BITS Pilani

Pilani|Dubai|Goa|Hyderabad

M Tech(Data Science & Engineering)

Introduction to Statistical Methods

Team ISM

innovate achieve lead

BITS Pilani

Session No 2

Probability basics , Conditional Probability and Bayes Theorem

(Session 2: 14th/15th May 2022)

Agenda

- Experiments, Counting Rules, and Assigning Probabilities
- Events and Their Probability
- Some Basic Relationships of Probability
- Conditional Probability
- Bayes' Theorem

Random Experiment

Term "**random experiment**" is used to describe any action whose outcome is not known in advance. Here are some examples of experiments dealing with statistical data:

- Tossing a coin
- Counting how many times a certain word or a combination of words appears in the text of the "King Lear" or in a text of Confucius
- counting occurrences of a certain combination of amino acids in a protein database.
- pulling a card from the deck

Sample spaces, sample sets and events

The ***sample space*** of a random experiment is a set S that includes all possible outcomes of the experiment.

For example, if the experiment is to throw a die and record the outcome, the sample space is S = { 1,2,3,4,5,6}

- Discrete sample spaces.
- Continuous sample spaces

Discrete Random Variables

- A discrete random variable is one which may take on only a countable number of distinct values such as 0,

  1, 2, 3, 4,....

- • Discrete random variables are usually (but not necessarily) counts.

Examples:

- number of children in a family
- the Friday night attendance at a cinema
- the number of patients a doctor sees in one day
- the number of defective light bulbs in a box of ten
- the number of “heads” flipped in 3 trials

Continuous Random Variable

- A continuous random variable is one which takes an infinite number of possible values.

- Examples:
  - ✓ height
  - ✓ weight
  - ✓ the amount of sugar in an orange
  - ✓ the time required to run a mile.

Event

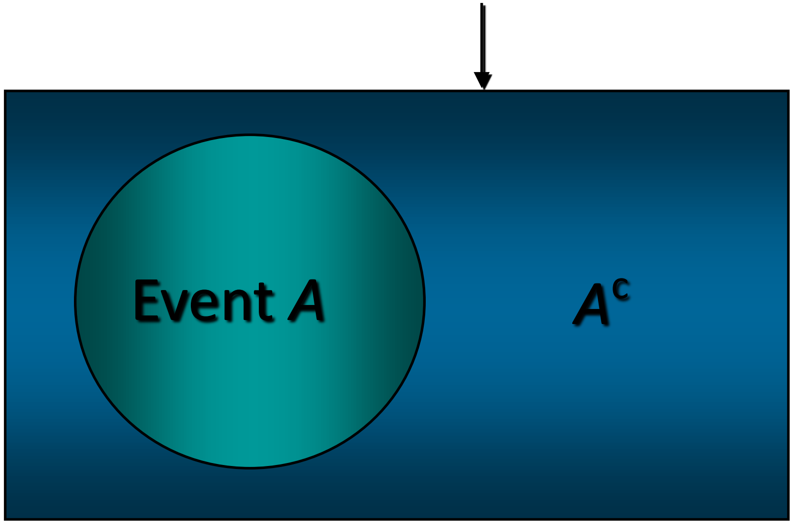
An **event** is a subset of the sample space of a random experiment.

An event is a set of outcomes of the experiment. This includes the *null* (empty) set of outcomes and the set of *all* outcomes. Each time the experiment is run, a given event *A* either *occurs*, if the outcome of the experiment is an element of *A*, or *does not occur*, if the outcome of the experiment is not an element of *A*.

Complement of an Event

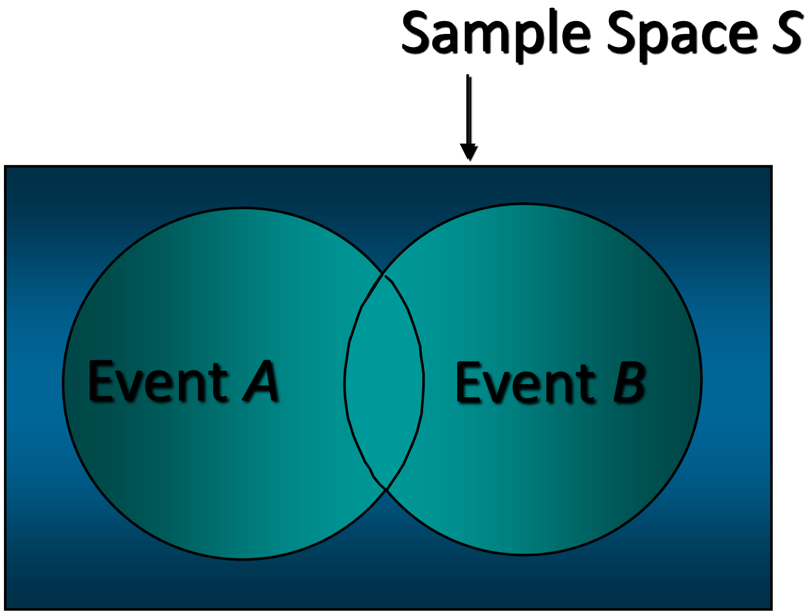
- The <u>complement</u> of event *A* is defined to be the event consisting of all sample points that are not in *A.*
- The complement of *A* is denoted by A^c.
- The <u>Venn diagram</u> below illustrates the concept of a complement.

Sample Space *S*



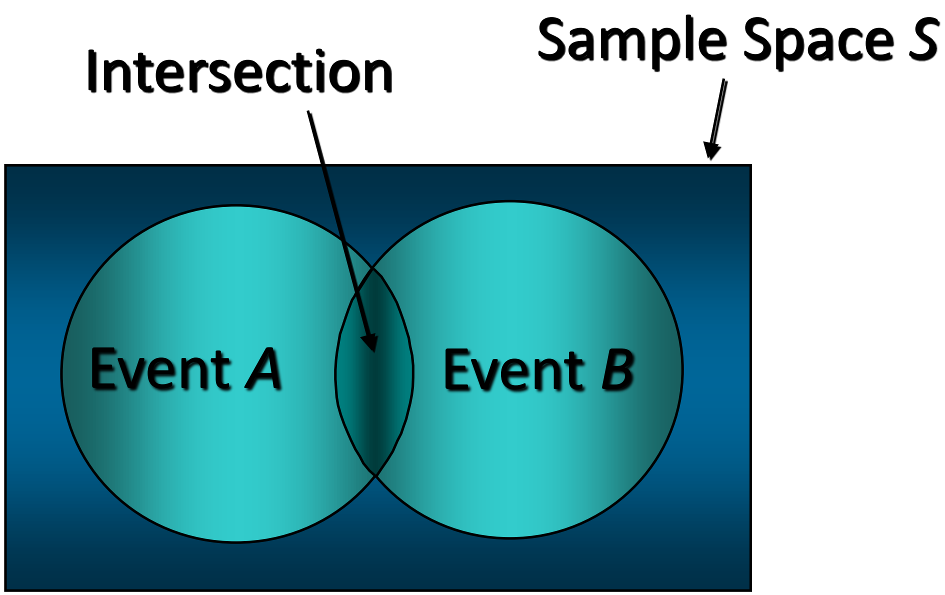
Union of Two Events

- The <u>union</u> of events *A* and *B* is the event containing all sample points that are in *A* or *B* or both.
- The union is denoted by *A* U *B*
- The union of *A* and *B* is illustrated below.



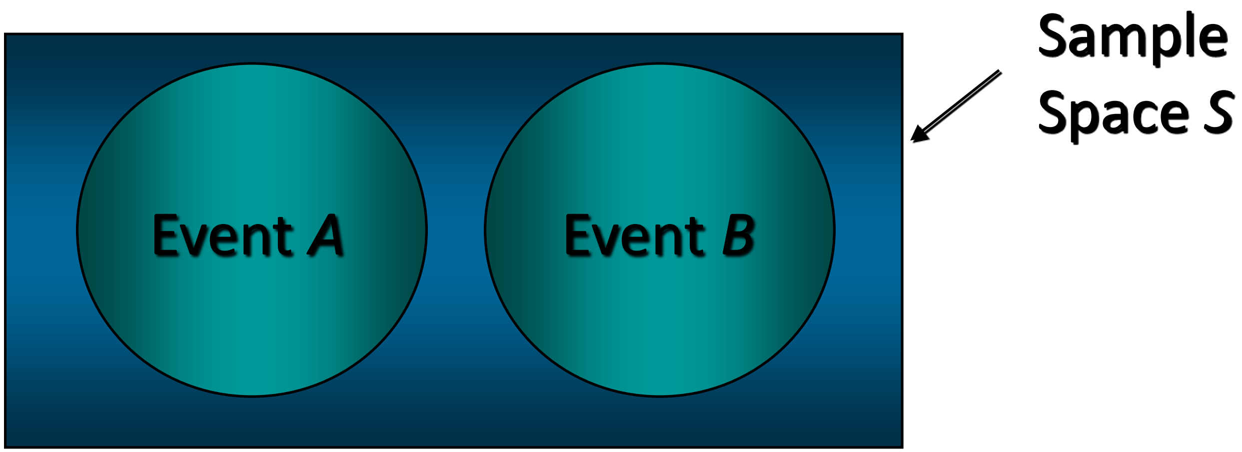
Intersection of Two Events

- The <u>intersection</u> of events *A* and *B* is the set of all sample points that are in both *A* and *B*.
- The intersection of *A* and *B* is the area of overlap in the illustration below.



Mutually Exclusive Events

- Two events are said to be mutually exclusive if the events have no sample points in common. That is, two events are mutually exclusive if, when one event occurs, the other cannot occur.



Independent & Dependent

Events are either

- *independent* (the occurrence of one event has no effect on the probability of occurrence of the other) or
- *dependent* (the occurrence of one event gives information about the occurrence of the other)

Classical Probability

Probability is a numerical measure of the likelihood that an event will occur.

CLASSICAL PROBABILITY

$$\text{Probability of an event} = \frac{\text{Number of favorable outcomes}}{\text{Total number of possible outcomes}}$$

Empirical Probability

Empirical or **relative frequency** is the second type of objective probability. It is based on the number of times an event occurs as a proportion of a known number of trials.

EMPIRICAL PROBABILITY The probability of an event happening is the fraction of the time similar events happened in the past.

In terms of a formula:

$$\text{Empirical probability} = \frac{\text{Number of times the event occurs}}{\text{Total number of observations}}$$

The empirical approach to probability is based on what is called the law of large numbers. The key to establishing probabilities empirically is that more observations will provide a more accurate estimate of the probability.

LAW OF LARGE NUMBERS Over a large number of trials, the empirical probability of an event will approach its true probability.

Axioms of Probability

Probability is a number that is assigned to each member of a collection of events from a random experiment that satisfies the following properties:

If S is the sample space and E is any event in a random experiment,

(1) $P(S) = 1$

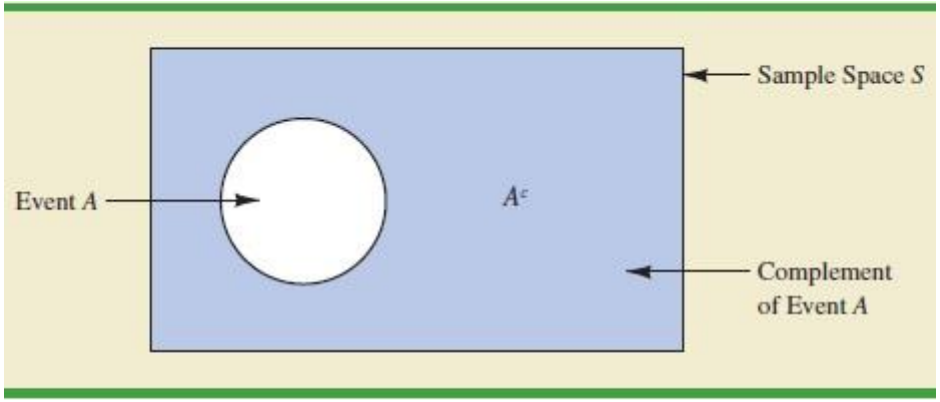
(2) $0 \leq P(E) \leq 1$

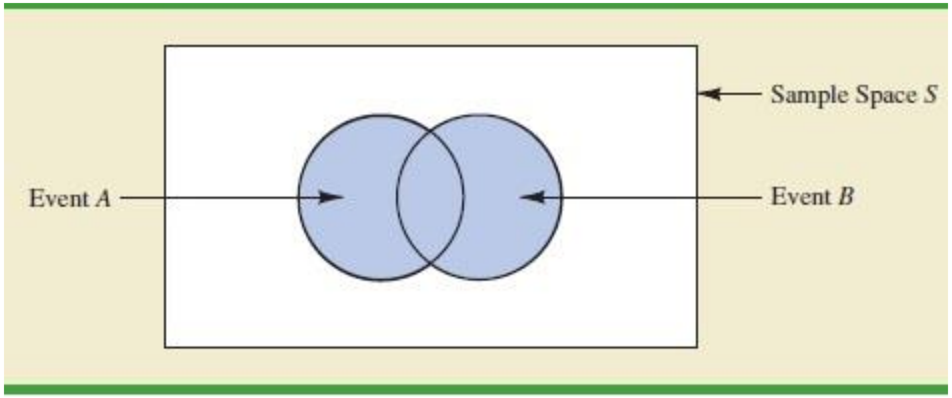
(3) For two events E_1 and E_2 with $E_1 \cap E_2 = \emptyset$

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

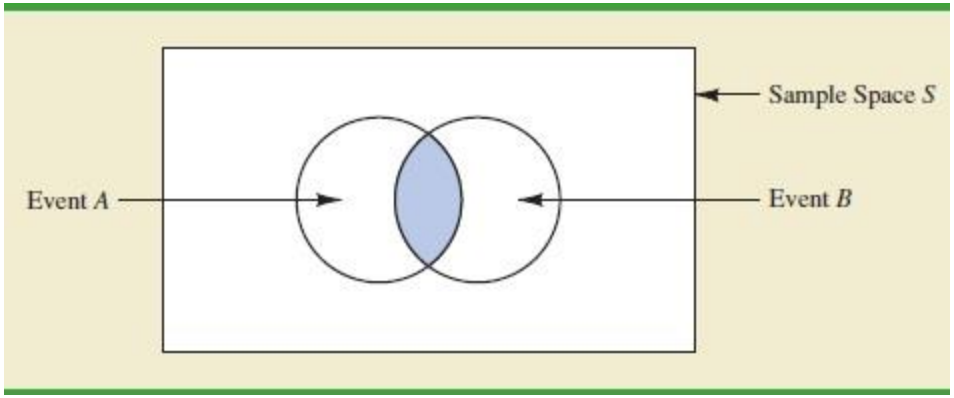
Axioms of Probability

$$P(A) = 1 - P(A^c)$$

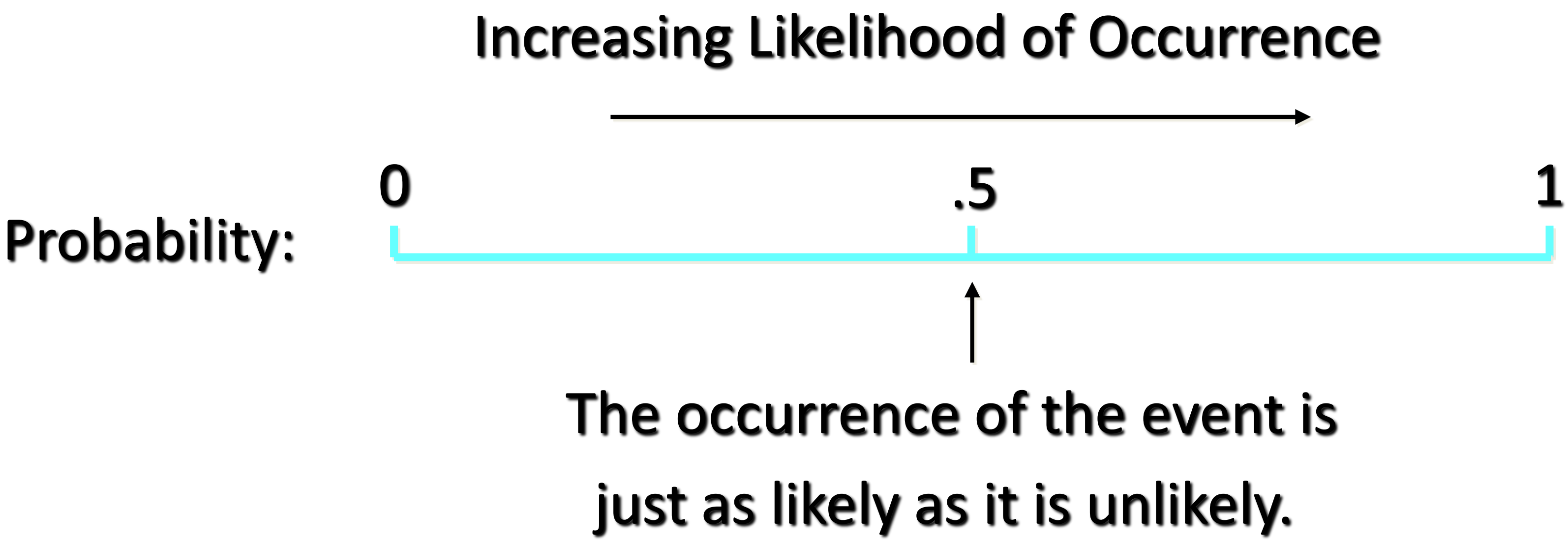




$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



Probability as a Numerical Measure of the Likelihood of Occurrence



Example: 1

The sales manager of an e commerce company says that 80% of those who visit their website for the first time do not buy any mobile. If a new customer visits the website, what is the probability that the customer would buy mobile

Example 2

	Blue	Black	Brown	Total
Software prog	35	25	20	80
Project Mgrs	7	8	5	20
Total	42	33	25	100

- If an employee is selected at random , what is the probability that he is a software prog?
- If an employee is selected at random , what is the probability that he is wearing a blue trouser

Example 3

A Survey conducted by a bank revealed that 40% of the accounts are savings accounts and 35% of the accounts are current accounts and the balance are loan accounts.

- What is the probability that an account taken at random is a loan account ?
- What is the probability that an account taken at random is **NOT** savings account ?
- What is the probability that an account taken at random is **NOT** a current account
- What is the probability that an account taken at random is a current account or a loan account?

Example 4

In a certain residential hub,60% of all households get internet service from the local cable company,80% get the television service from that company, and 50% get both services from that company.

If a household is randomly selected, what is the probability that it gets at least one of these two services from the company, and what is the probability that it gets exactly one of these services from the company?

EXAMPLE 5

A problem in statistics is given to 3 students P,Q and R whose chances of solving it are ½.3/4 and ¼ respectively.

- What is the probability that the problem is solved?

EXAMPLE 6

A speaks truth in 80% cases and B speaks in 60% cases. What percentage of cases are they likely to contradict each other in stating the same fact.

EXAMPLE 7

If two dice are thrown , what is the probability that the sum is

a) Greater than 8
b) Less than 6
c) Neither 7 nor 11

Example

$$P(X = 2) = P\{(1, 1)\} = \frac{1}{36}$$

$$P(X = 3) = P\{(1, 2), (2, 1)\} = \frac{2}{36}$$

$$P(X = 4) = P\{(1, 3), (2, 2), (3, 1)\} = \frac{3}{36}$$

$$P(X = 5) = P\{(1, 4), (2, 3), (3, 2), (4, 1)\} = \frac{4}{36}$$

$$P(X = 6) = P\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\} = \frac{5}{36}$$

$$P(X = 7) = P\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\} = \frac{6}{36}$$

$$P(X = 8) = P\{(2, 6), (3, 5), (4, 4), (5, 3), (6, 2)\} = \frac{5}{36}$$

$$P(X = 9) = P\{(3, 6), (4, 5), (5, 4), (6, 3)\} = \frac{4}{36}$$

$$P(X = 10) = P\{(4, 6) (5, 5), (6, 4)\} = \frac{3}{36}$$

$$P(X = 11) = P\{(5, 6), (6, 5)\} = \frac{2}{36}$$

$$P(X = 12) = P\{(6, 6)\} = \frac{1}{36}$$

If two dice are thrown , what is the probability that the sum is

a) Greater than 8

b) Less than 6

c) Neither 7 nor 11

If two dice are thrown , what is the probability that the sum is

a) Greater than 8

b) Less than 6

c) Neither 7 nor 11

EXAMPLE 8

The probability that ‘A’ will be alive 10 years hence is 5/8 and that B will be alive is ¾. Find the probability that

a) At least one is alive
b) Exactly one is alive
c) None are alive

Conditional Probability

Conditional Probability

- We will use the notation $P(A \mid B)$ to represent the **conditional probability of *A* given that the event *B* has occurred.** B is the "conditioning event"

  As an example, consider the event A that a randomly selected student at your university obtained all desired classes during the previous term's registration cycle. Presumably $P(A)$ is not very large.

However, suppose the selected student is an athlete who gets special registration priority (the event B). Then $P(A \mid B)$ should be substantially larger than $P(A)$, although perhaps still not close to 1.

The Definition of Conditional Probability

Definition

For any two events A and B with $P(B) > 0$, the conditional probability of A given that B has occurred is defined by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \tag{2.3}$$

The Multiplication Rule for $P(A \cap B)$

The definition of conditional probability yields the following result, obtained by multiplying both sides of Equation (2.3) by $P(B)$.

The Multiplication Rule

The Multiplication Rule

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

This rule is important because it is often the case that $P(A \cap B)$ is desired, whereas both $P(B)$ and $P(A \mid B)$ can be specified from the problem description.

Conditional Probability

Independent Events

Example

Suppose that A and B are events with probabilities:

$P(A)=1/3$, $P(B)=1/4$, $P(A \cap B)=1/10$

Find each of the following:

- $P(A \mid B)$
- $P(B \mid A)$
- $P(A' \mid B')$

Solution

$P(A) = 1/3$, $P(B) = 1/4$, $P(A \cap B) = 1/10$

1. $P(A \mid B) = P(A \cap B)/P(B)$

   $= 1/10/1/4 = 4/10$

2. $P(B \mid A) = P(A \cap B)/P(A)$

   $= 1/10/1/3 = 3/10$

3. $P(A' \mid B') = P(A' \cap B')/P(B') = P((A \cup B)')/(1-P(B))$

   $= (1-P(A \cup B))/(1 – P(B))$

   $= (1 – (P(A)+P(B)-P(A \cap B)))/(1-P(B))$

   $= (1 – (1/3+1/4-1/10))/(1-1/10)$

   $=(1-29/60)/9/10 = 31/60/9/10 = 31/54.$

Example

Let E = the sum of the faces is even and S_2=the second die is a 2

Find

1. $P(S_2 \mid E)$

2. $P(E \mid S_2)=$

Solution

Let E=the sum of the faces is even and S_2=the second die is a 2

1. $P(S_2 \mid E) =$
   $P(S_2 \cap E) / P(E) =$
   $3/18 = 1/6$
2. $P(E \mid S_2) =$
   $3/6 = 1/2$

$$S = \begin{Bmatrix} (1,1) & (2,1) & (3,1) & (4,1) & (5,1) & (6,1) \\ (1,2) & (2,2) & (3,2) & (4,2) & (5,2) & (6,2) \\ (1,3) & (2,3) & (3,3) & (4,3) & (5,3) & (6,3) \\ (1,4) & (2,4) & (3,4) & (4,4) & (5,4) & (6,4) \\ (1,5) & (2,5) & (3,5) & (4,5) & (5,5) & (6,5) \\ (1,6) & (2,6) & (3,6) & (4,6) & (5,6) & (6,6) \end{Bmatrix}$$

EXAMPLE : 9

From a city population , the probability of selecting a male or a smoker is 7/10, a male smoker is 2/5, and a male if a smoker is already selected is 2/3.

Find the probability of selecting

a) A non – smoker

b) A male

c) A smoker, if a male is first selected

Example : 10

Suppose that of all individuals buying a certain digital camera, 60% include an optional memory card in their purchase, 40% include an extra battery, and 30% include both a card and battery. Consider randomly selecting a buyer and let

A = {memory card purchased} and
B = {battery purchased}.

Then $P(A) = .60$, $P(B) = .40$, $P(\text{both purchased}) = P(A \cap B) = .30$

Solution

- Given that the selected individual purchased an extra battery, the probability that an optional card was also purchased .

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{.30}{.40} = .75$$

That is, of all those purchasing an extra battery, 75% purchased an optional memory card. Similarly,

P(battery|memorycard)

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{.30}{.60} = .50$$

Example 11

A news magazine publishes three columns entitled "ART"(A),"BOOKS"(B) and "CINEMA:(C). Reading habits of a randomly selected reader with respect to these columns are

Read Regularly	A	B	C	A & B	A & C	B&C	A&B&C
Probability	0.14	0.23	0.37	0.08	0.09	0.13	0.05

a) Find the probability that he follows ART given that he read BOOKS regularly.

b) Find the probability that he follows ART given that he regularly follows at least BOOKS or CINEMA.

c) Find the probability that he follows ART given that he regularly follows at least one.

d) Find the probability that he follows atleast ART or BOOKS given that he follows cinema regularly.

Example-11

a) Find the probability that he follows ART given that he read BOOKS regularly.

b) Find the probability that he follows ART given that he regularly follows at least BOOKS or CINEMA

c) Find the probability that he follows ART given that he regularly follows at least one.

d) Find the probability that he follows atleast ART or BOOKS given that he follows cinema regularly.

For example, consider three events A_1, A_2, and A_3. The triple intersection of these events can be represented as the double intersection $(A_1 \cap A_2) \cap A_3$. Applying our previous multiplication rule to this intersection and then to $A_1 \cap A_2$ gives

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 | A_1 \cap A_2) \cdot P(A_1 \cap A_2) \\ &= P(A_3 | A_1 \cap A_2) \cdot P(A_2 | A_1) \cdot P(A_1) \qquad (2.4) \end{aligned}$$

Thus the triple intersection probability is a product of three probabilities, two of which are conditional.

The Law of Total Probability

Let $A_1, \ldots, A_k$ be mutually exclusive and exhaustive events. Then for any other event B,

$$P(B) = P(B|A_1)P(A_1) + \cdots + P(B|A_k)P(A_k)$$
$$= \sum_{i=1}^{k} P(B|A_i)P(A_i) \qquad (2.5)$$

EXAMPLE

- Three persons A,B and C are competing for the post of CEO of a company. The chances of they becoming CEO are 0.2,0.3 and 0.4 respectively.
- The chances of they taking employees beneficial decisions are 0.50,0.45 and 0.6 respectively
- What are the chances of having employees beneficial decisions after having new CEO

Example

An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3.

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively.

What is the probability that a randomly selected message is spam?

EXAMPLE

- Three persons A,B and C are competing for the post of CEO of a company. The chances of they becoming CEO are 0.2,0.3 and 0.4 respectively.
- The chances of they taking employees beneficial decisions are 0.50,0.45 and 0.6 respectively
- What are the probabilities of chances of having employees' beneficial decisions after having A as new CEO
- Similarly, B
- Similarly, C

Example

An individual has 3 different email accounts. Most of her messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3.

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively.

What is the probability that a randomly selected spam message belongs to account # 1

Example

In a city patients visits three doctors A,B and C in the ratio 6:5:7.The chance that these doctors refers a case to a specialist are 54%,60% and 55% respectively.

If a patient visits a specialist , what is the probability that he is referred by doctor A

Bayes' Theorem

Bayes' Theorem

Let $A_1, A_2, \ldots, A_k$ be a collection of k mutually exclusive and exhaustive events with *prior* probabilities $P(A_i)$ $(i = 1, \ldots, k)$. Then for any other event B for which $P(B) > 0$, the *posterior* probability of A_j given that B has occurred is

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{k} P(B|A_i) \cdot P(A_i)} \quad j = 1, \ldots, k \qquad (2.6)$$

WEATHER	PLAY	WEATHER	PLAY
SUNNY	YES	SUNNY	YES
OVERCAST	NO	OVERCAST	NO
RAINY	NO	RAINY	NO
SUNNY	YES	SUNNY	YES
SUNNY	NO	RAINY	NO
RAINY	YES	OVERCAST	YES

Thanks