Bayes's Formula

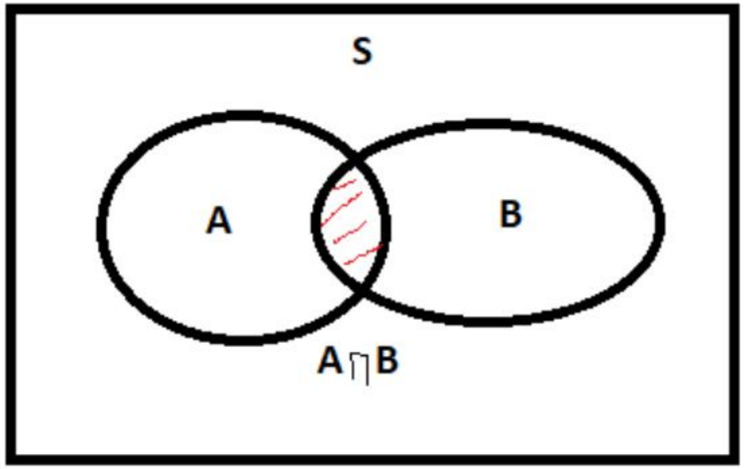
Recall

- **Conditional Probability**

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \textit{where } P(B) > 0$$

Conditional probability of $P(A|B)$ is undefined when $P(B) = 0$.

If P(B)=0, it means that the event B never occurs, so it does not make sense to talk about the probability of A given B.



- $P(A \mid B) = P(A \cap B)/P(B)$ **then**

$$P(A \cap B) = P(A \mid B)P(B)$$

- **If** $E_1, \ldots, E_n$ **are events then**

$$P(E_1 \cap E_2 \cap \cdots \cap E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2), \ldots, P(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1})$$

Practice Exercises

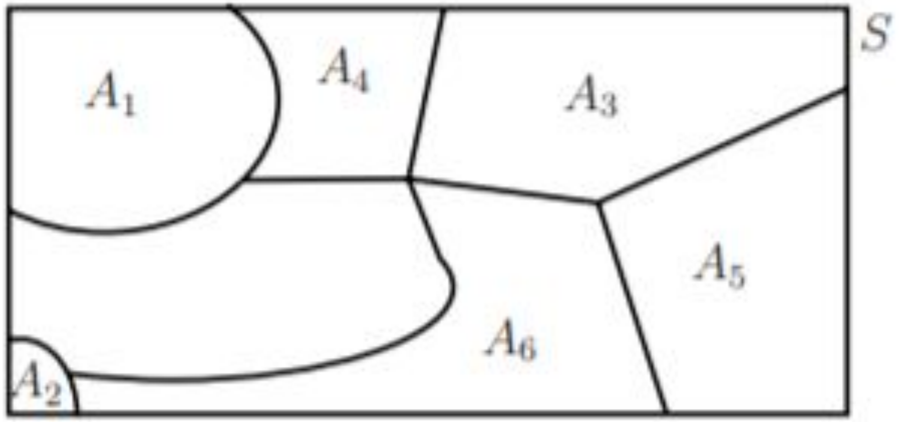
- Suppose an urn has 5 White balls and 7 Black balls. Each ball that is selected is returned to the urn along with an additional ball of the same color. Suppose draw 3 balls. **Question**: What is the probability that you get 3 white balls.

  $$P\ (3\ white\ balls)$$
  $$=\ P\ (1st\ W)\ P\ (2nd\ W\ |1st\ W)\ P\ (3nd\ W\ |\ 1st\ \&\ 2ndW)$$

- Phan wants to take a Biology course or a Chemistry course. Given that the students take Biology, the probability that they get an A is is $4/5$. While the probability of getting an A given that the student took Chemistry is $1/7$. If Phan makes a decision on the course to take randomly, what's probability of **getting an A in Chem**?
  - "Let B = "Takes Biology" and C = "Takes Chemistry" and A = "gets an A", then
  - $P(A \cap C) = P(C)P(A|C) = \frac{1}{2} \times \frac{1}{7}$

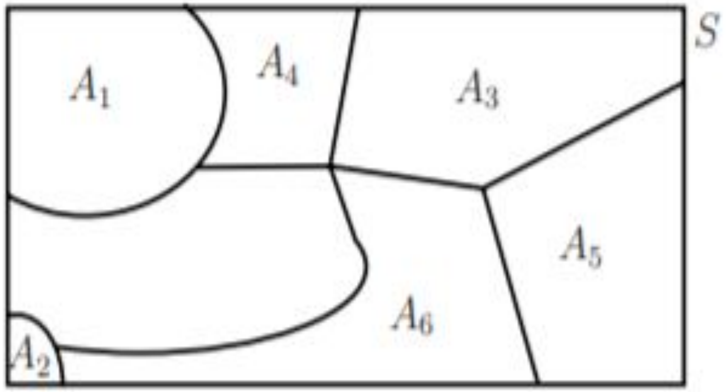
A partition of a sample space

- **The collection of events $A_1, A_2, ... A_n$ is said to partition a sample space S if**
  1. $A_1 \cup A_2 \cup ... \cup A_n = \mathbf{S}$ [Exhaustive]
  2. $A_i \cap A_j = \phi$, for all $i \neq j$ [non-overlapping]
  3. $A_i \neq \phi$, for all i [non-empty]



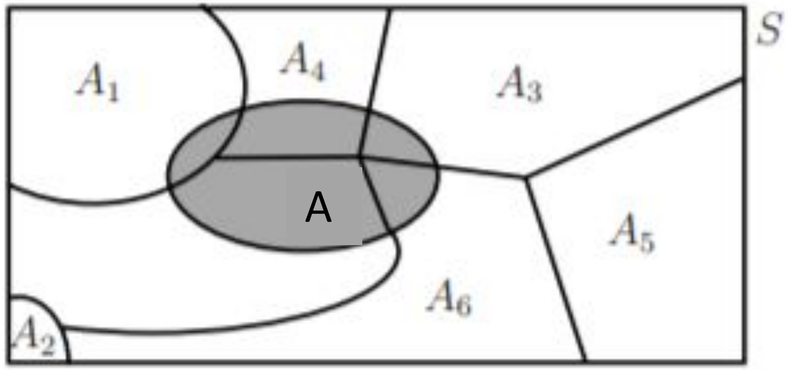
A partition of a sample space

- **The collection of events $A_1, A_2, ... A_n$ is said to partition a sample space S if**
  1. $\mathrm{A_1} \cup \mathrm{A_2} \cup ... \cup \mathrm{A_n} = \mathbf{S}$ [Exhaustive]
  2. $\mathrm{A_i} \cap \mathrm{A_j} = \boldsymbol{\phi}$, **for all** $\mathrm{i} \neq \mathrm{j}$ [non-overlapping]
  3. $\mathrm{A_i} \neq \boldsymbol{\phi}$, **for all** i [non-empty]



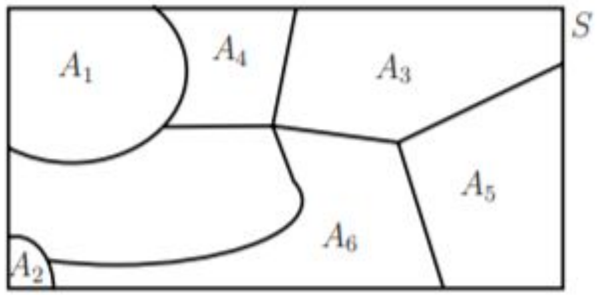
- **If A is any event within S then we can express $\boldsymbol{A}$ as the union of subsets:**

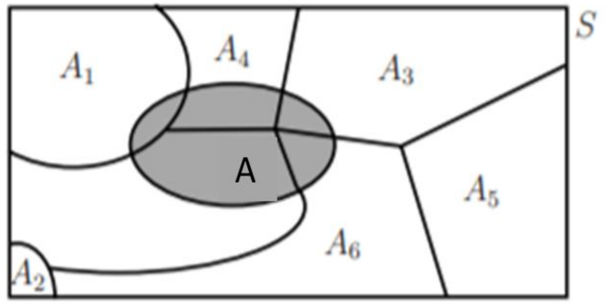
$$\boldsymbol{A} = (\boldsymbol{A} \cap A_1) \cup (\boldsymbol{A} \cap A_2) \cup \cdots \cup (\boldsymbol{A} \cap A_n)$$



A partition of a sample space

- The collection of events $A_1, A_2, \ldots A_n$ is said to partition a sample space S if
  1. $\mathrm{A_1} \cup \mathrm{A_2} \cup \ldots \cup \mathrm{A_n} = \mathbf{S}$ [Exhaustive]
  2. $\mathrm{A_i} \cap \mathrm{A_j} = \phi$, for all $\mathrm{i} \neq \mathrm{j}$ [non-overlapping]
  3. $\mathrm{A_i} \neq \phi$, for all i [non-empty]

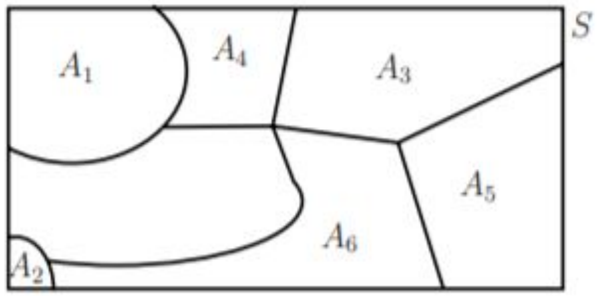


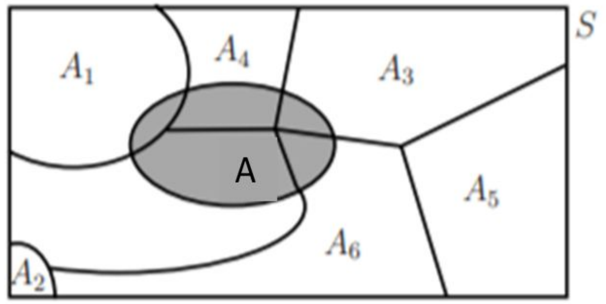


- $\boldsymbol{A} = (\boldsymbol{A} \cap A_1) \cup (\boldsymbol{A} \cap A_2) \cup \cdots \cup (\boldsymbol{A} \cap A_n)$
- Each of the bracket term $(\boldsymbol{A} \cap \boldsymbol{A_1}), (\boldsymbol{A} \cap A_2), \ldots (\boldsymbol{A} \cap A_n)$ are mutually exclusive.
  - if one occurs then none of the others can occur
  - $\boldsymbol{P}(\boldsymbol{A}) = \boldsymbol{P}(\boldsymbol{A} \cap A_1) \cup \boldsymbol{P}(\boldsymbol{A} \cap A_2) \cup \cdots \cup \boldsymbol{P}(\boldsymbol{A} \cap A_n)$

A partition of a sample space

- The collection of events $A_1, A_2, ... A_n$ is said to partition a sample space S if
- $\boldsymbol{A} = (\boldsymbol{A} \cap A_1) \cup (\boldsymbol{A} \cap A_2) \cup \cdots \cup (\boldsymbol{A} \cap A_n)$
- Each of the bracket term $(\boldsymbol{A} \cap \boldsymbol{A_1}), (\boldsymbol{A} \cap A_2), \ldots (\boldsymbol{A} \cap A_{\boldsymbol{n}})$ are mutually exclusive.
  - if one occurs then none of the others can occur
  - $\boldsymbol{P(A)} = \boldsymbol{P}(\boldsymbol{A} \cap A_1) \cup \boldsymbol{P}(\boldsymbol{A} \cap A_2) \cup \cdots \cup \boldsymbol{P}(\boldsymbol{A} \cap A_n)$
  - $\boldsymbol{P(A)} = \sum_{j=1}^{n} \boldsymbol{P(A \cap A_j)}$
  - $\boldsymbol{P(A)} = \sum_{j=1}^{n} \boldsymbol{P\left(A \middle| A_j\right) P(A_j)}$, in terms of conditional probabilities

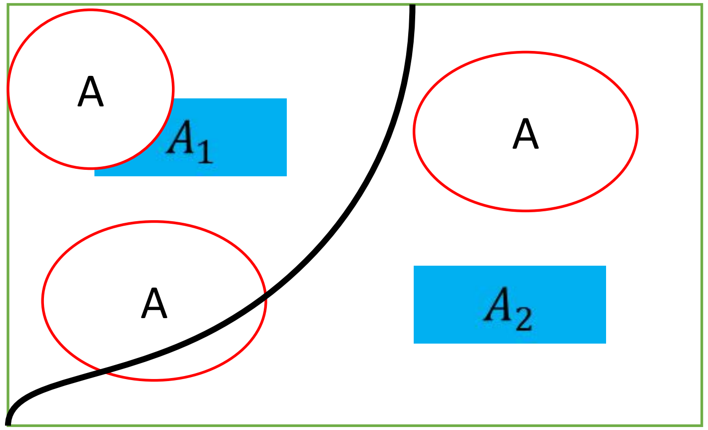




Total Probability Theorem

- **We can relate the unconditional probability of an event A with its conditional probabilities**
- **By considering a partition of the sample space S with mutually exclusive and exhaustive events $A_1, A_2, \dots, A_n$.**

S: Sample Space



$$\mathbf{P(A)} = \sum_{j=1}^{n} P(A|A_j)P(A_j)$$

A_1 and A_2 are mutually exclusive and exhaustive events

This is the theorem of Total Probability

Total Probability Theorem

- Suppose a class contains 60% girls and 40% boys. Suppose that 30% of the girls have long hair, and 20% of the boys have long hair. A student is chosen uniformly at random from the class. What is the probability that the chosen student will have long hair?
- Solution
  - Let A_1 be the set of girls and A_2 be the set of boys.
  - Then {A1, A2} is a partition of the class.
  - Let B be the set of all students with long hair and we are interested in P(B).

$$\boldsymbol{P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)}$$
$$\boldsymbol{= (0.3)(0.6) + (0.2)(0.4) = 0.26}$$

Total Probability Theorem

- In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that she knows the correct answer and $(1 - p)$ the probability that she guesses.
- Assume that a student who guesses at the answer will be correct with probability $1/m$, where m is the number of multiple-choice alternatives.
- What is the conditional probability that a student knew the answer to a question given that she answered it correctly?

- Solution
  - A: Student knows the answer.
  - B: Student guess the answer.
  - C: correct answer
  - $P(A) = p, P(B) = 1 - p$
  - $P(C|A) = 1, P(C|B) = \frac{1}{m}$
  - $P(A|C) = \frac{P(A \cap C)}{P(C)}$?
  - $P(C) = P(C|A)P(A) + P(C|B)P(B)$

$$= (1)(p) + \left(\frac{1}{m}\right)(1 - p) = \frac{mp - p + 1}{m}$$

  - $P(A \cap C) = P(C|A)P(A) = (1)(p)$
  - $P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{pm}{mp - p + 1}$

Bayes's Formula

- $P(H \cap D) = P(D|H).P(H)$, P(H)>0
- $P(D \cap H) = P(H|D).P(D)$, P(D)>0
- So,

$$P(H|D).P(D) = P(D|H).P(H)$$

- Dividing through by $P(D)$ we get

$$\boxed{P(H|D) = \frac{P(D|H).P(H)}{P(D)}}$$

Bayes' law

Applications of Bayes' Theorem: Machine Learning

1. Naive Bayes Classifier: Find posterior probabilities of Classes
2. Bayesian Belief Networks : Reasoning with uncertain knowledge

This law is important in the cases where we want to know the number on the left, and we do know (or can guess) the numbers on the right.

Bayes's Formula

- $$P(H|D) = \frac{P(D|H).P(H)}{P(D)}$$
- P(H)
  - Initial probability that hypothesis H holds before we have observed the training data.
- P(D)
  - Prior probability of training data D, i.e., the probability of D given no knowledge about which hypothesis holds.
- P(D|H)
  - Probability of D given H
- P(H|D)
  - Probability of H given D (posterior probability)

Evaluation of posteriori probability in terms of a priori information

Bayes's Formula

- **Question**
  - A laboratory blood test is 99 percent effective in detecting a certain disease when it is, in fact, present.
  - However, the test also yields a “false positive” result for 0.5% percent of the healthy persons tested.
  - If 0.1 percent of the population actually has the disease, what is the probability a person has the disease given that his test result is positive?

- **Solution**
  - A: Person has the disease.
  - B: Person does not have the disease.
  - C: Test result is positive
  - $P(A|C) = \frac{P(A\cap C)}{P(C)} = \frac{P(C|A)P(A)}{P(C)}$
  - $P(A) = \frac{0.1}{100} = 0.001, P(B) = 0.999$
  - $P(C|A) = 0.99, \ P(C|B) = \frac{0.5}{100} = 0.005$
  - $P(C) = P(C|A)P(A) + P(C|B)P(B)$

$= (0.99)(0.001) + (0.005)(0.999)$

$$P(A|C) = \frac{(0.99)(0.001)}{(0.99)(0.001) + (0.005)(0.999)}$$

Bayes's Formula

- $$P(H|D) = \frac{P(D|H).P(H)}{P(D)}$$
- Question
  - A simple binary communication channel carries messages by using only two signals, say 0 and 1.
  - We assume that, for a given binary channel, 40% of the time a 0 is transmitted
  - The probability that a transmitted 0 is correctly received is 0.95, and the probability that a transmitted 1 is correctly received is 0.90
  - Determine
    1. the probability of a 1 being received, and
    2. given a 1 is received, the probability that 1 was transmitted.

- Solution
  - A: Event of transmitting 1
  - A': Event of receiving 1
  - B: Event of transmitting 0
  - B': Event of receiving 0
- P(B) = 0.4, then P(A) = 0.6
- P(B'|B) = 0.95, then P(A'|B) = 0.05
- P(A'|A) = 0.90, then P(B'|A) = 0.10

1. $\mathrm{P(A') = P(A'|A)P(A) + P(A'|B)P(B)}$
   $= 0.90 \times 0.6 + 0.05 \times 0.4 = 0.56$
2. $\mathrm{P(A|A')} = \dfrac{P(A'|A)P(A)}{P(A')}$

$$= \frac{0.90 \times 0.6}{0.56}$$

Practice Exercises

- A certain letter is equally likely to be in any one of three different folders. Let p_i be the probability that you will find the letter upon making a quick examination of folder i if the letter is, in fact, in folder $i, i = 1, 2,3$; $(p_i < 1)$. Suppose you look in folder 1 and do not find the letter. What is the probability that the letter is in folder 1?
- A new computer program consists of two modules. The first module contains an error with probability 0.2. The second module is more complex; it has a probability of 0.4 to contain an error, independently of the first module. An error in the first module alone causes the program to crash with probability 0.5. For the second module, this probability is 0.8. If there are errors in both modules, the program crashes with probability 0.9. Suppose the program crashed. What is the probability of errors in both modules?

Naive Bayes Classifier

- Assume target function $f : X \rightarrow V$, where each instance x described by attributes $(a_1, a_2 , \dots , a_n)$.
- Most probable value of $f(x)$ is:

$$V_{MAP} = argmax_{v_j \in V}\, p\left(v_j \middle| a_1, a_2, \dots, a_n\right)$$

$$= argmax_{v_j \in V} \frac{p\left(a_1, a_2, \dots, a_n \middle| v_j\right) p(v_j)}{\boxed{p(a_1, a_2, \dots, a_n)}}$$

$$= argmax_{v_j \in V}\, p\left(a_1, a_2, \dots, a_n \middle| v_j\right) p(v_j)$$

Traning Dataset

Age	Income	Student	Credit rating	Buys compter ?
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
$31 \dots 40$	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
$31 \dots 40$	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
> 40	medium	yes	fair	yes
≤ 30	medium	yes	excellent	yes
$31 \dots 40$	medium	no	excellent	yes
$31 \dots 40$	high	yes	fair	yes
> 40	medium	no	excellent	no

$X = (age \leq 30, income = medium, Student = yes$,*Credit-rating = fair)?*

$f : X \rightarrow \{yes, no\}$

Constant

Naive Bayes Classifier

- $= argmax_{v_j \in V} p(a_1, a_2, \dots, a_n | v_j) p(v_j)$

- **Naive Bayes assumption:**

$$p(a_1, a_2, \dots, a_n | v_j) = \prod_i p(a_i | v_j)$$

- **Which gives, Naive Bayes classifier:**

$$V_{NB} = p(v_j) \prod_i p(a_i | v_j)$$

Traning Dataset

Age	Income	Student	Credit rating	Buys compter ?
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
$31\dots40$	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
$31\dots40$	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
> 40	medium	yes	fair	yes
≤ 30	medium	yes	excellent	yes
$31\dots40$	medium	no	excellent	yes
$31\dots40$	high	yes	fair	yes
> 40	medium	no	excellent	no

$X = (age \leq 30, income = medium, Student = yes$, Credit-rating = fair$)$?

$argmax_{v_j \in V} p(v_j) \times p(age \leq 30 | v_j) \times p(income = medium | v_j)$
$\times p(, Student = yes | v_j) p($Credit-rating = fair$| v_j)$

Naive Bayes Classifier

$$argmax_{v_j \in V} p(v_j) \times p(age \le 30|v_j) \times p(income = medium|v_j) \times p(,Student = yes|v_j) p(Credit\text{-}rating = fair|v_j)$$

$$P(age = '\le 30' | Buyscomputers = 'yes') = 2/9 = 0.222$$
$$P(age = '\le 30' | Buyscomputers = 'no') = 3/5 = 0.600$$

$$P(Income = 'medium' | Buyscomputers = 'yes') = 4/9 = 0.444$$
$$P(Income = 'medium' | Buyscomputers = 'no') = 2/5 = 0.4$$

$$P(Student = 'yes' | Buyscomputers = 'yes') = 6/9 = 0.0.667$$
$$P(Student = 'yes' | Buyscomputers = 'no') = 1/5 = 0.2$$

$$P(Creditrating = 'fair' | Buyscomputers = 'yes') = 6/9 = 0.0.667$$
$$P(Creditrating = 'fair'' | Buyscomputers = 'no') = 2/5 = 0.4$$

X belongs to class 'buys computer = 'yes'.

$\boldsymbol{P(X|C_i) \times P(C_i)}$:

$P(X|Buyscomputers = 'yes') \times P(Buyscomputer = 'yes') = 0.028\ (0.044 \times 9/14)$

$P(X|Buyscomputers = 'no') \times P(Buyscomputer = 'no') = 0.007\ (0.019 \times 5/14)$

Traning Dataset

Age	Income	Student	Credit rating	Buys compter ?
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
$31 \ldots 40$	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
$31 \ldots 40$	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
> 40	medium	yes	fair	yes
≤ 30	medium	yes	excellent	yes
$31 \ldots 40$	medium	no	excellent	yes
$31 \ldots 40$	high	yes	fair	yes
> 40	medium	no	excellent	no

Reference

- **Statistics with Economics and Business Applications, Chapter 3 Probability and Discrete Probability Distributions**
- **Modern Business Statistics, Slides by John Loucks**
- lecture notes on Probability Theory by Phanuel Mariano