Introduction to Statistical Methods

ISM Team

innovate achieve lead

BITS Pilani
Pilani Campus

Session 3:
Conditional Probability, Independent events and, Total Probability
(7-12-2024 / 8-12-2024)

IMP Note to Self

Start Recording

Important Note for Students

- All submissions for graded components must be the result of your original effort.
- It is strictly prohibited to copy and paste verbatim from any sources, whether online or from your peers.
- The use of unauthorized sources or materials, as well as collusion or unauthorized collaboration to gain an unfair advantage, is also strictly prohibited. Please note that we will not distinguish between the person sharing their resources and the one receiving them for plagiarism, and the consequences will apply to both parties equally.

Important Note for Students

- In cases where suspicious circumstances arise, such as identical verbatim answers or a significant overlap of unreasonable similarities in a set of submissions, will be investigated, and severe punishments will be imposed on all those found guilty of plagiarism.

Revise

- Introduction to Probability
- Review of Set Theory
- Counting Principles
- Definition of Probability
- Axioms of Probability
- Addition Rule of Probability

Contact Session 3

Contact Session 3: Module 2(Conditional Probability & Bayes theorem)

Contact Session	List of Topic Title	Reference
CS - 3	Introduction to conditional probability,indepents events, Total probability	T1 & T2
HW	Problems on conditional probability	T1 & T2
Lab		

Agenda

- Conditional Probability
- Independent Events
- Total Probability

Conditional Probability

- The probabilities assigned to various events depend on the experimental situation when the assignment is made.
- Subsequent to the initial assignment, partial information relevant to the outcome of the experiment may become available.
- Such information may cause us to revise some of our probability assignments.

Conditional Probability: Example

- We examine how the information "an event A has occurred" affects the probability assigned to B.
- For example, B might refer to an individual having a particular disease in the presence of certain symptoms.
- If a blood test is performed on the individual and the result is negative , then the probability of having the disease will change (it should decrease, but not usually to zero, since blood tests are not infallible).
- We will use the notation to represent the conditional probability of B given that the event A has occurred. A is the "conditioning event."

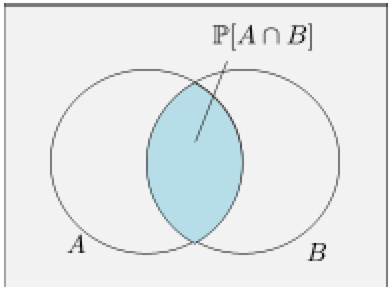
Conditional Probability

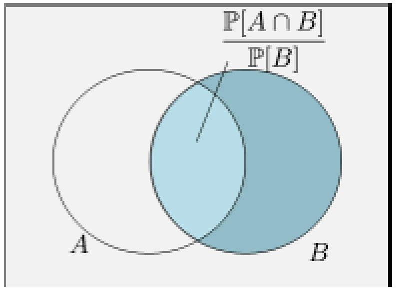
Let A and B be two events in sample space. The **conditional probability** that event B occurs given that event A has occurred and it is denoted by

$$P(B|A) = \frac{P(A \text{ and } B)}{P(A)} \quad \text{OR} \quad P(A|B) = \frac{P(A \cap B)}{P(B)}$$

It can also be written as $P(A \cap B) = P(A)\,P(B|A) \qquad P(A) \neq 0$

$= P(B)\,P(A|B) \qquad P(B) \neq 0$





Conditional Probability

Let A,B and C be three events in a sample space S,

then $\mathbf{P(A \cap B \cap C) = P(A)\, P(B|A)\, P(C|A \cap B)}$

and it is called **Multiplication Rule.**

Conditional Probability

In general, $A_1, A_2, \ldots, A_n$ are events in S then

$$P(A_1 \cap A_2 \cap \cdots \cap A_n)$$

$$= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Example: 1

Consider randomly selecting a student at a certain university, and let A denote the event that the selected individual has a Visa credit card and B be the analogous event for a Master Card. Suppose that $P(A) = 0.5$, $P(B) = 0.5$ and $P(A \cap B) = 0.25$. Calculate and interpret each of the following probabilities

a. $P\left({}^{B}/_{A}\right)$

b. $P\left({}^{B'}/_{A}\right)$

c. $P\left({}^{A'}/_{B}\right)$

d. Given that the selected individual has at least one card, what is the probability that he or she has a Visa card?

Solution:

let A denote the event that the selected individual has a Visa credit card and B be the analogous event for a Master Card

$P(A) = 0.5$, $P(B) = 0.5$ and $P(A \cap B) = 0.25$.

a. $P({}^{B}/_{A}) = \frac{P(A \cap B)}{P(A)} = \frac{0.25}{0.5}$

b. $P\left({}^{B'}/_{A}\right) = \frac{P\left(A \cap B'\right)}{P(A)} = \frac{P(A) - P(A \cap B)}{P(A)} = \frac{0.5 - 0.25}{0.5} = 0.5$

c. $P\left({}^{A'}/_{B}\right) = \frac{P\left(A' \cap B\right)}{P(B)} = \frac{P(B)-P(A \cap B)}{P(B)} = \frac{0.5-0.25}{0.5} = 0.5$

d. $P\left({}^{A}/_{A \cup B}\right) = \frac{P[(A \cap (A \cup B))]}{P(A \cup B)} = \frac{P[A]}{P(A)+P(B)-P(A \cap B)} = \frac{0.5}{0.5+0.5-0.25} =$ 0.67

Example: 2

Following table contains the different age group peoples who have defaulted and not defaulted on Loans.

		Age			
		Young	Middle-Aged	Senior citizens	Total
Loan Default	No	10503	27368	259	38130
	Yes	3586	4851	120	8557
	Total	14089	32219	379	46687

a) What is the probability that a person will not default on the loan given he/she is middle-aged?
b) What is the probability that a person is middle-aged given he/she has not defaulted on the loan?

Solution:

Let A be denote the event that a person will not default on the loan and B be the event that he/she is middle-aged.

$$P(A) = \frac{38130}{46687} = 0.82, P(B) = \frac{32219}{46687} = 0.69, \mathrm{P}(\mathrm{A} \cap B) = 0.59$$

a) $P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{0.59}{0.69} = 0.85.$

b) $P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{0.59}{0.82} = 0.72$

Example:3

A certain shop repairs both audio and video components. Let A denote the event that the next component brought in for repair is an audio component, and let B be the event that the next component is a compact disc player (so the event B is contained in A). Suppose that $P(A) = 0.6$ and $P(B) = 0.05$. Then find $P(B/A)$.

Solution:

Let A denote the event that the next component brought in for repair is an audio component and

Let B be the event that the next component is a compact disc player

P(A) = 0.6 and P(B) = 0.05 and given that $B \subseteq A$

From sets operations $A \cap B = B$

then $P(A \cap B) = P(B) = 0.05$

$$P(^B/_A) = \frac{P(A \cap B)}{P(A)} = \frac{0.05}{0.6} = 0.0833$$

INDEPENDENT EVENTS

innovate achieve lead

An important result from the conditional probability:

If B has no effect on A, then, $P\ (A/B) = P(A)$

Also $P\ (B/A) = P(B)$

and we say the events are independent. (Statistically Independent)

i.e., The probability of A does not depend on B.

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

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

or $$\boxed{P(A \cap B) = P(A) \times P(B)}$$

Example: 1

Two sets of cards with a letter on each card as follows are placed into separate bags:

Bag 1:	I	L	J	A	U	
Bag 2:	L	R	H	E	C	A

Bala randomly picked one card from each bag. Find the probability that:

a) He picked the letters ‘J’ and ‘R’.
b) Both letters are ‘L’.
c) Both letters are vowels.

Solution:

a) Probability that she picked J and R $= \frac{1}{5} . \frac{1}{6} = \frac{1}{30}$

b) Probability that both letters are L $= \frac{1}{5} . \frac{1}{6} = \frac{1}{30}$

c) Probability that both letters are vowels$= \frac{2}{5} . \frac{2}{6} = \frac{2}{15}$

Example: 2

In a group of 300 all adults, 272 are right-handed, 3 adults are selected <u>with replacement</u>. Find

a) P(all 3 are right- handed)

b) P(all 3 left-handed)

c) P(At least one right-handed)

Solution:

P(one right-handed) = 272/300 =0.907

P(one left-handed) = (300-272)/300 = 28/300 = 0.093

a) P(3 right-handed) = (272/300) (272/300)(272/300)

$= (272/300)^3 = 0.745$

b) P(3 left-handed) = (28/300)(28/300)(28/300) = $(28/300)^3$

= 0.0008

c) P(At least one right-handed) = 1 – P(All 3 left-handed)

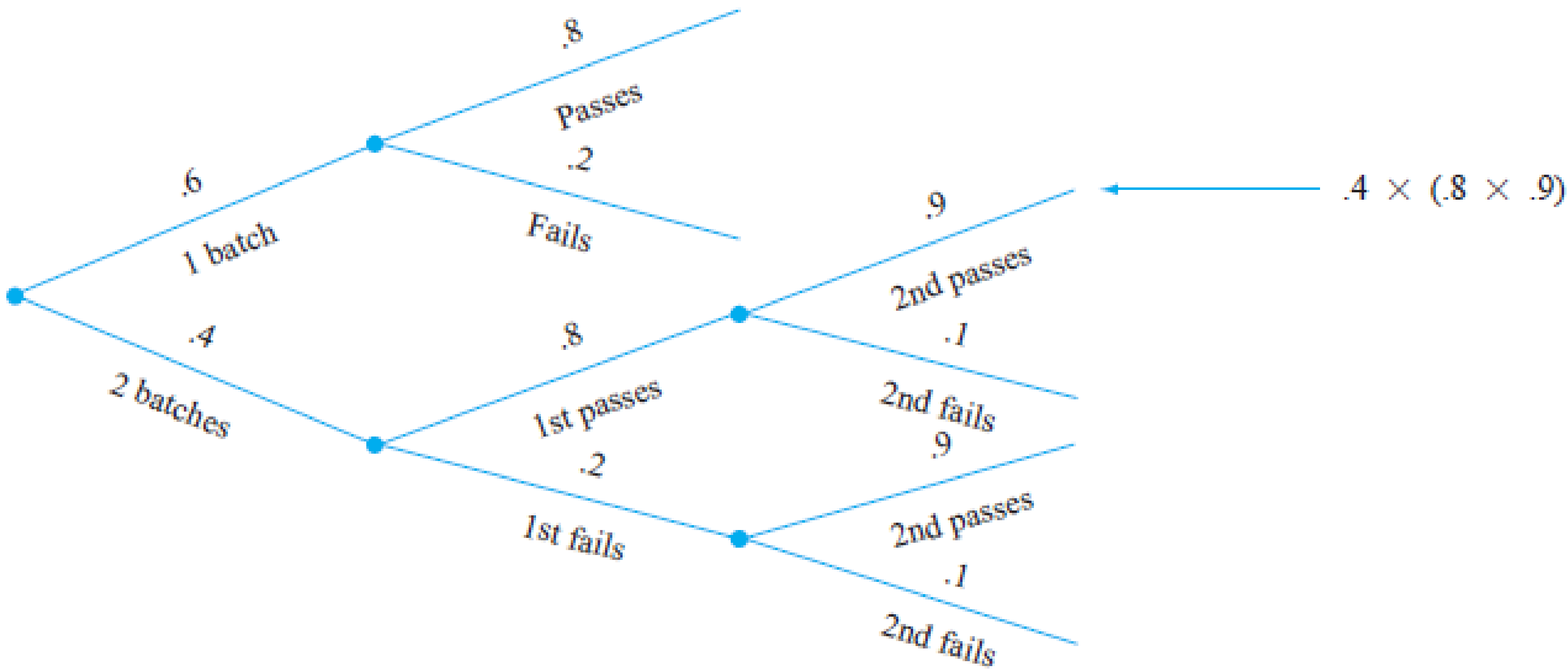
= 1 - 0.0008

= 0.9992

Example: 3

Each day, Monday through Friday, a batch of components sent by a first supplier arrives at a certain inspection facility. Two days a week, a batch also arrives from a second supplier. Eighty percent of all supplier 1's batches pass inspection, and 90% of supplier 2's do likewise. What is the probability that, on a randomly selected day, two batches pass inspection? We will answer this assuming that on days when two batches are tested.

Solution:



$$
\begin{aligned}
P(\text{two pass}) &= P(\text{two received} \cap \text{both pass}) \\
&= P(\text{both pass} \mid \text{two received}) \cdot P(\text{two received}) \\
&= [(.8)(.9)](.4) = .288
\end{aligned}
$$

Example: 4

A data science team is working on a model to predict whether an email is spam or not. They are using two independent features: The presence of the word "offer“ (Feature A) and the presence of a suspicious link (Feature B).

The probability that an email contains the word "offer“ (Feature A) is 0.6.

The probability that an email contains a suspicious link (Feature B) is 0.4.

a) What is the probability that an email contains both the word "offer" and a suspicious link?
b) What is the probability that an email contains either the word "offer" or a suspicious link or both?
c) What is the probability that an email contains neither the word "offer" nor a suspicious link?

Solution:

a) Probability of both features being present (A and B):

Since the events are independent $P(A \cap B) = P(A).P(B)$

$$\therefore P(A \cap B) = (0.6)(0.4) = 0.24$$

b) Probability of at least one feature being present (A or B or both):

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

$$P(A \cup B) = 0.6 + 0.4 - 0.24 = 0.76$$

c) Probability of neither feature being present (neither A nor B):

$\overline{P(A \cup B)} = 1 - P(A \cup B) = 1 - 0.76 = 0.24$

Miscellaneous examples

1. In a jar with 5 red, 6 blue and 2 white marbles. Two marbles are selected, find the probability that both are red if:

a) If two marbles are selected with replacement.

b) If two marbles are selected without replacement.

2. In batch of 6400 light bulbs, 80 are defective. If 12 light bulbs are selected from the batch without replacements, find probability that all are good.

3. In a city with two airports, 100 flights were surveyed. 20 of those flights departed late.

- 45 flights in the survey departed from airport A; 9 of those flights departed late.
- 55 flights in the survey departed from airport B; 11 flights departed late.
- Are the events "depart from airport A" and "departed late" independent?

Contd..

4. If A and B are two events with P(A)= 1/3, P(B)= 1/2 and $P(A \cap B) = 1/4$ Find

$$P\left(\frac{A}{B}\right), P\left(\frac{B}{A}\right), \mathrm{P}\left(\frac{\overline{\mathrm{A}}}{\overline{B}}\right), \mathrm{P}\left(\frac{\overline{\mathrm{B}}}{\overline{A}}\right), \mathrm{P}\left(\frac{\mathrm{A}}{\overline{B}}\right)$$

5. Three students A,B,C write an examination. Their chances of passing are $\frac{1}{4}, \frac{1}{2}$ and $\frac{1}{3}$ respectively. Find the probability that

(i) all of them pass,

(ii) at least one of them passes,

(iii) at least two of them pass.

Total Probability

innovate achieve lead

The Law of Total Probability:

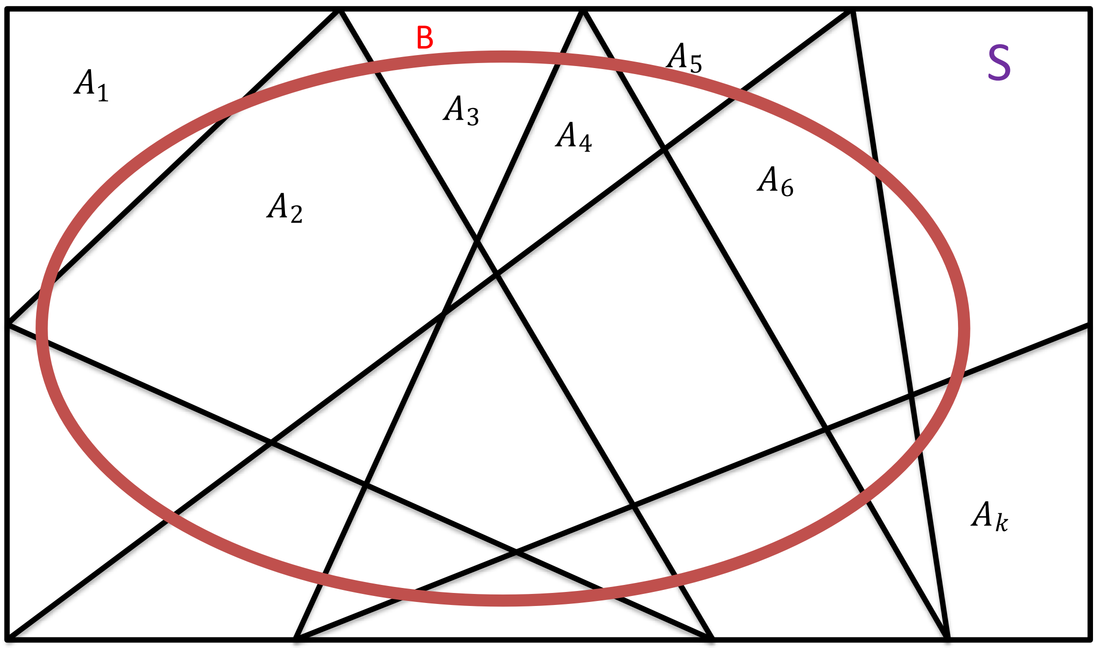
Let $A_1, A_2, \dots A_k$ be mutually exclusive and exhaustive events.

Then for any other event B,

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \dots + P(B|A_k)P(A_k)$$

$$P(B) = \sum_{i=1}^{k} P(B|A_i)P(A_i)$$

$S = A_1 \cup A_2 \cup \ldots \cup A_k$ and $A_1 \cap A_2 \cap \ldots \cap A_k = \emptyset$



$$\therefore \boldsymbol{B} = \boldsymbol{B} \cap \boldsymbol{S} = \boldsymbol{B} \cap \{\boldsymbol{A}_1 \cup \boldsymbol{A}_2 \cup \boldsymbol{A}_3, \ . \quad . \quad . \cup \boldsymbol{A}_k\}$$

Proof:

We have $S = \{A_1 \cup A_2 \cup A_3 \dots \cup A_k\}$ and $A \subset S$

$$\therefore B = B \cap S = B \cap \{A_1 \cup A_2 \cup A_3, \dots \cup A_k\}$$

Using distributive law in the R.H.S., we get

$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_k)$

Since $B \cap A_i$ $(i = 1\ to\ n)$ are mutually exclusive. So by addition rule of probability

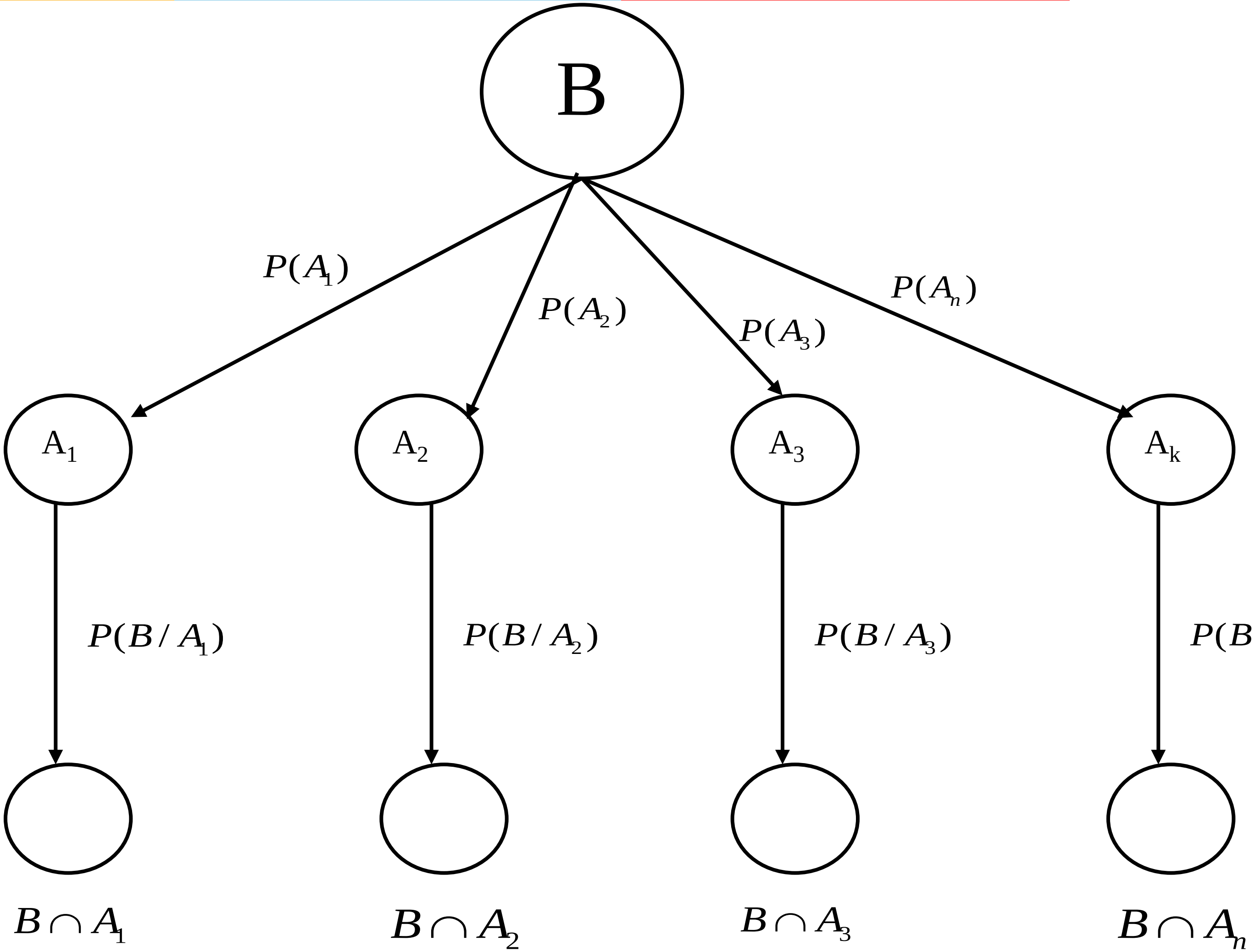
$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k)$$

$$i.e\ P(B) = \sum_{i=1}^{k} P(B \cap A_i)$$

But $P(B \cap A_i) = P(A_i).P(B|A_i)$.

$$\therefore P(B) = \sum_{i=1}^{k} P(A_i).P(B|A_i)$$

The Theorem of Total Probability
(tree diagram)



Example: 1

At a certain gas station, 40% of the customers use regular gas (A_1), 35% use plus gas (A_2), and 25% use premium (A_3). Of those customers using regular gas, only 30% fill their tanks (event B). Of those customers using plus, 60% fill their tanks, whereas of those using premium, 50% fill their tanks.

What is the probability that the next customer fills the tank?

Solution:

Probabilities of customers using regular gas

$$\boldsymbol{P(A_1) = 40\% = 0.4}$$

Probabilities of customers using plus gas

$$\boldsymbol{P(A_2) = 35\% = 0.35}$$

Probabilities of customers using premium gas

$$\boldsymbol{P(A_3) = 25\% = 0.25}$$

Also given with conditional probabilities of full gas tank

$$P\left({}^{B}/_{A_1}\right) = 30\% = 0.3$$

$$P\left({}^{B}/_{A_2}\right) = 60\% = 0.6$$

$$P\left({}^{B}/_{A_3}\right) = 50\% = 0.5$$

The probability of next customer filling the tank is

$$P(B) =$$

$$= P(A_1)\,P\left({}^{B}/_{A_1}\right) + P(A_2)P\left({}^{B}/_{A_2}\right) + P(A_3)P\left({}^{B}/_{A_3}\right)$$

$$= (0.4 \times 0.3) + (0.35 \times 0.6) + (0.25 \times 0.5)$$

$$= 0.455$$

Example:2

A simple binary communication channel carries messages by using only two signals, say 0 and 1. For a given binary channel, 40% of the time a 1 is transmitted; the probability that a transmitted 0 is correctly received is 0.90, and the probability that a transmitted 1 is correctly received is 0.95.

Determine

(a) The probability of a 1 being received, and

(b) Given a 1 is received, the probability that 1 was transmitted.

Solution:

Let A = event that 1 is transmitted,

$\bar{A}$ = event that 0 is transmitted ,

B = event that 1 is received

$\bar{B}$ = event that 0 is received.

The information in the given problem can be stated as:

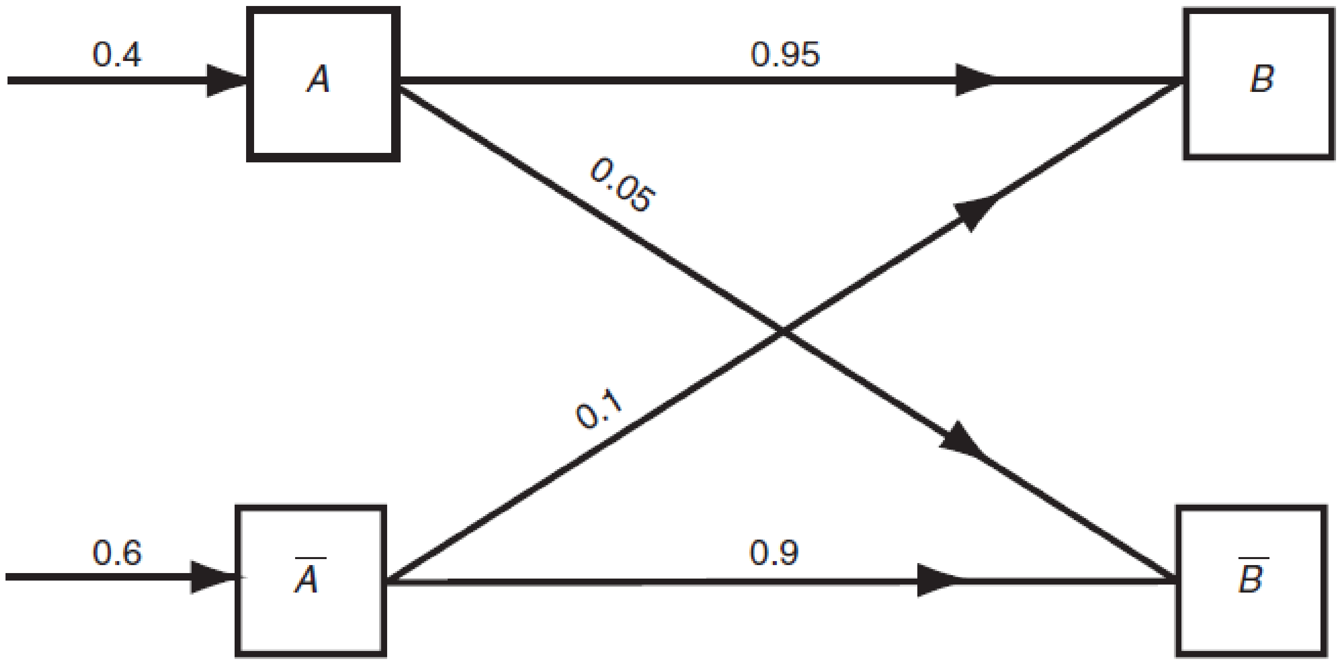
$P(A) = 0.4$ $P(\bar{A})$=0.6 , $P(B|A) = 0.95, P(\bar{B}|\mathrm{A}) = 0.05$

$P(\bar{B}|\bar{A})$=0.90, $P(B|\bar{A})$=0.10.

To find:

a) $P(B)$

b) $P(A|B)$



Probabilities associated with a binary channel

a) Since A and $\bar{A}$ are mutually exclusive and exhaustive, it follows from the theorem of total probability,

$$P(B) = P(B|A)P(A) + P(B|\overline{A})P(\overline{A}) = 0.95(0.4) + 0.1(0.6) = 0.44.$$

b) P(A/B) can be found as: P(A ∩B)/P(B) = P(A)P(B|A)/P(B)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{0.95(0.4)}{0.44} = 0.863.$$

Example: 3

A data science team is developing a predictive model to determine whether a user will click on an online advertisement. The likelihood of a user clicking on an ad depends on the type of device they are using. The team has categorized the devices into three types: desktop, tablet, and mobile.

The probabilities of a user using each type of device are as follows:

- The probability that a user uses a desktop is 0.5.
- The probability that a user uses a tablet is 0.2.
- he probability that a user uses a mobile device is 0.3.

Also,

- The probability of clicking on an ad given that the user is on a desktop is 0.04.
- The probability of clicking on an ad given that the user is on a tablet is 0.06.
- The probability of clicking on an ad given that the user is on a mobile device is 0.1.

What is the overall probability that a user will click on an ad?

Solution

The total probability that a user clicks on an ad can be found using the law of total probability, which states:

$$P(C) = P(C|D)P(D) + P(C|T)P(T) + P(C|M)P(M).$$

Where:

- $P(C)$-the probability of clicking on an ad.
- $P(C|D)$, $P(C|T)$ and $P(C|M)$ are the conditional probabilities of clicking on an ad given the device type.
- $P(D), P(T)$ and $P(M)$ are the probabilities of using a desktop, tablet, and mobile device, respectively.

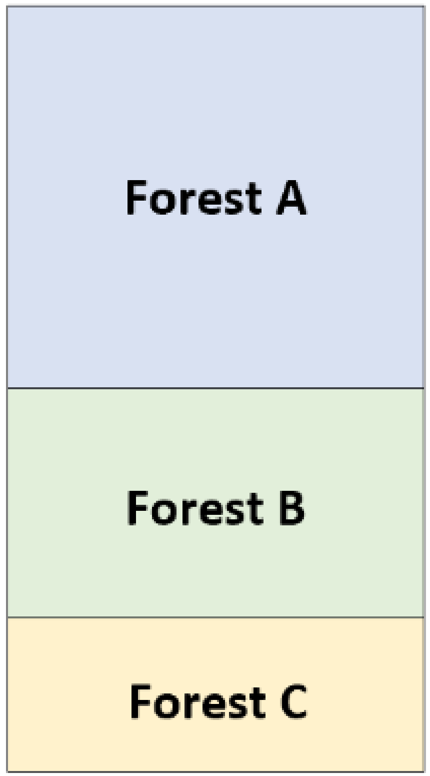
$$\therefore P(C) = (0.04 * 0.5) + (0.06 * 0.2) + (0.1 * 0.3) = 0.062$$

So, the overall probability that a user will click on an ad is 0.062.

Example: 4

Forest A occupies 50% of the total land in a certain park and 20% of the plants in this forest are poisonous. Forest B occupies 30% of the total land and 40% of the plants in it are poisonous. Forest C occupies the remaining 20% of the land and 70% of the plants in it are poisonous. If we randomly enter this park and pick a plant from the ground, what is the probability that it will be poisonous?

Given:



Occupies 50% of total land
20% of plants are poisonous

Occupies 30% of total land
40% of plants are poisonous

Occupies 20% of total land
70% of plants are poisonous

Solution

If we let P(*P*) = the probability of the plant being poisonous, and $P(B_i)$ be the probability that we've entered one of the three forests, then we can compute the probability of a randomly chosen plant being poisonous as:

$$P(P) = \Sigma P(P|B_i)*P(B_i)$$

$$P(P) = P(P|B_1)*P(B_1) + P(P|B_2)*P(B_2) + P(P|B_3)*P(B_3)$$

$$P(P) = (0.20)*(0.50) + (0.40)*(0.30) + (0.70)*(0.20)$$

$$P(P) = \mathbf{0.36}$$

Example: 5

A person has undertaken a mining job. The probabilities of completion of the job on time with and without rain are 0.42 and 0.90 respectively. If the probability that it will rain is 0.45, then determine the probability that the mining job will be completed on time.

Solution:

Let A be the event that the mining job will be completed on time and B be the event that it rains. We have,

$P(B) = 0.45$,

$P(\text{no rain}) = P(B') = 1 - P(B) = 1 - 0.45 = 0.55$

By multiplication law of probability,

$P(A|B) = 0.42$

$P(A|B') = 0.90$

Since, events B and B′ form partitions of the sample space S, by total probability theorem, we have

$$P(A) = P(B)\,P(A|B) + P(B')\,P(A|B')$$
$$= 0.45 \times 0.42 + 0.55 \times 0.9$$
$$= 0.189 + 0.495 = 0.684$$

So, the probability that the job will be completed on time is 0.684.

Glossary

- **Mutually Exclusive Events:** Events that cannot occur simultaneously. In the context of the total law of probability, the sum of the probabilities of all mutually exclusive events equals 1.
- **Condition Probability $\boldsymbol{P(A|B)}$**: The probability of event A occurring given that event B has occurred. It quantifies the dependence of A on B.
- **Independent Events:** Two events are independent if the occurrence of one does not affect the probability of the other. For independent events A and B, $P(A \cap B) = P(A).P(B)$.
- **Total Law of Probability:** A fundamental rule in probability theory that expresses the total probability of an event as the sum of the probabilities of that event occurring given different mutually exclusive scenarios or conditions. It is particularly useful for breaking down complex probability problems into simpler parts.

  $$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \ldots + P(B|A_k)P(A_k)$$

Important Formulae

Counting Rule for Combinations

$$C_n^N = \binom{N}{n} = \frac{N!}{n!(N-n)!}$$

Counting Rule for Permutations

$$P_n^N = n!\binom{N}{n} = \frac{N!}{(N-n)!}$$

Computing Probability Using the Complement

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

Addition Law

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

Important Formulae

Conditional Probability

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

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

Multiplication Law

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

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

Multiplication Law for Independent Events

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

Practice problems

Example 1.

An office has 4 secretaries handling respectively 20%, 60%,15% and 5% of the files of all government reports. The probability that they misfile such reports are respectively 0.05, 0.1, 0.1 and 0.05. Find the probability that the misfiled report can be blamed on the first secretary.

Ans: 0.1143

Example 2.

In a class 70% are boys and 30% are girls. 5% of boys and 3% of girls are irregular to the classes. What is the probability of a student selected at random is irregular to the classes and what is the probability that the irregular student is a girl?

Ans: P(Irregular Student) = 0.044, P(Irregular and Girl) = 0.009

Example 3.

Three machines A, B and C produce respectively 60%, 30% and 10% of the total number of items of a factory. The percentage of defective outputs of these machines are 2%, 3% and 4%. An item is selected at random and is found to be defective.

(i) Find the probability that the item was not defective and produced by machine C?

Ans: 0.096

(ii) What is the probability that the item was produced by machine C or B? Ans: 0.4

Example: 4

A card is randomly drawn from an incomplete deck of cards from which the ace of diamonds is missing.

1. What is the probability that the card is “clubs”? (Ans. 13/51)
2. What is the probability that the card is a “queen”? (Ans. 4/51)
3. Are the events “clubs” and “queen” independent (Ans. No)

Example: 5

Three companies produce the same tool and supply it to the market. Company A produces 30% of the tools for the market and the remaining 70% of the tools are produced in companies B and C. 98% of the tools produced in company A, 95% of the tools produced in company B and 97% of the tools produced in company C are not defective. What percent of tools should be produced by companies B and C so that a tool picked at random in the market will have a probability of being non defective equal to 96%?

Ans: 0.164

Example: 6

A sample of 500 respondents was selected in a large metropolitan area to study consumer behavior, with the following results:

ENJOYS SHOPPING FOR CLOTHING	GENDER		
	Male	Female	Total
Yes	136	224	360
No	104	36	140
Total	240	260	500

a. Suppose the respondent chosen is a female. What is the probability that she does not enjoy shopping for clothing? (Ans: 0.1384)

b. Suppose the respondent chosen enjoys shopping for clothing. What is the probability that the individual is a male?(Ans: 0.56)

Example: 7

A standard deck of cards is being used to play a game. There are four suits (hearts, diamonds, clubs, and spades), each having 13 faces (ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, and king), making a total of 52 cards. This complete deck is thoroughly mixed, and you will receive the first 2 cards from the deck without replacement.

a. What is the probability that both cards are queens? (Ans: 0.0045)

b. What is the probability that the first card is a 10 and the second card is a 5 or 6? (Ans: 0.012)

c. If you were sampling with replacement, what would be the answer in (a)? (Ans: 0.0059)

IMP Note to Self

Thank you