Probability Theory

Mustafa Shiple

Agenda

Introduction

Primary goals

Areas of growth

Timeline

Summary

Definitions

Sample Space

Sample Space : {1,2,3,4,5,6}

Countable sample space : {1,2,3,4,5,6}
uncountable : $0 < S < 5$

A={  }, B={  }

Unions : $A \cup B = \{1,3,4,5,6\}$

Intersection (joint) : $A \cap B = \{3,5\}$

Complements : $A^c = \{2,4,6\}$

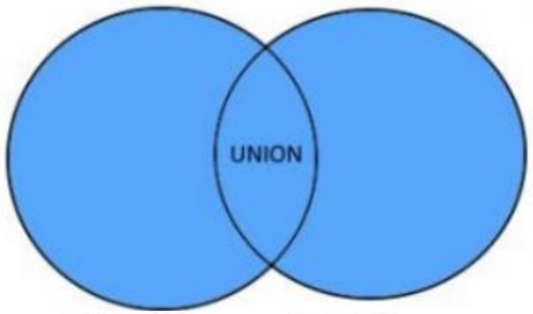
Definitions

A={  ,  ,  } , B={  ,  ,  ,  }

Event : *set of outcomes of the experiment (subset of sample space).*

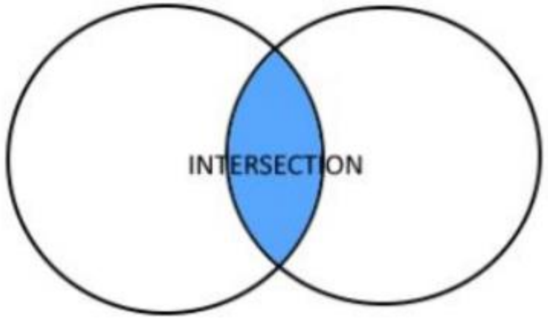
Marginal probability: *the probability of a single event occurring, independent of other events.*

$$P(A) = \frac{3}{6} = 0.5$$



Union probability: *the probability of two events.*

$$P(A \cup B) = \frac{5}{6}$$



Intersection (joint) probability: *the probability of two events.*

$$P(A \cap B) = \frac{2}{6}$$

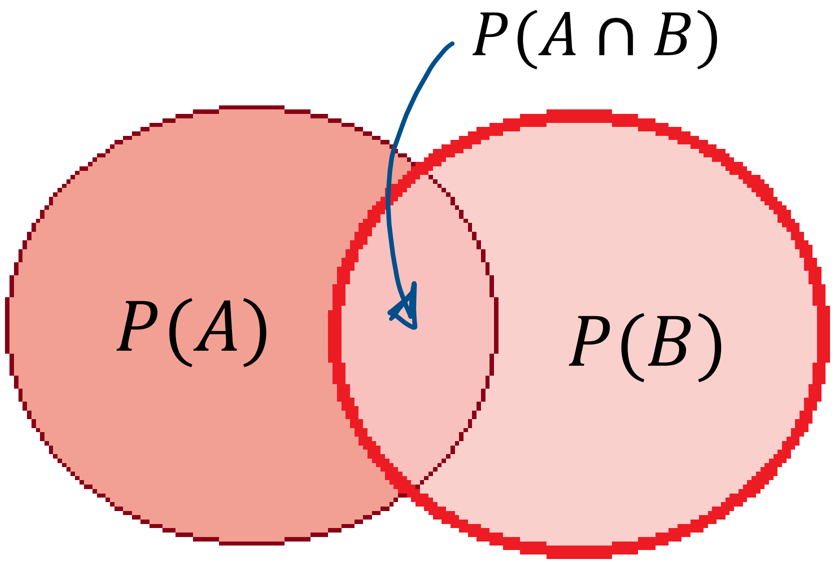
Definitions

A={ , ,  }, B={ , , , }

Conditional probability:

Given event

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{6} \times \frac{6}{4} = 0.5$$



Definitions

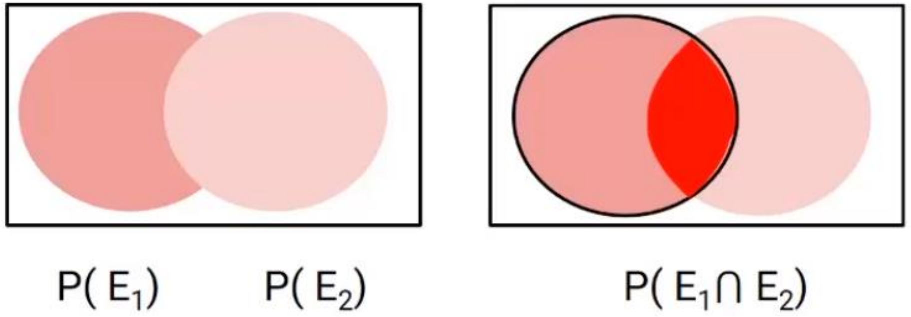
Example : calculate the probability of tossing two fair dice were the first one get (2) and the sum of two dice less than 5

$$P(D_2 = 2 | D_1 + D_2 \leq 5)$$

Event 1 $= P(D_2 = 2)$
Event 2 $= P(D_1 + D_2 \leq 5)$

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_1)}$$

$P(D_1 + D_2 \leq 5) =$



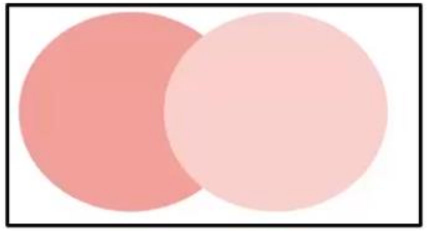
D1 →, D2 ↓

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

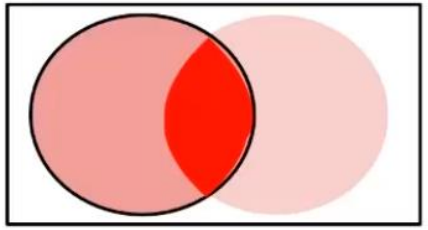
Definitions

Example : calculate the probability of tossing two fair dice were the first one get (2) and the sum of two dice less than 5

$$P(D_2 = 2 | D_1 + D_2 \leq 5)$$



$P(E_1)$ $P(E_2)$



$P(E_1 \cap E_2)$

Event 1 $= P(D_2 = 2)$
Event 2 $= P(D_1 + D_2 \leq 5)$

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

$$P(E_2 = D_1 + D_2 \leq 5) = \frac{10}{36}$$

$$P(E_1 \cap E_2) = \frac{3}{36}$$

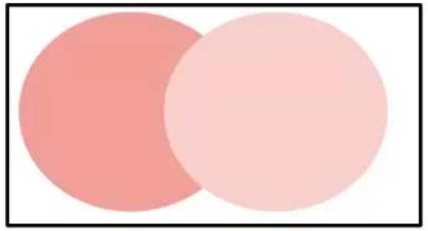
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

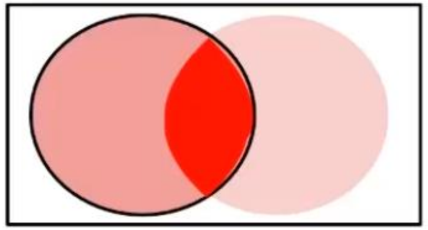
	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Definitions

Example : calculate the probability of tossing two fair dice were the first one get (2) and the sum of two dice less than 5

$$P(D_2 = 2|D_1 + D_2 \leq 5)$$





$P(E_1)$ $P(E_2)$ $P(E_1 \cap E_2)$

$Event\ 1 = P(D_2 = 2)$
$Event\ 2 = P(D_1 + D_2 \leq 5)$

$$P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$$

$$P(E_2 = D_1 + D_2 \leq 5) = \frac{10}{36}$$

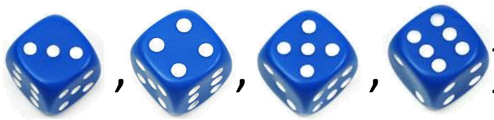
$$P(E_1 \cap E_2) = \frac{3}{36}$$

$$P(E_1|E_2) = \frac{3/36}{10/36} = 0.3$$

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Definitions

A={  } , B={  }

Conditional probability:

Given event

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2}{6} \times \frac{6}{4} = 0.5$$

Multiplication Rule:

$$P(A \cap B) = P(B)\, P(A|B)$$

Intersection (joint)

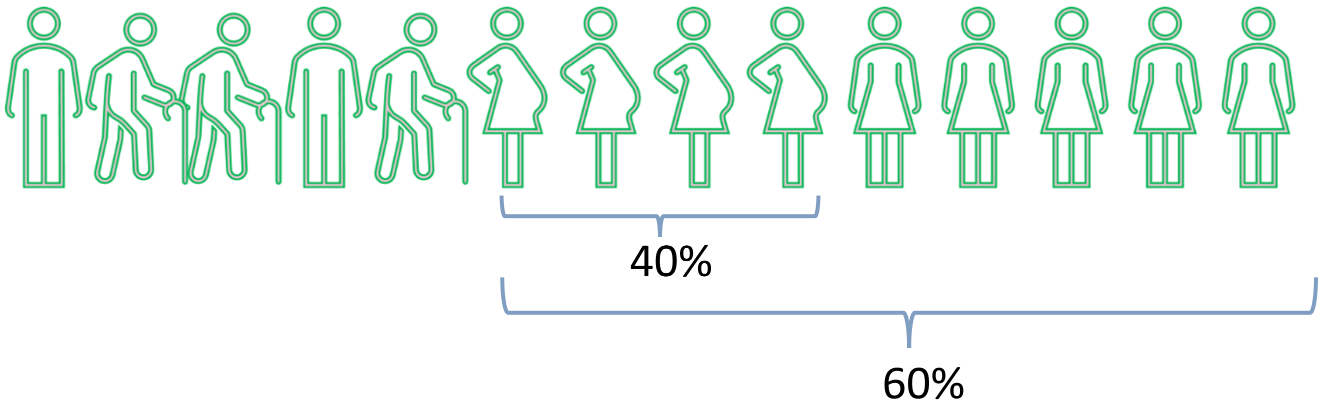
Conditional

select someone from the entire group who has two characteristics.

pull out a subgroup that has one of the characteristics already, and you want the probability that someone from that subgroup has a second characteristic.

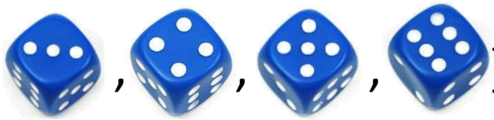
Exercise

Suppose there is an event 60% of attendance are woman (W) and 40% of them are pregnant (Pr). What the chance of a person you select from this event is a woman AND pregnant



$$P(W \cap Pr) = P(W)\, P(W|Pr)$$
$$= 0.6 \times 0.4 = 0.24$$

Definitions

A={  } , B={  }

Additional Rule:

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

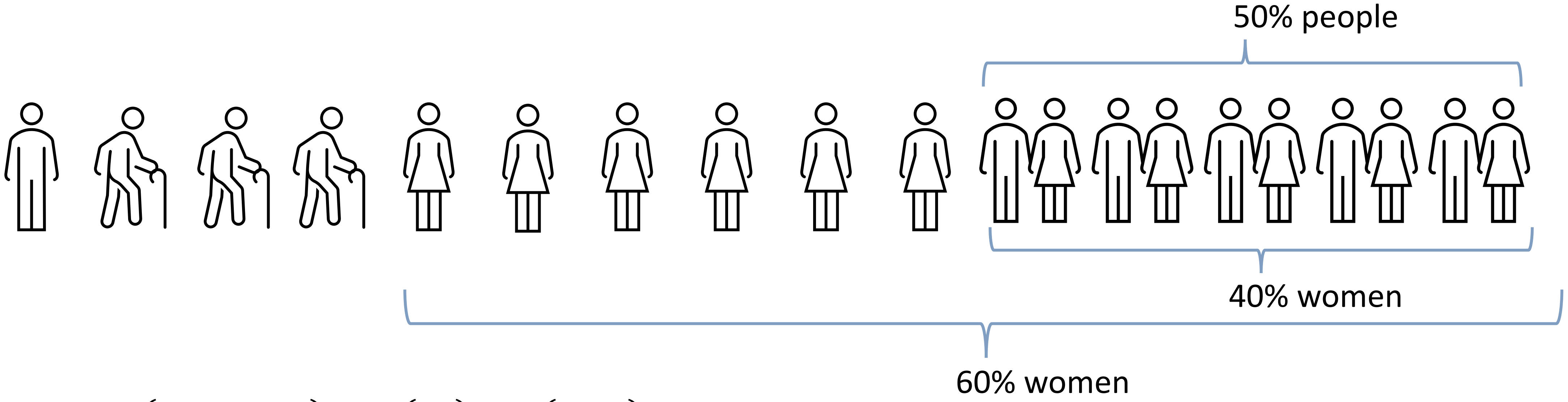
union

Intersection
(joint)

select someone from the entire group who has two characteristics.

Exercise

Suppose there is an event 60% of attendance are woman (W) and 50% of class are married (M). What the chance of a person you select from this event is a woman OR married



$$P(W \cap MW) = P(W) \times P(MW)$$
$$= 0.6 \times 0.4 = 0.24$$
$$P(W \cup M) = P(W) + P(MP) - P(W \cap MW)$$
$$= 0.6 + 0.5 - 0.24 = 0.86$$

Bayes Theorem

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

$$P(A \cap B) = P(B)\, P(A|B) = P(B \cap A)$$

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

Bayes Theorem

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

Likelihood

prior

evidence

$$P(C_1 \mid x_1 \& x_2 \& x_3 \& x_4) = \frac{P(x_1 | C_1) * P(x_2 | C_1) * P(x_3 | C_1) * P(x_4 | C_1) * P(C_1)}{P(x_1) * P(x_2) * P(x_3) * P(x_4)}$$

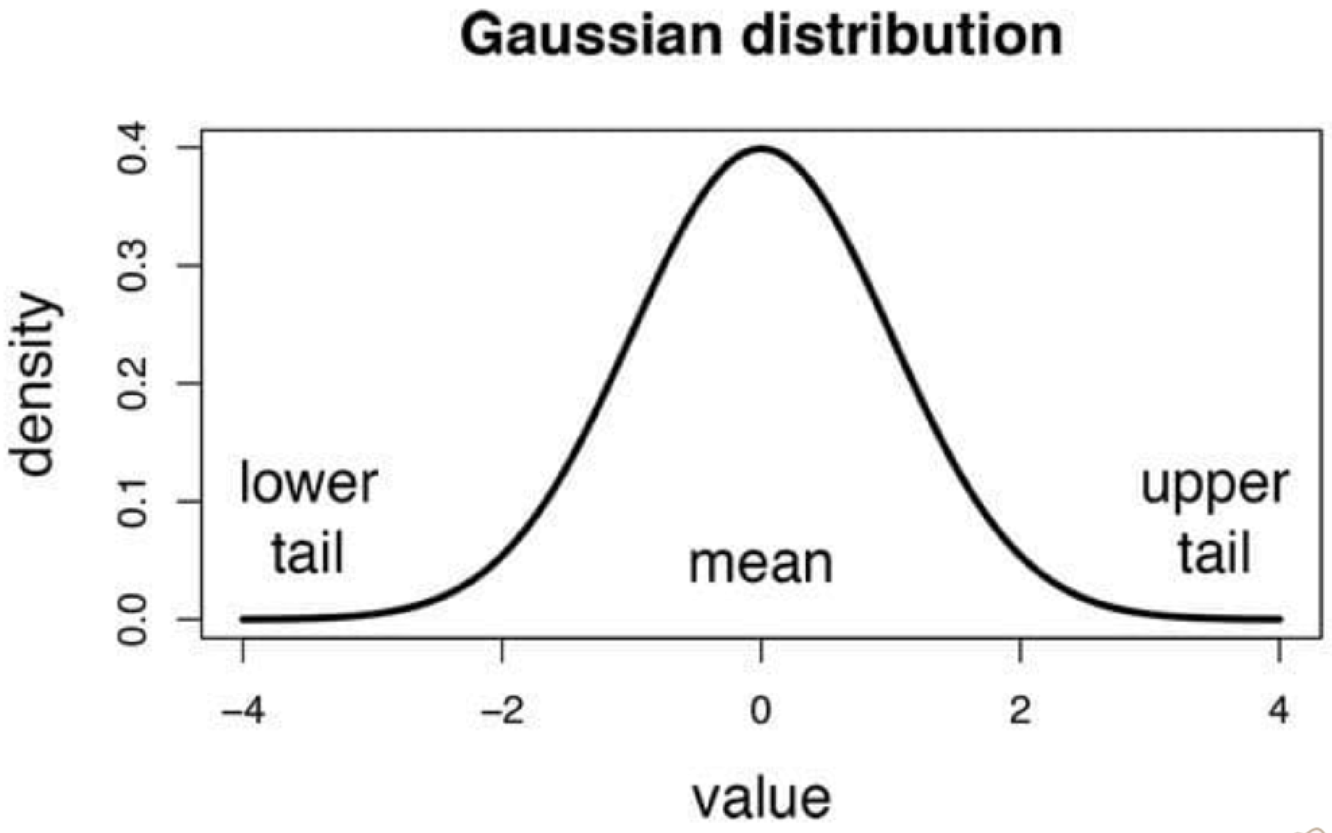
Gaussian Naïve Bayes Classifier:

continuous values associated with each feature are assumed to be distributed according to a **Gaussian distribution**

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

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

$$\sigma = \left[\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2\right]^{0.5}$$



Multinomial Naïve Bayes Classifier

Feature vectors represent the frequencies with which certain events have been generated by a **multinomial distribution**.
This is the event model typically used for document classification.

Bernoulli Naïve Bayes Classifier:

In the multivariate Bernoulli event model, features are independent booleans (binary variables) describing inputs.

binary term occurrence (i.e. a word occurs in a document or not)

Lab time

Annual revenue growth

Thank you