI

- Methods to cope with uncertainty:
  - Fuzzy Logic
  - Probability Theory

Uncertainty in AI - Fuzzy Sets

- Crisp Sets:
  - Standing = {x | x is a student standing in your class}
  - Sitting = {y | y is a student sitting in your class}
- The intersection of the two above sets is an empty set. A student belongs to one or the other.

$$Standing \cap Sitting = \varnothing$$

Uncertainty in AI - Fuzzy Sets

Notation Note:
x such that x is a student standing in your class
used for set notation

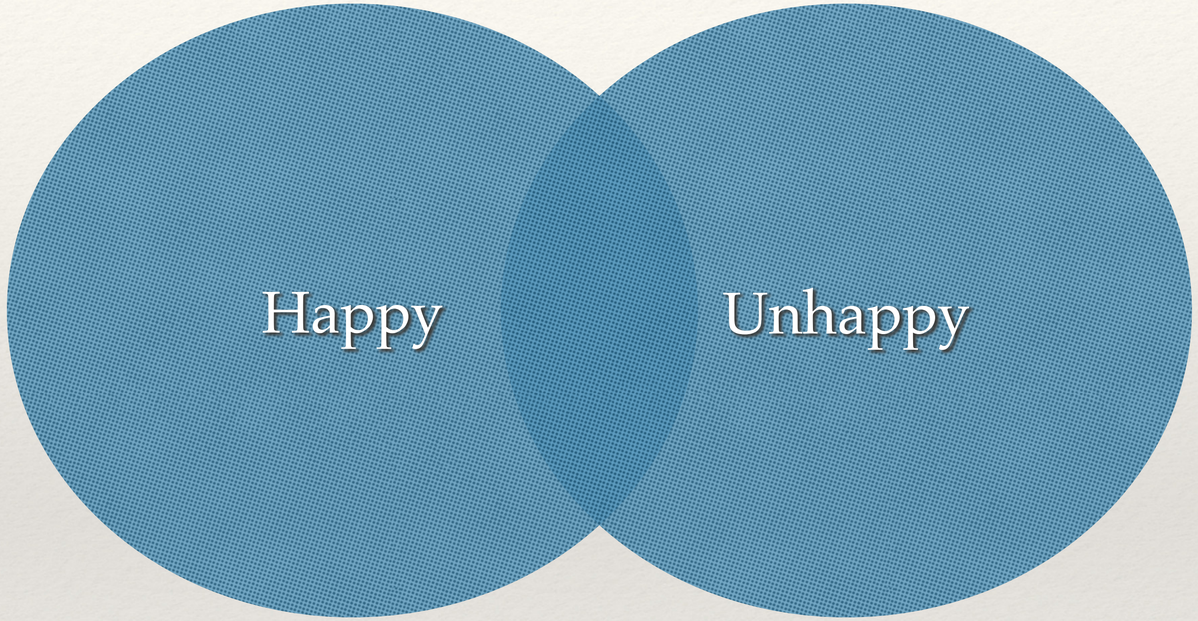
- Crisp Sets:
  - Standing = {x | x is a student standing in your class}
  - Sitting = {y | y is a student sitting in your class}
- The intersection of the two above sets is an empty set. A student belongs to one or the other.

$$Standing \cap Sitting = \varnothing$$

Fuzzy Sets

- Are you happy with your job?
- Are you unhappy with your job?

Fuzzy Sets



It is possible to be Happy and Unhappy about your job at the same time

Uncertainty in AI

- What about concepts like ‘tallness’? Is someone that is 5 foot, someone that would be considered tall?

Uncertainty in AI

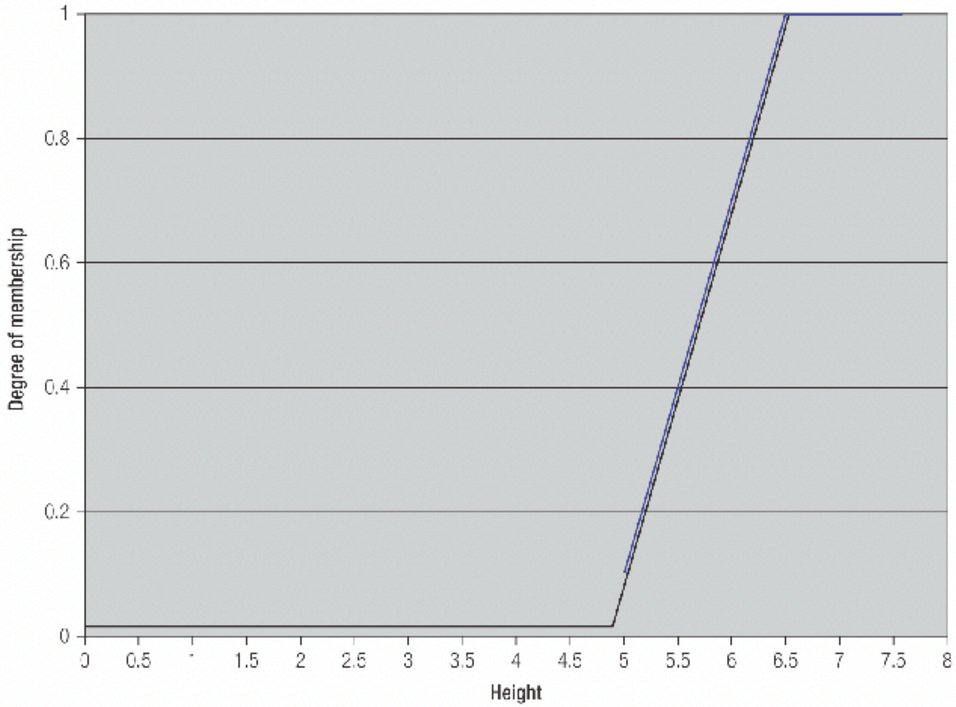


Figure 8.2

Membership function for the set of tall people.

Probability Theory

Probability Theory

- Sample Space for an experiment S is the set of all possible outcomes

Probability Theory

- Sample space of flipping one coin = {H, T}
- Sample space of flipping two coins = {(H,H), (H,T), (T, H), (T,T)}

Probability Theory

- Sample space of flipping one coin = {H, T}
- Probability of Heads = 0.5
  - P(H) = {H, T} = 0.5
- Probability of Tails = 0.5
  - P(T) = {H, T} = 0.5

Probability Theory

- Sample space of flipping two coins = {(H,H), (H,T), (T, H), (T,T)}
- Probability of outcomes:
  - P(H,H) = {(H,H), (H,T), (T, H), (T,T)} = 1/4 = 0.25
  - P(H, T) = {(H,H), (H,T), (T, H), (T,T)} = 1/4 = 0.25
  - P(T, H) = {(H,H), (H,T), (T, H), (T,T)} = 1/4 = 0.25
  - P(T, T) = {(H,H), (H,T), (T, H), (T,T)} = 1/4 = 0.25

Probability Theory

- Rolling a single die
  - What's the sample set?
  - What's the probability that each event will occur?

Probability Theory

- Rolling a single die
  - What's the sample set? {1, 2, 3, 4, 5, 6}
  - What's the probability that each event will occur?
    - P(1) = {1, 2, 3, 4, 5, 6} = 1/6 = 16.67%
    - P(2) = {1, 2, 3, 4, 5, 6} = 1/6 = 16.67%
    - …

Probability Theory

- Rolling 2 dice
  - What's the sample set?
  - What's the probability that each event will occur?

Probability Theory

- Rolling 2 dice
  - What's the sample set?
    - {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
    - (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
    - (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}
  - What's the probability that each event will occur?

Probability Theory

- Rolling 2 dice:
  - P(1,1) =
    - {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
    - (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
    - (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}
  - P(1,1) = P(Event/Number of items in sample space) = 1/36

Probability Theory

- Rolling a 2 dice:
  - P(Sum of two dice = 4)
    - {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
    - (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
    - (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}
  - P(Sum of two dice = 4) = ??

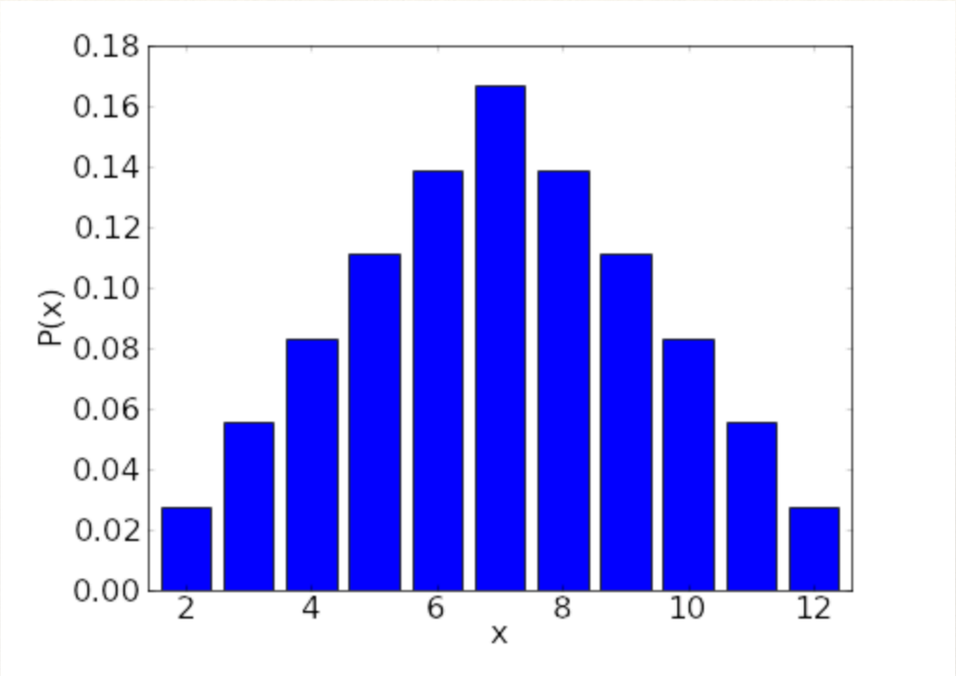
Probability Theory

- Rolling a 2 dice:
  - P(Sum of two dice = 4)
    - {(1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
    - (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
    - (5,1), (5,2), (5,3), (5,4), (5,5), (5,6), (6,1), (6,2), (6,3), (6,4), (6,5), (6,6)}
  - P(Sum of two dice = 4) = 3/36 = 8.3%

Probability Theory

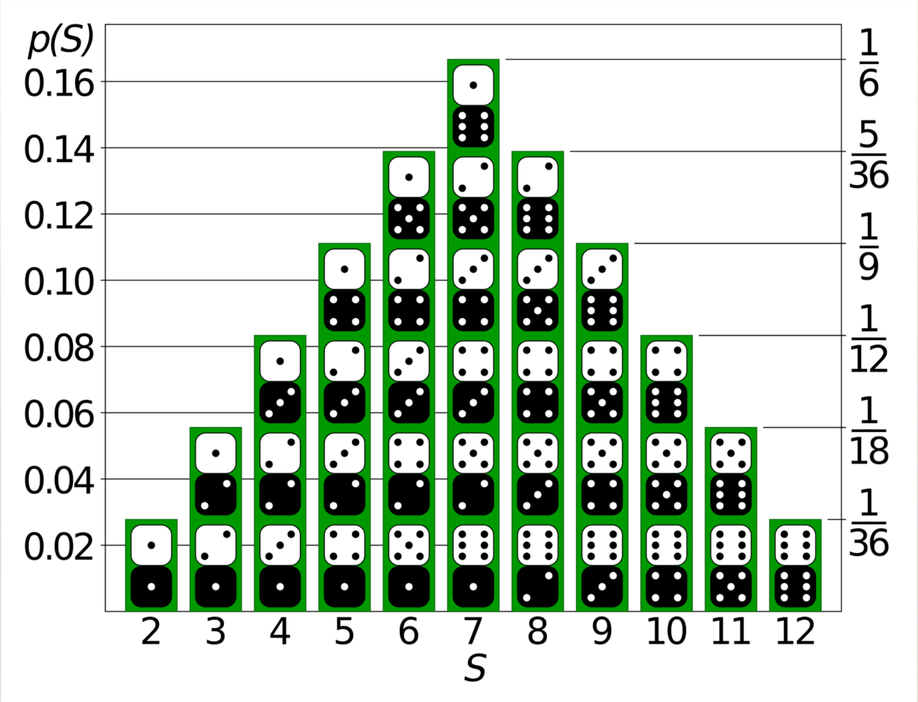
- Probability Distribution for the sum of two dice is?
- If we know the probability distribution, we can determine:
  - What sum has the highest probability?
  - What sum has the lowest probability?
  - Cumulative distribution (i.e., what's the probability of rolling less than 8)

Probability Distribution



https://mathinsight.org/image/two_dice_distribution

Probability Distribution



https://math.stackexchange.com/questions/1204396/why-is-the-sum-of-the-rolls-of-two-dices-a-binomial-distribution-what-is-define

Probability Distribution

Three basic axioms of probability:

1. For any event E: P(E) >= 0

2. P(S) = 1 where S is the sample space of an experiment

3. If the events E_1 and E_2 are mutually exclusive then $P(E_1 \cap E_2) = P(E_1) + P(E_2)$

Class Exercise

- Probability Distribution for 1d6 and 2d6 have been covered
- What is the sample space for:
  - Rolling 1d4
  - Rolling 2d4
- What is the probability distribution:
  - Rolling 1d4
  - Rolling 2d4
  - Rolling 3d4

Bayesian Classifier

- Probability distribution
  - Rolling 1d4
  - Rolling 2d4
  - Rolling 3d4
- Given the following numbers, what is the probability that it came from each of the above distributions?
  - 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12

Bayesian Classifier

- Distributions: 1d4, 2d4, 3d4
- Given the following numbers, what is the probability that it came from each of the above distributions?
  - 1:
    - Probability of 1 Given 1d4 = ?
    - Probability of 1 Given 2d4 = ?
    - Probability of 1 Given 3d4 = ?

Bayesian Classifier

- Distributions: 1d4, 2d4, 3d4
- Given the following numbers, what is the probability that it came from each of the above distributions?
  - 2:
    - Probability of 2 Given 1d4 = ?
    - Probability of 2 Given 2d4 = ?
    - Probability of 2 Given 3d4 = ?

Bayesian Classifier

- Distributions: 1d4, 2d4, 3d4
- Given the following numbers, what is the probability that it came from each of the above distributions?
  - Complete for each of the following numbers: 1 - 12

Completed Answer Next Slide

Probability Distribution

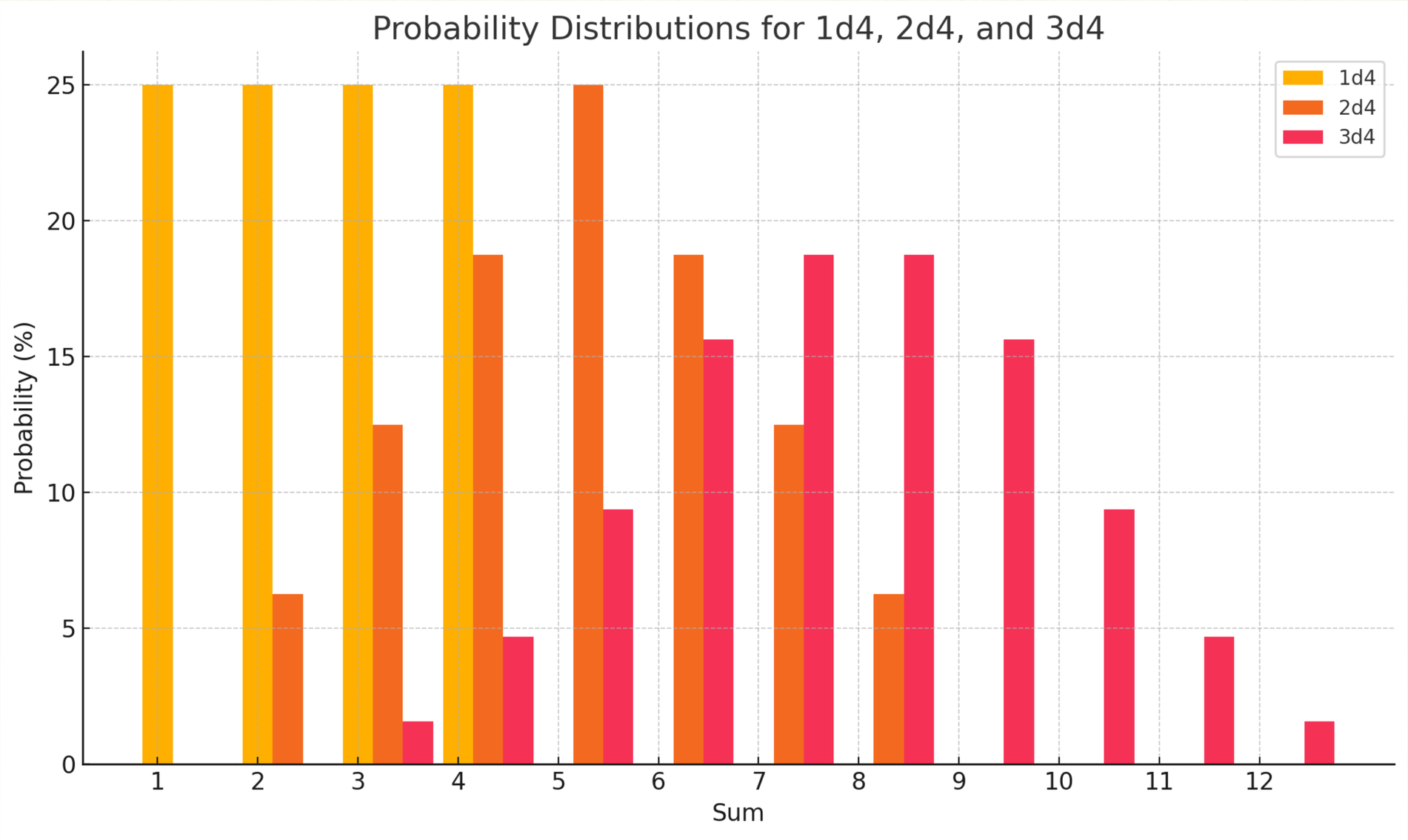
	1d4	2d4	3d4
1	25	0	0
2	25	6.25	0
3	25	12.5	1.6
4	25	18.75	4.7
5	0	25	9.4
6	0	18.75	15.6
7	0	12.5	18.8
8	0	6.25	18.8
9	0	0	15.6
10	0	0	9.4
11	0	0	4.7
12	0	0	1.6

Which distribution did it come from?

- Can we determine if given an outcomes what distribution it came from?

Bayesian Classifier - Which Distribution Generated the outcome?

Outcome	1d4	2d4	3d4
1	25	0	0
2	25	6.25	0
3	25	12.5	1.6
4	25	18.75	4.7
5	0	25	9.4
6	0	18.75	15.6
7	0	12.5	18.8
8	0	6.25	18.8
9	0	0	15.6
10	0	0	9.4
11	0	0	4.7
12	0	0	1.6



Please note the X axis is slightly misaligned

D4 Distributions

- Does a formula exist, if we know the number of d4 we are using, what the probability distribution is?

Please note the X axis is slightly misaligned

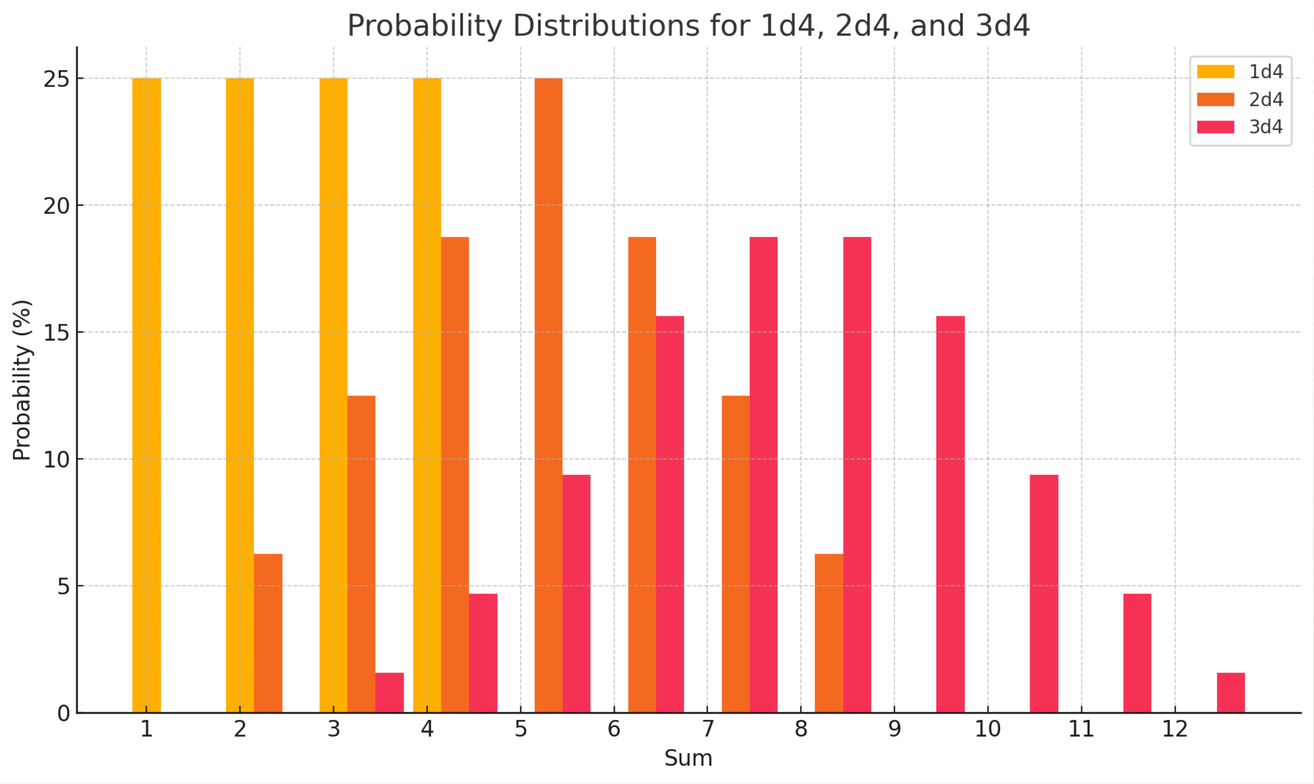
D4 Distributions

- If a formula did exist, we could easily calculate the probability of an outcome given the number of d4's that were used.
- For example, calculateD4Distribution(X, numberOfD4s)
  - calculateD4Distribution(1, 1) = .25
  - calcuateD4Distribution(1,2) = 0
  - calcuateD4Distribution(2,2) = 6.25

Please note the X axis is slightly misaligned

The formula would return the values to the right

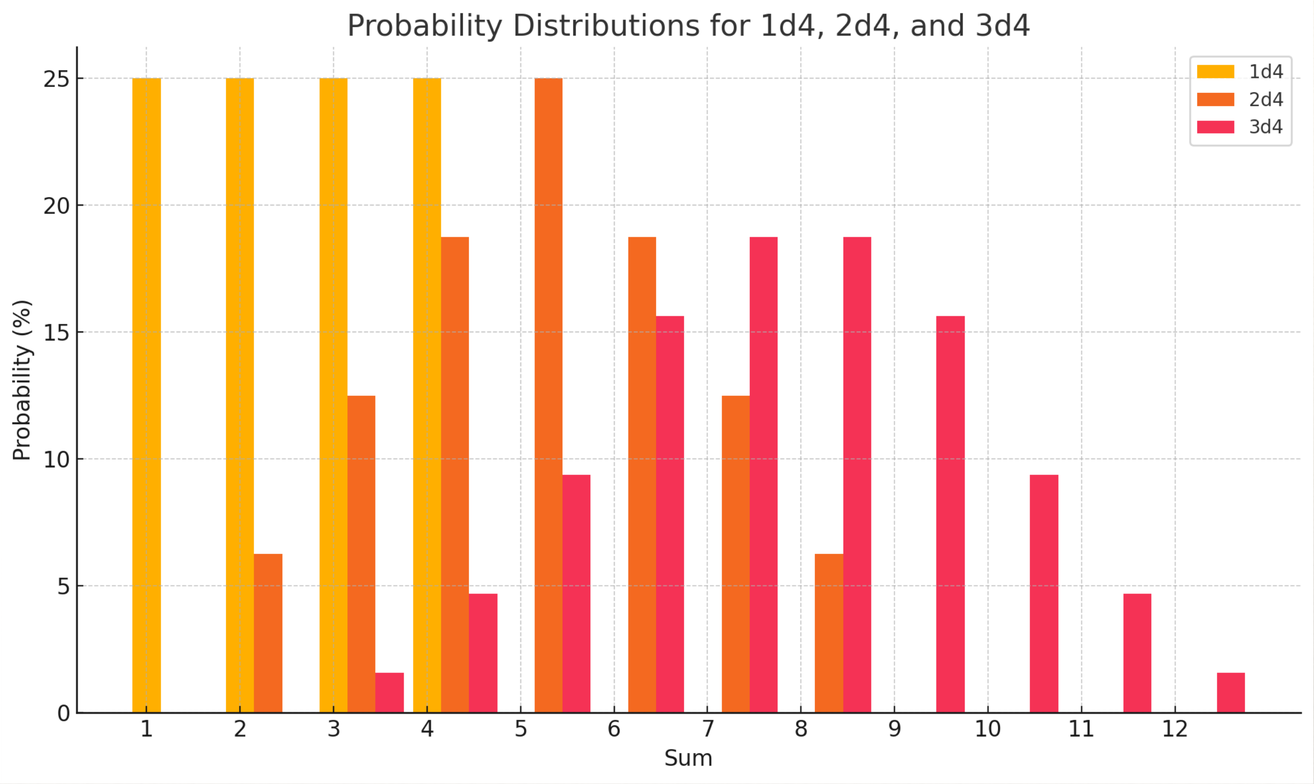
- calculateD4Distribution(1, 1) = .25
- calcuateD4Distribution(1,2) = 0
- calcuateD4Distribution(2,2) = 6.25



Please note the X axis is slightly misaligned

The formula would return the values to the right

- calculateD4Distribution(1, 1) = .25
- calcuateD4Distribution(1,2) = 0
- calcuateD4Distribution(2,2) = 6.25



Please note the X axis is slightly misaligned

No Parametric distribution like this exists!!
We could however write a program that statically returns a value based on a static mapping

Gaussian Distribution

- No formula exists to determine the parametric distribution based on the number of d4's, however parametric distributions do exist, and one of those is the Gaussian distribution.
- Gaussian distribution is determined by the mean (average sample) and standard deviation (how spread out the values are)

Please note the X axis is slightly misaligned

Gaussian Distribution

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

x is the variable

σ is the standard deviation for the distribution

μ is the mean for the distribution

π is a pre-defined constant

e is a pre-defined constant

Please note the X axis is slightly misaligned

Gaussian Distribution

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

Can relatively easily convert the above into a function that programmatically determines the probability of x occurring given a gaussian distribution defined by 2 parameters: μ, σ

Please note the X axis is slightly misaligned

Gaussian Distribution

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

calculateDistribution(x, average, standardDeviation)

Please note the X axis is slightly misaligned

Gaussian Distribution

What if we have more than one feature?

Pack all of the features into a ‘feature vector’

Please note the X axis is slightly misaligned

Gaussian Distribution

What if we have more than one feature?

Pack all of the features into a ‘feature vector’

For example:
Height of an athlete

Just One feature!!

More than one feature would be:
Height, Weight, Shoe size

Three Features!!

Please note the X axis is slightly misaligned

Gaussian Distribution

What if we have more than one feature?

Pack all of the features into a ‘feature vector’

Data Representation

For example:
Height of an athlete

[70]

More than one feature would be:
Height, Weight, Shoe size

[70,200,13]

Please note the X axis is slightly misaligned

Multivariate Gaussian Distribution

$$f(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^k |\mathbf{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu) \right)$$

Please note the X axis is slightly misaligned

```python
import numpy as np

def multivariate_gaussian(x, mu, covarianceMatrix):
    k = len(mu)
    covarianceMatrix_det = np.linalg.det(covarianceMatrix)
    covarianceMatrix_inv = np.linalg.inv(covarianceMatrix)
    norm_factor = 1.0 / np.sqrt((2 * np.pi) ** k * covarianceMatrix_det)
    diff = x - mu
    exponent = -0.5 * np.dot(np.dot(diff.T, covarianceMatrix_inv), diff)
    return norm_factor * np.exp(exponent)
```

Parameters:
x (np.array): The k-dimensional data point (column vector).
mu (np.array): The mean vector (column vector).
covarianceMatrix (np.array): The covariance matrix (k x k matrix).

```python
import numpy as np

def multivariate_gaussian(x, mu, covarianceMatrix):
    k = len(mu)
    covarianceMatrix_det = np.linalg.det(covarianceMatrix)
    covarianceMatrix_inv = np.linalg.inv(covarianceMatrix)
    norm_factor = 1.0 / np.sqrt((2 * np.pi) ** k * covarianceMatrix_det)
    diff = x - mu
    exponent = -0.5 * np.dot(np.dot(diff.T, covarianceMatrix_inv), diff)
    return norm_factor * np.exp(exponent)
```

```python
# Example usage:
x = np.array([1.0, 2.0])
mu = np.array([0.0, 0.0])
covMatrix = np.array([[1.0, 0.5], [0.5, 1.0]])


# Calculate the multivariate Gaussian value
result = multivariate_gaussian(x, mu, covMatrix)
print(result)
```

Data

- Split the data into 60% training, 20% validation and 20% testing
- Training data is used to train the model
- Validation is used to tweak the hyper/parameters of the model
- Testing is used “1” time to determine how well your model generalizes to data it hasn’t seen before

Data

- Determining Mean and Standard Deviation for the Training Data

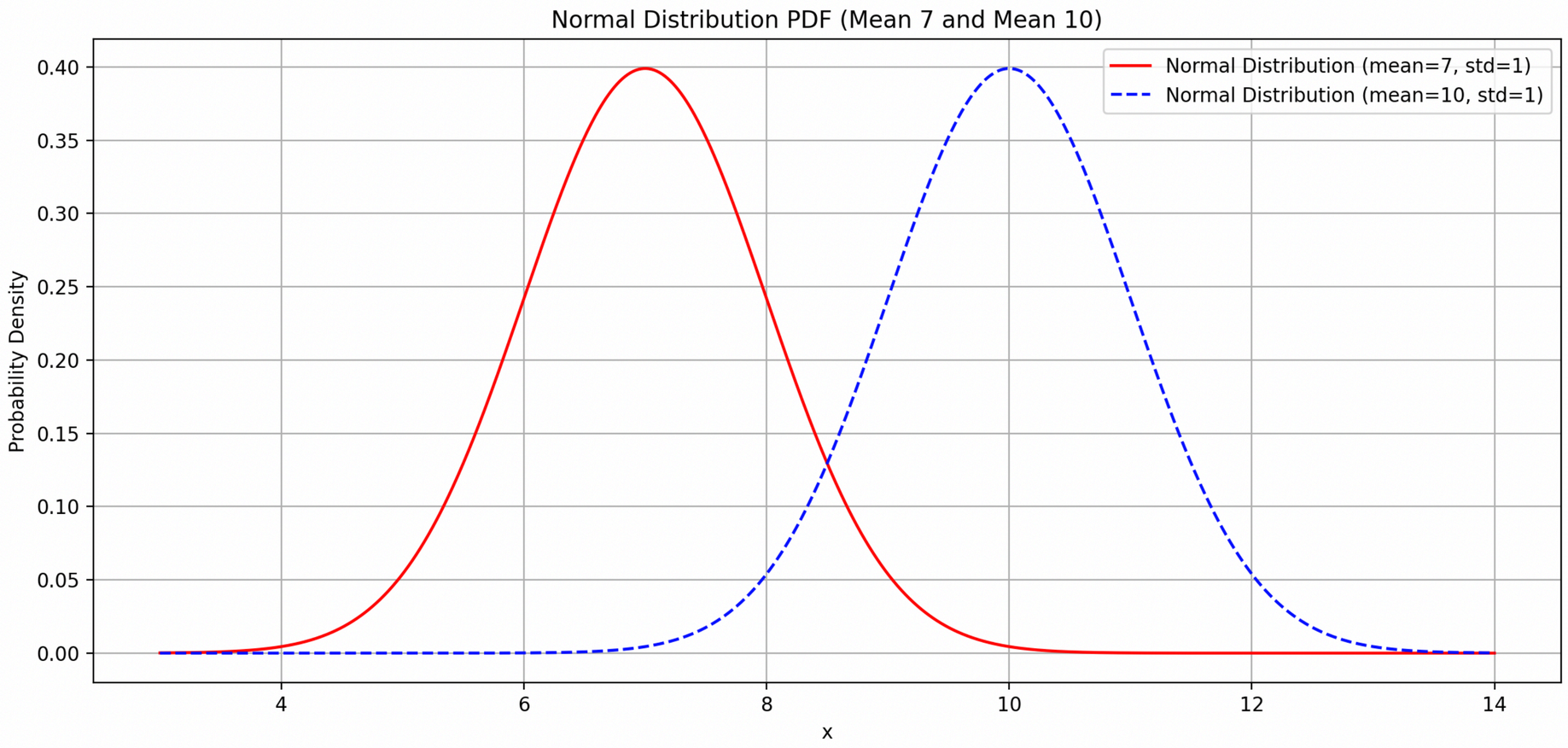
Calculating mean and standard deviation

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

n = number of data points in the sample
x_i is the ith sample

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Bayesian Classifier

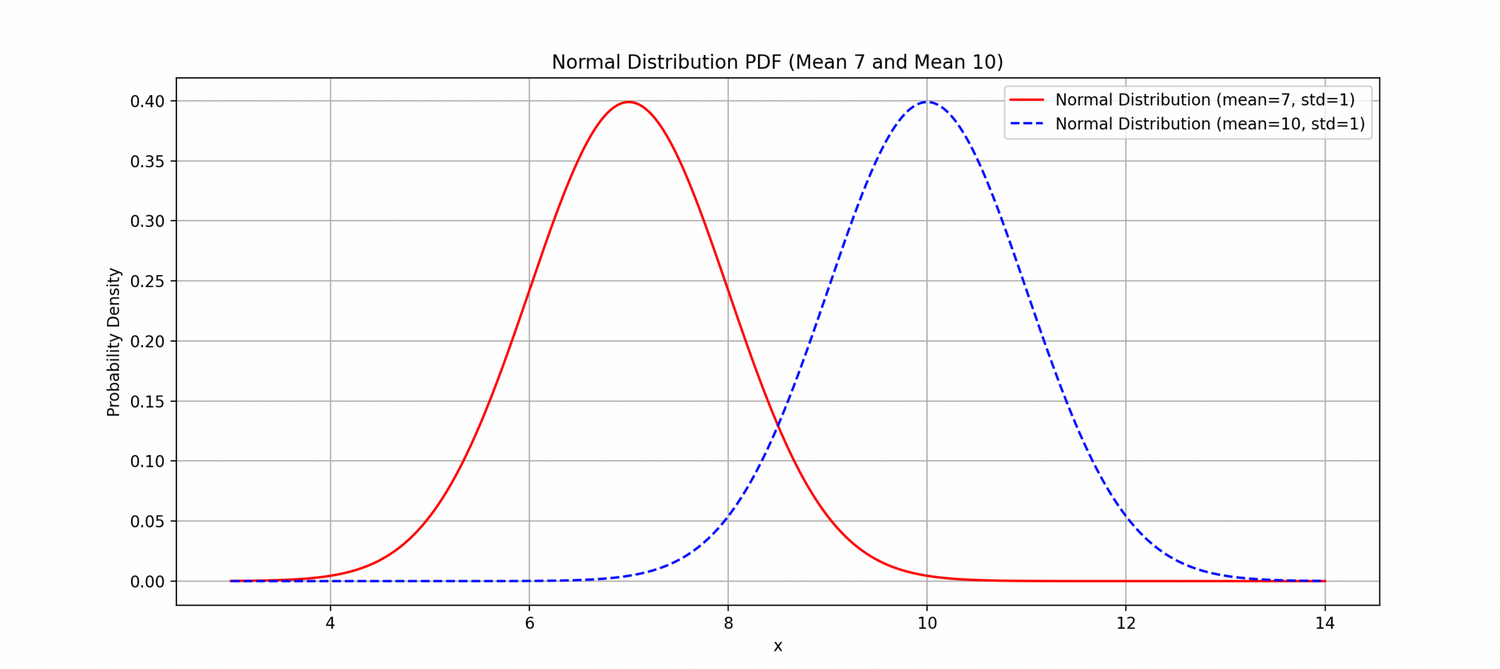


MatPlotLib

https://www.w3schools.com/python/matplotlib_intro.asp

Probability Density Functions (PDF)

The Probability Density Function (PDF) represents the likelihood of different outcomes in a continuous random variable.



Probability Density Functions (PDF)

In python, can calculate the PDF for a normal distribution using the norm.pdf command:

```
myValue = norm.pdf(x, mean, std_dev)
```

```
from scipy.stats import norm
```

Bayesian Classifier

```
probabilityFromA= norm.pdf(x, A_mean, A_std_dev)

probabilityFromB= norm.pdf(x, B_mean, B_std_dev)

If (probabilityFromA >  probabilityFromB):

    Predict class A
else:

   Predict class B
```

Bayesian Classifier Example

```
valueToQuery = 6
probFromA = norm.pdf(valueToQuery, mean1, std_dev)
promFromB = norm.pdf(valueToQuery, mean2, std_dev)


print(f"Probability that {valueToQuery} is from A = [{probFromA:.5f}]")
print(f"Probability that {valueToQuery} is from B = [{promFromB:.5f}]")

if(probFromA > promFromB):
    print("Predicting {valueToQuery} is from Class A")
else:
    print("Predicting {valueToQuery} is from Class B")
```

Bayesian Classifier Example

Probability that 6 is from A = [0.24197]
Probability that 6 is from B = [0.00013]
Predicting 6 is from Class A

Bayesian Classifier Example

```
valueToQuery = 9
probFromA = norm.pdf(valueToQuery, mean1, std_dev)
promFromB = norm.pdf(valueToQuery, mean2, std_dev)
```

```
print(f"Probability that {valueToQuery} is from A = [{probFromA:.5f}]")
print(f"Probability that {valueToQuery} is from B = [{promFromB:.5f}]")
```

```
if(probFromA > promFromB):
    print("Predicting {valueToQuery} is from Class A")
else:
    print("Predicting {valueToQuery} is from Class B")
```

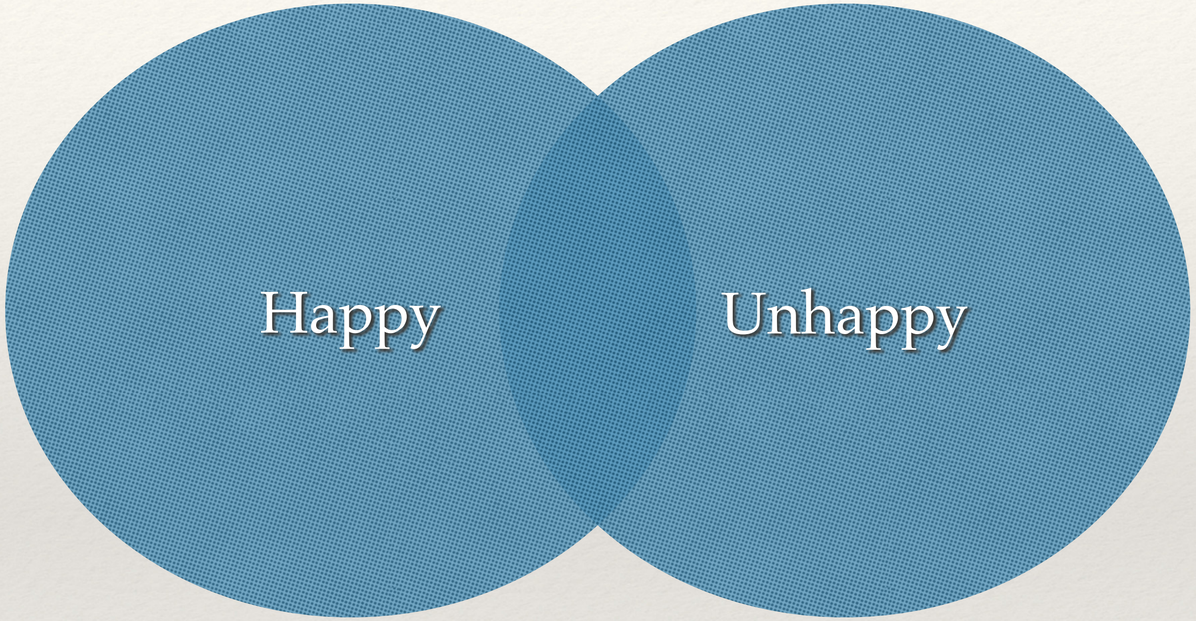
Bayesian Classifier Example

Probability that 9.5 is from A = [0.01753]
Probability that 9.5 is from B = [0.35207]
Predicting 9.5 is from Class B

More Probability

Union

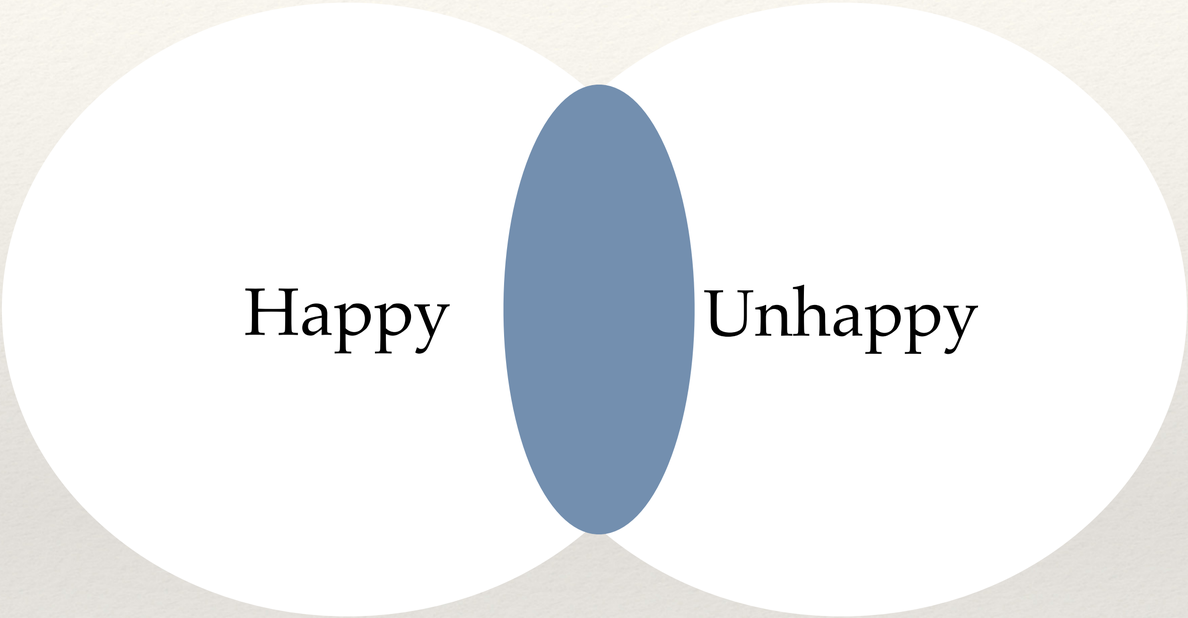
- Union ∪



$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Intersect

- Intersect ∩



Class Exercise

A single card is drawn from a standard 52-card deck. Let D be the event that the card drawn is a diamond, and let F be the event that the card drawn is a face card. Find the indicated probabilities:

1) P(D) =
2) P(F) =
3) P(F') =
4) P(D') =
5) $P(D \cap F)$ =
6) $P(D' \cap F)$ =
7) $P(D \cup F)$ =
8) $P(D' \cup F)$ =
9) $P(D \cap F')$ =
10) $P(D' \cap F')$ =
11) $P(D \cup F)$=
12) $P(D' \cup F')$=

Playing Cards

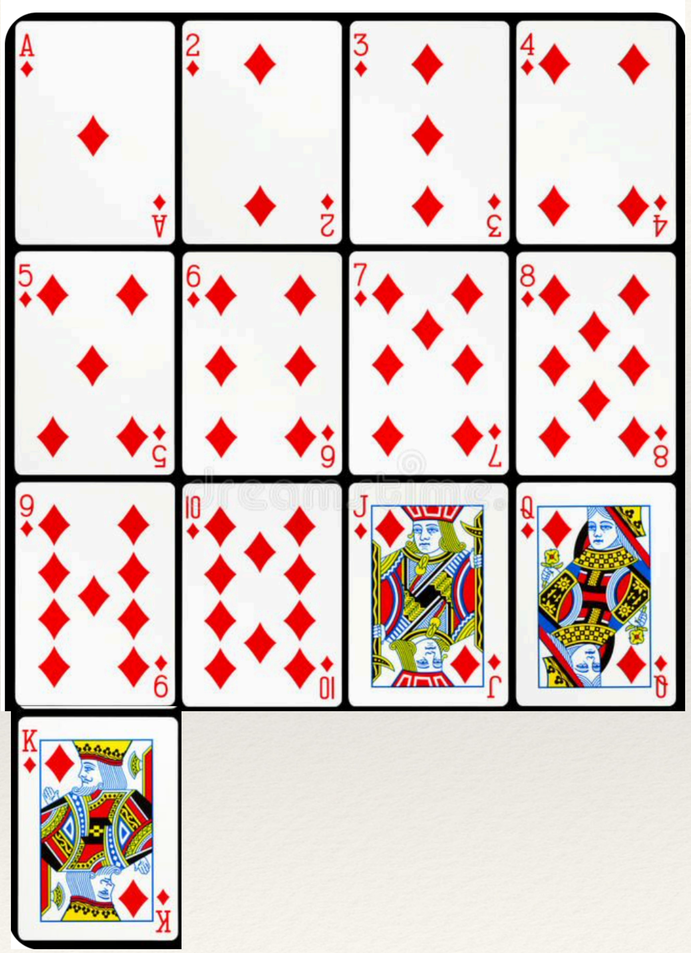


Image retrieved from: https://in.pinterest.com/pin/768778598882188678/

Class Exercise

A single card is drawn from a standard 52-card deck. Let D be the event that the card drawn is a diamond, and let F be the event that the card drawn is a face card. Find the indicated probabilities:

1) P(D) =
2) P(F) =
3) P(F') =
4) P(D') =
5) $P(D \cap F)$ =
6) $P(D' \cap F)$ =
7) $P(D \cup F)$ =
8) $P(D' \cup F)$ =
9) $P(D \cap F')$ =
10) $P(D' \cap F')$ =
11) $P(D \cup F)$=
12) $P(D' \cup F')$=

Class Exercise

A single card is drawn from a standard 52-card deck. Let D be the event that the card drawn is a diamond, and let F be the event that the card drawn is a face card. Find the indicated probabilities:

1) P(D) = .25
2) P(F) = 12/52 = .2308
3) P(F') = 40/52 = .7692
4) P(D') = .75
5) $P(D \cap F)$ = 3/52 = .0577
6) $P(D' \cap F)$ = 9/52 = .1731
7) $P(D \cup F)$ = P(D) + P(F) – $P(D \cap F)$ = 13/52 + 12/52 – 3/52 = 22/52 = .42307
8) $P(D' \cup F)$= P(D') + P(F) – $P(D' \cap F)$ = 39/52 + 12/52 – 9/52 = 42/52 = .80769
9) $P(D \cap F')$= 10/52 = .1923
10) $P(D' \cap F')$= (39-9)/52 = .5769
11) $P(D \cup F)$= (D) + P(F) – $P(D \cap F)$= 13/52 + 12/52 – 3/52 = 22/52 = .42307
12) $P(D' \cup F')$= P(D') + P(F') - $P(D' \cap F')$ = 39/52 + 40/52 – 30/52 = 49/52 = .9423:

Conditional Probability

P(A | B) = Probability event A occurs, given event B occurs

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

$P(A \cap B)$ is the probability of both A and B occurring

$P(B)$ is the probability of event B occurring

Independent Events

The occurrence of one event does not affect the probability of the other event

$P(A \mid B)$ = Probability event A occurs, given event B occurs

If the events A and B are independent of each other, then $P(A \mid B) = P(A)$

Independent Events

The occurrence of one event does not affect the probability of the other event

$P(A \mid B)$ = Probability event A occurs, given event B occurs

If the events A and B are independent of each other, then $P(A \mid B) = P(A)$

What are some examples of independent events?

Independent Events Examples

Rolling 2 dice

Drawing 2 cards (with replacement)

It raining and you rolling a dice?

Conditional Probability

Let's say you would like to know the probability of drawing a face card (A) given that you know the card is a diamond (B)

Event A: drawing a face card

Event B: drawing a diamond

Conditional Probability

This used to be P(F)

Let's say you would like to know the probability of drawing a face card (A) given that you know the card is a diamond (B) This used to be P(D)

Event A: drawing a face card

Event B: drawing a diamond

Conditional Probability

Identify $P(A \cap B)$

A: Drawing a face card

B: Drawing a diamond

So $P(A \cap B) = 3/52$

Conditional Probability

Identify $P(B)$

B: Drawing a diamond

So $P(B) = 13/52$

Conditional Probability

Apply the formula $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)} = \dfrac{3/52}{13/52} = 3/13$

Probability

To transfer into a particular technical department, a company requires an employee to pass a screening test. A maximum of 3 attempts are allowed at 6-month intervals between trials. From past records it is found that 40% pass on the first trial; of those that fail the first trial and take the test a second time, 60% pass; and of those that fail on the second trial and take the test a third time, 20% pass.

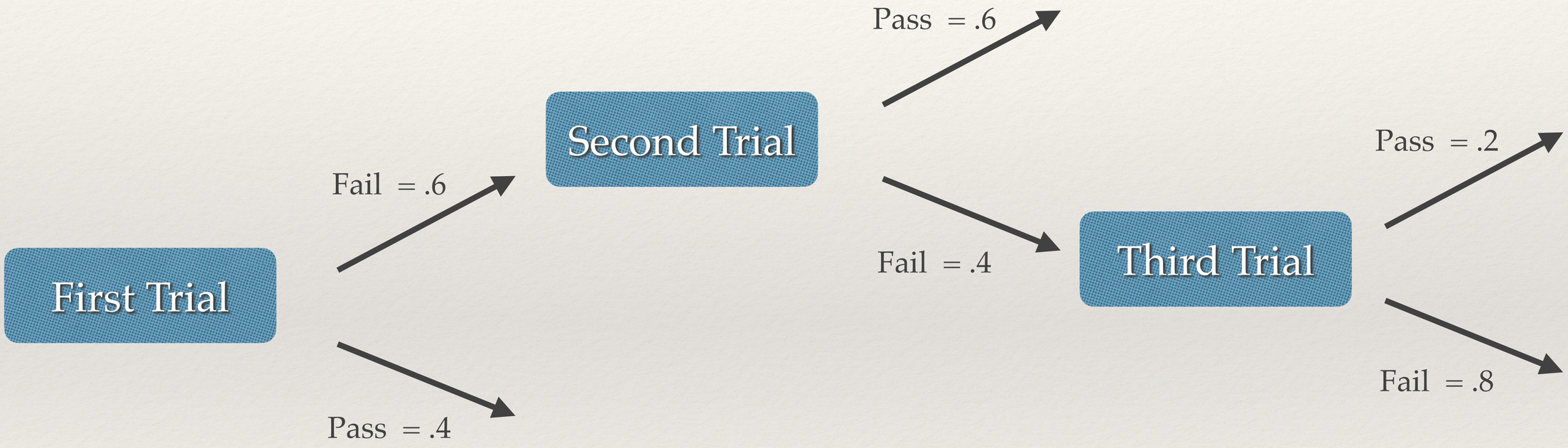
Probability

P(pass on first) = 0.4

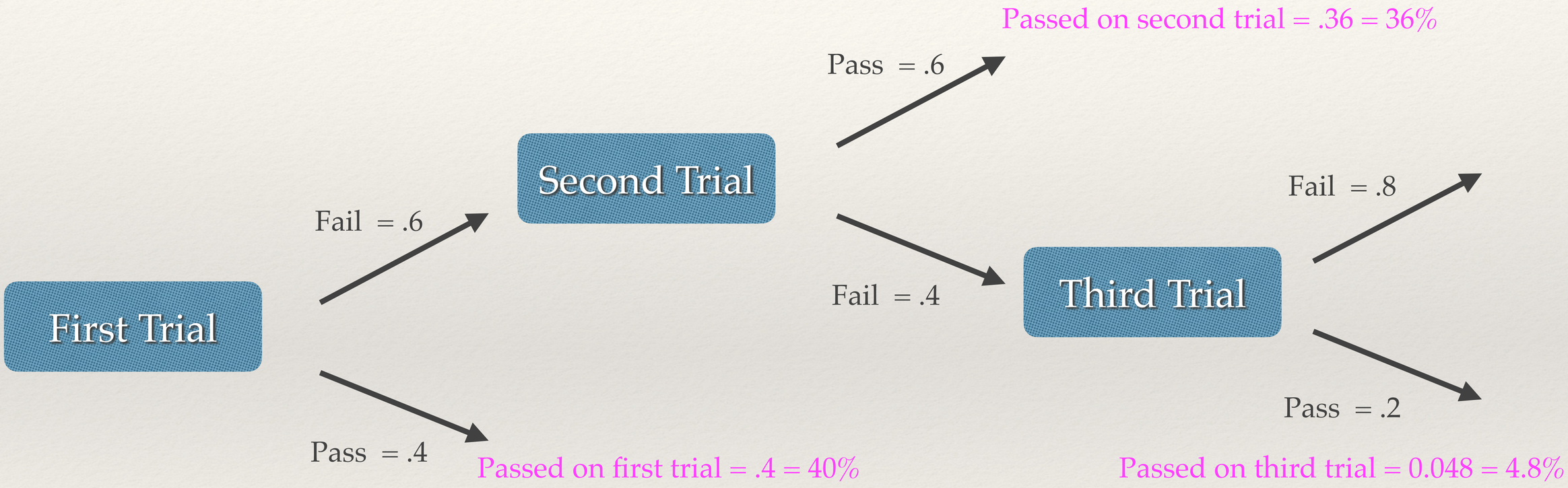
P(pass on second | failed first) = .6

P(pass on third | failed second) = .2

Conditional Probability



Conditional Probability



Bayes Theorem

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

P(A | B) is the posterior probability, the probability of event A given that B has occurred
P(B | A) is the likelihood, the probability of observing B given that A is true
P(A) is the prior probability of A, i.e., the initial belief about A before any evidence
P(B) is the marginal probability or the total probability of B, which accounts for all possibilities of B

Bayes Theorem

A brief physical exam is administered to all new inmates at a prison. Suppose that 80% of all healthy individuals pass this exam, 60% of all individuals with minor ailments pass and 30% of all prisoners with serious ailments also pass. Suppose that 25% of these new prisoners are in good health, 50% have minor ailments, and 25% have major health issues. Given that an inmate passes this physical event, what is the conditional probability that the inmate is in good physical condition?

Bayes Theorem Example

Let's define the events:

- **H**: The inmate is in good health.
- **M**: The inmate has minor ailments.
- **S**: The inmate has serious ailments.
- **P**: The inmate passes the physical exam.

We are interested in finding the conditional probability that an inmate is in good health given that they passed the exam $P(H \mid P)$

Bayes Theorem Example

Prior probabilities:

$P(H) = 0.25$ 25% of the prisoners are in good health

$P(M) = 0.50$ 50% of the prisoners have minor ailments

$P(S) = 0.25$ 25% of prisoners have serious ailments

Likelihoods:

$P(P \mid H) = 0.80$ 80% of healthy individuals pass the exam

$P(P \mid M) = 0.60$ 60% of individuals with minor ailments pass the exam

$P(P \mid S) = 0.30$ 30% of individuals with serious ailments pass the exam

Bayes Theorem Example

Total probability of passing the exam:

$$P(P) = P(P \mid H)P(H) + P(P \mid M)P(M) + P(P \mid S)P(S)$$

$$P(P) = (0.80 \times 0.25) + (0.60 \times 0.50) + (0.30 \times 0.25)$$

$$P(P) = 0.20 + 0.30 + 0.075 = 0.575$$

Bayes Theorem Example

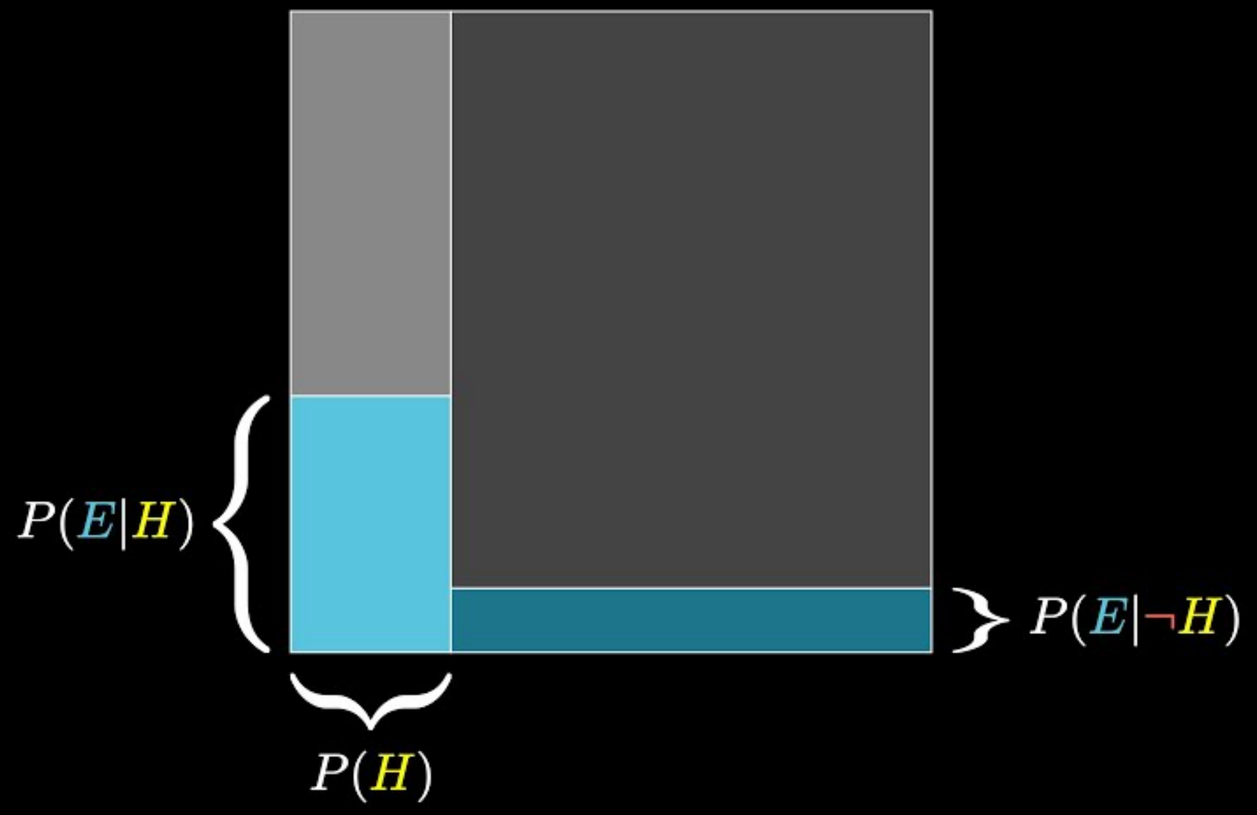
Now, using **Bayes' Theorem**, we can calculate the conditional probability $P(H \mid P)$:

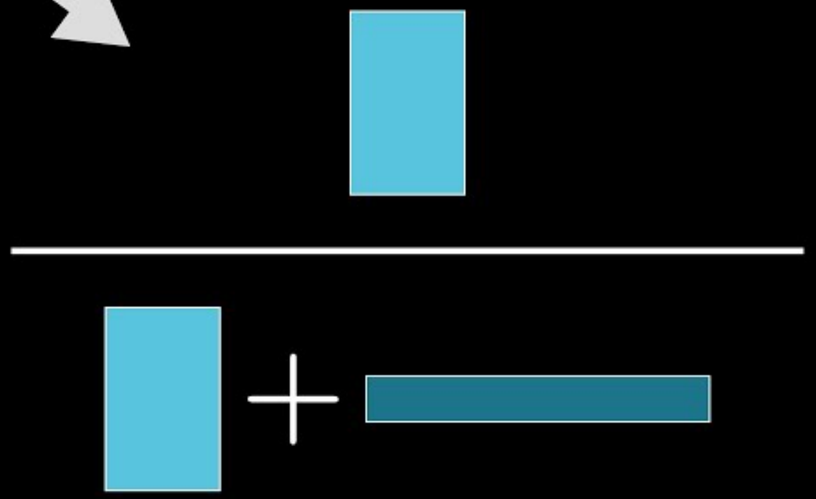
$$P(H \mid P) = \frac{P(P \mid H) \cdot P(H)}{P(P)}$$

Substitute the known values:

$$P(H \mid P) = \frac{(0.80 \times 0.25)}{0.575} = \frac{0.20}{0.575} \approx 0.3478$$

This is Bayes' rule





END