MAT 3103: Computational Statistics and Probability

Chapter 3: Probability

Probability:

A probability is a number that reflects the chance or likelihood that a particular event will occur. Probabilities can be expressed as proportions that range from 0 to 1, and they can also be expressed as percentages ranging from 0% to 100%. A probability of 0 indicates that there is no chance that a particular event will occur, whereas a probability of 1 indicates that an event is certain to occur. The probability that the sun will rise from the north tomorrow is 0, whereas the probability that an individual currently alive will die one day is 1. A probability of 0.45 (45%) indicates that there are 45 chances out of 100 for the event to occur.

Application of probability for engineering:

Application areas: Modeling of text and web data, network traffic modeling, probabilistic analysis of algorithms and graphs, reliability modeling, simulation algorithms, data mining, and speech recognition.

The mathematical methods that we will use to analyze these applications will include basic principles of probability such as Bayes rule, conditional probability, random variables, expectation, and Markov chains. These are only the few examples you can actually develop logical thinking for series and sequence programming. We use probability to measure the success or failure of something. Same goes for the use of probability in programming. Programmers use probability to measure the success of the program before running it.

In modern computer science, software engineering, and other fields, the need arises to make decisions under uncertainty. Presenting probability and statistical methods, simulation techniques, and modeling tools, Probability and Statistics for Computer Scientists helps students solve problems and make optimal decisions in uncertain conditions, select stochastic models, compute probabilities and forecasts, and evaluate performance of computer systems and networks. If you combine probability and statistics with a computer science curriculum, you get a data science curriculum! If you have interest in becoming a data scientist (or any related profession), then that should be motivation for you.

Probability Theory is one of the most important courses in Electrical Engineering. In most Universities you are taught an Introduction to Probability Theory and Statistics and then in other classes you are learning about more specific areas depending on the needs of each class. For instance, if you are going to choose a) Control Systems, b) Signals, c) Information Theory, d) Communications, then you are certainly going to need probabilities in most of your classes. But that is not the case if you are going to choose Energy or Motors.

Probability is particularly relevant in the area of quality engineering and being able to design reliable circuits where you have to take into account the tolerance of various components, and the predicted lifetimes. What is the mean-time-to-failure (MTTF) of your system? Where can you get by with a 20% tolerance capacitor instead of a more expensive 5% one? Communications theory and error correction are all about probability and statistics, where the maximum amount of information is jammed into the smallest amount of bandwidth, while dealing with possibly very noisy channels.

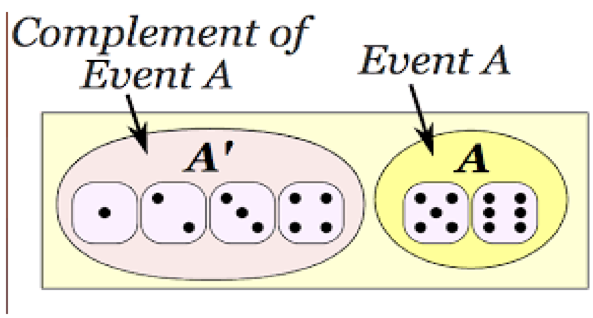
Some relevant terms which are needed to understand probability:

- **Experiment:** The work or activity that generates the results to be studied.
- **Random experiment:** It is an experiment whose outcomes cannot be predicted with certainty in advance, and these outcomes depend on chance.
- **Outcome:** The result of an experiment.
- **Sample space:** Set of all possible outcomes. It is usually denoted by the symbol S.

The above-mentioned terms can be better understood from the following set of examples:

Random experiment	**Outcome**	**Sample space**
Tossing a coin (once)	Head (H), tail (T)	S = {H, T}
Tossing a coin (twice)	H, T	S = {HH, HT, TH, TT}

- **Event:** Any subset of a sample space is an event. It is denoted by capital letters, e.g., A/B/C. In an experiment of rolling a die, an event can be of getting any of the numbers from 1 to 6 on its uppermost face. E.g., A = getting number 4 when a die is rolled. A is an event.
- **Mutually exclusive events:** The events are said to be mutually exclusive when they do not occur simultaneously. If a student is in class, he/she cannot be at shopping mall in the same time. If a ball is white, it cannot be red.
- **Equally likely events:** Events are said to be equally likely, when there is equal chance of occurring. In rolling a die, all six faces are equally likely to come.
- **Exhaustive outcomes:** All possible outcomes of a random experiment are exhaustive outcomes. In the sample space (S) given above the outcomes HH, HT, TH and TT are exhaustive outcomes.
- **Favorable outcomes:** Number of outcomes in favor of an event is known as favorable outcomes. It is denoted by m ($\leq$ n).
- **Complementary events:** The complement of an event A, denoted A' or A^c, is the event not A. If the probability of an event, A, is P(A), then the probability that the event would not occur (also called the complementary event) is 1 – P(A).

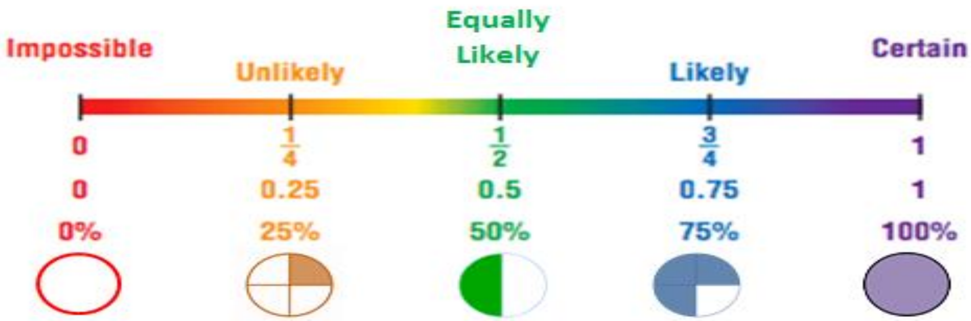


- **Independent events:** Two or more events are said to be independent when the occurrence of one trial does not affect the other. If a coin is tossed one by one, then in a trial the head or tail may come which never describes anything what event will come in second trial. So, the second trial is completely independent to that of the first trial.

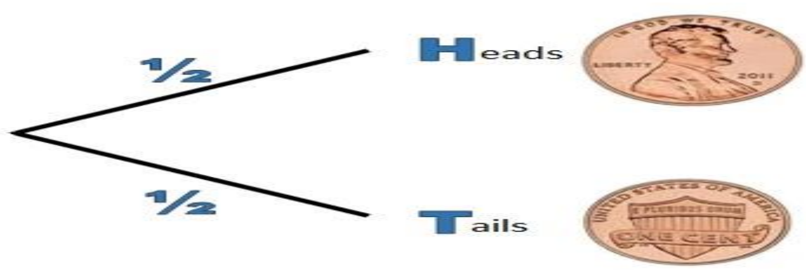
Probability: If a random experiment shows n exhaustive, mutually exclusive and equally likely outcomes and if m ($\leq$ n) outcomes are in favor of an event A, then the probability of an event A is measured by:

$$P(A) = \frac{\text{Number of favourable outcomes}}{\text{Total number of possible outcomes}} = \frac{n(A)}{n(S)} = \frac{m}{n}.$$

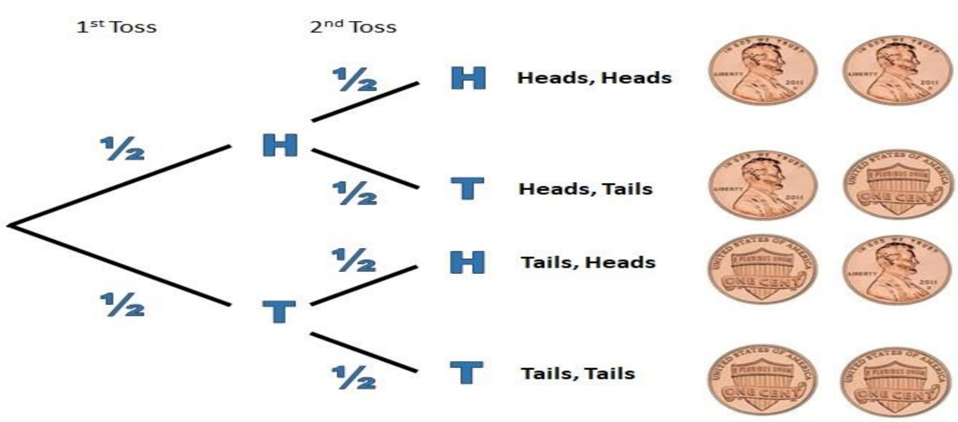
Interpretation rule of probability:



Tree diagram: A diagram that represents probabilities. If a fair coin is tossed once, then



If a fair coin is tossed twice, then



Conditional probability: The conditional probability of an event A in relationship to an event B is the probability that event A occurs given that event B has already occurred, denoted as $P(A/B)$, read as the probability of A given B.

$$P(A/B) = \frac{P(A\cap B)}{P(B)}, P(B) \neq 0.$$

Additive law of probability: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Multiplicative law of probability: P(AB) = P(A) P(B), where A and B are independent events.

With replacement: Selecting with replacement is exactly what it sounds like – we are picking something out of a bag (bucket, drawer, group, etc.), putting it back in there (i.e., the replacement), and picking another one of those something out.

Without replacement: In some experiments, the sample space may change for the different events. For example, a marble may be taken from a bag with 20 marbles and then a second marble is taken without replacing the first marble. The sample space for the second event is then 19 marbles instead of 20 marbles. This is called probability without replacement.

Example 3.1: A fair coin is tossed once. What is the probability that a head will be shown?

Solution: A fair coin means, the probability of observing head (H) equal to the probability of observing tail (T). A coin is tossed once, so the sample space is S = {H; T}. Let us now define the event that head will be shown. Thus, the event set A = {H} and

$$P(\text{A}) = \frac{m}{n} = \frac{1}{2}.$$

Example 3.2: Two fair coins are tossed once. What is the probability that (a) both coins will show head (b) at least one coin will show tail (c) at most (or at best) one coin will show head, and (d) none of the coins will show head?

Solution: Sample space for two coins is S = {HH; HT; TH; TT} and the total cases, n = 4.

(a) Both the coin show heads (H), that is, the event set A = {HH}. Thus m = 1 and $P(\text{A}) = \frac{m}{n} = \frac{1}{4}$.

(b) At least one coin will show tail (T), that is, the event set B = {HT; TH; TT} and m = 3. Thus

$$P(\text{B}) = \frac{m}{n} = \frac{3}{4}.$$

(c) At best one coin will show head, that is, C = {HT; TH; TT} and $P(\text{C}) = \frac{m}{n} = \frac{3}{4}$.

(d) None of the coins will show head, that is, the event set D = {TT} and P(D) = $\frac{m}{n} = \frac{1}{4}$.

Example 3.3: Three fair coins are tossed once. What is the probability that (a) at least two coins will show head (b) first or third coin will show head (c) first coin shows head given that third coin shows head?

Solution: Sample space for three coins is

<table>
<tr><td rowspan="2">First coin</td><td colspan="4">Second coin</td><td rowspan="4"></td></tr>
<tr><td colspan="2">H</td><td colspan="2">T</td></tr>
<tr><td>H</td><td colspan="2">TT</td><td colspan="2">HT</td></tr>
<tr><td>T</td><td colspan="2">TH</td><td colspan="2">TT</td></tr>
<tr><td rowspan="2">Third coin</td><td colspan="4">First two coins</td><td>Sample space</td></tr>
<tr><td>HH</td><td>HT</td><td>TH</td><td>TT</td><td rowspan="3">S = {HHH; HHT; HTH; HTT; THH; THT; TTH; TTT}</td></tr>
<tr><td>H</td><td>HHH</td><td>HHT</td><td>HTH</td><td>HTT</td></tr>
<tr><td>T</td><td>THH</td><td>THT</td><td>TTH</td><td>TTT</td></tr>
</table>

The all possible outcomes, n = 8.

(a) Let A be an event that at least two coins will show head. So, A = {HHH; HHT; HTH; THH},

$$P(\text{A}) = \frac{m}{n} = \frac{4}{8} = \frac{1}{2}.$$

Now, the first coin will show head, that is, B = {HHH; HHT; HTH; HTT} and P (B) = $\frac{4}{8}$,

the third coin will show head, that is, C = {TTH; THH; HTH; HHH} and P (C) = $\frac{4}{8}$,

$$\text{So, } P(\text{B} \cap \text{C}) = \frac{2}{8}.$$

(b) $P(\text{B} \cup \text{C}) = P(\text{B}) + P(\text{C}) - P(\text{B} \cap \text{C}) = \frac{4}{8} + \frac{4}{8} - \frac{2}{8} = \frac{6}{8}$.

(c) $P(B|C) = \frac{\text{P(B} \cap \text{C)}}{\text{P(C)}} = \frac{\frac{2}{8}}{\frac{4}{8}} = \frac{2}{4}$.

Problem 3.4: A die is rolled. Find the probability that an even number is obtained.

Solution: Let us first write the sample space: S = {1, 2, 3, 4, 5, 6}.

Let, E = an even number is obtained. So, E = {2, 4, 6}.

Then, $P(\text{E}) = \frac{\text{n(E)}}{\text{n(S)}} = \frac{m}{n} = \frac{3}{6}$.

Problem 3.5: Two bits are produced one by one using an electronic device. The device is such that it produces Fine (F) bit 50% times and Noisy (N) bits 50% times. Find the probability that (a) both bits are fine, (b) one bit is fine, and (c) at least one bit is fine.

Solution: Let us first construct the sample space, S.

First bit	Second bit		Sample space
	F	N	S = {FF, FN, NF, NN}
F	FF	FN	
N	NF	NN	

(a) Let A = both bits are fine = {FF}. Then, $P(\mathrm{A}) = \frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{1}{4}$.

(b) Let B = one is fine bit = {FN, NF}. Then, $P(\mathrm{B}) = \frac{\mathrm{n(B)}}{\mathrm{n(S)}} = \frac{2}{4}$.

(c) Let C = at least one bit is fine = {FF, FN, NF}. Then, $P(\mathrm{C}) = \frac{\mathrm{n(C)}}{\mathrm{n(S)}} = \frac{3}{4}$.

Problem 3.6: There are 100 vehicles at a car park. 60 of them are cars, 30 are vans and the remaining are lorries. If every vehicle is equally likely to leave, find the probability of: (a) van leaving first, and (b) lorry leaving first.

Solution: (a) Let A be the event of a van leaving first. Then, $P(\mathrm{A}) = \frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{30}{100}$.

(b) Let B be the event of a lorry leaving first. Then, $P(\mathrm{B}) = \frac{\mathrm{n(B)}}{\mathrm{n(S)}} = \frac{10}{100}$.

Problem 3.7: In a box there are 30 bulbs. The bulbs are identified by identity number 1 to 30. One bulb is selected randomly. Find the probability that the selected bulb has the identity number (a) either multiple of 3 or 5, and (b) even under the condition that it is multiple of 3.

Solution: (a) Let A = multiple of 3 = {3, 6, 9, 12, 15, 18, 21, 24, 27, 30}. So, $P(\mathrm{A}) = \frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{10}{30}$.

B = multiple of 5 = {5, 10, 15, 20, 25, 30}. Then, $P(\mathrm{B}) = \frac{\mathrm{n(B)}}{\mathrm{n(S)}} = \frac{6}{30}$.

Now, A∩B = {15, 30}. Then, $P(\mathrm{A\cap B}) = \frac{\mathrm{n(A\cap B)}}{\mathrm{n(S)}} = \frac{2}{30}$.

So, $P(\mathrm{A\cup B}) = P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{A\cap B}) = \frac{10}{30} + \frac{6}{30} - \frac{2}{30} = \frac{14}{30} = \frac{7}{15}$.

(b) Let, D = even no. = {2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30}.

Now, D∩A = {6, 12, 18, 24, 30}. Then, $P(\mathrm{D\cap A}) = \frac{\mathrm{n(D\cap A)}}{\mathrm{n(S)}} = \frac{5}{30}$.

So, $P\,(\mathrm{D}|\mathrm{A}) = \frac{P(\mathrm{D\cap A})}{P(\mathrm{A})} = \frac{5/30}{10/30} = \frac{5}{10}$.

Problem 3.8: The probability that Lima passes mathematics is $\frac{7}{8}$ and the probability that she passes statistics is $\frac{3}{4}$. If the probability of passing both the courses is $\frac{5}{6}$, what is the probability that Lima will pass at least one of the courses?

Solution: Let, A be the event of passing mathematics and B be the event of passing statistics.

So, $P(A\cup B) = P(A) + P(B) - P(A\cap B) = \frac{7}{8} + \frac{3}{4} - \frac{5}{6} = 0.79$.

Problem 3.9: Three consecutive phone calls are monitored. The calls may be either Voice call (V) or Data call (D), where P(V) = P(D) = 50%. Voice call means someone is speaking and Data call means it carries a signal. Find the probability that out of the three calls, there will be (a) no data calls, (b) at least one voice call, and (c) at best one voice call.

Solution: Let us first construct the sample space, S.

First call	Second call	
	V	D
V	VV	VD
D	DV	DD

Third call	First two calls				Sample space
	VV	VD	DV	DD	S = {VVV, VVD, VDD, VDV, DDV, DVV, DVD, DDD}
V	VVV	VVD	VDV	VDD	
D	DVV	DVD	DDV	DDD	

(a) P (no data calls) = P(VVV) = $\frac{1}{8}$.

(b) P (at least one voice call) = P (VVV, VVD, VDD, VDV, DDV, DVV, DVD) = $\frac{7}{8}$.

(c) P (at best one voice call) = P (VDD, DDV, DVD, DDD) = $\frac{4}{8}$.

Problem 3.10: Three consecutive phone calls are monitored. The calls may be either Voice call (V) or Data call (D), if P(V) = 10% and P(D) = 90%, Find the probability that out of the three calls, there will be (a) no data calls, (b) at least one voice call, and (c) at best one voice call.

Solution: Given, P(V) = 10% = 0.1 and P(D) = 90% = 0.9.

(a) P (no data calls) = P (VVV) = P(V) P(V) P(V) = (0.1) (0.1) (0.1) = 0.001.

(b) P (at least one voice call) = 1 − P (no voice call) = 1 − P (DDD)

= 1 − P(D) P(D) P(D) = 1 − (0.9) (0.9) (0.9) = 1 – 0.729 = 0.271.

(c) P (at best one voice call) = P (VDD, DDV, DVD, DDD)

= P (VDD) + P (DDV) + P (DVD) + P (DDD)

= P(V) P(D) P(D) + P(D) P(D) P(V) + P(D) P(V) P(D) + P(D) P(D) P(D)

= (0.1) (0.9) (0.9) + (0.9) (0.9) (0.1) + (0.9) (0.1) (0.9) + (0.9) (0.9) (0.9)

= 0.972.

Problem 3.11: Eighty-five per cent e-mails sent from a cyber cafe reach to the destination properly. Once 3 mails are checked randomly, Find the probability that **(**a**)** all 3 reach properly, **(**b**)** two reach properly, **(**c**)** at least one reaches properly and **(**d**)** at best two reach properly.

Solution: Let, R = reach properly the e-mail and N = not reach properly the e-mail.

Sent of 3 e-mails can occur in n = 2^3 = 8 ways. The sample space is-

S = {RRR, RRN, RNR, RNN, NRR, NRN, NNR, NNN}

Given, P(R) = 0.85 and P(N) = 1− 0.85 = 0.15. So, R and N are not equally likely.

a) Let, A: all 3 reach properly; where, A = {RRR}

P(A) = P(RRR) = 0.85×0.85×0.85 = 0.6141.

b) Let, B: two reach properly; where, B = {RRN, RNR, NRR}

P(B) = P(RRN)+ P(RNR)+ P(NRR)

= (0.85×0.85×0.15) + (0.85×0.15×0.85) + (0.15×0.85×0.85) = 0.3251.

c) Let, C: at least one reaches properly;

where, C = {RRR, RRN, RNR, RNN, NRR, NRN, NNR} $\bar{C}$ = {NNN}

P(C) =1− P ($\bar{C}$) = 1− P(NNN) = 1− (0.15×0.15×0.15) = 0.9966.

d) Let, D: at best two reach properly;

where, D = {RRN, RNR, RNN, NRR, NRN, NNR, NNN}, $\bar{D}$ = {RRR}

P(D) =1− P ($\bar{D}$) = 1−P(RRR) = 1− (0.85×0.85×0.85) = 0.3859.

Problem 3.12: There are 7 Vivo and 5 LG mobile sets in a box. Two sets are drawn at random. Find the probability that (a) both are Vivo sets, and (b) one set is Vivo, other is LG.

Solution: There are total 7 + 5 = 12 sets. Two sets can be selected in 12_{c_2} ways.

(a) Let A = both sets are Vivo. A can occur in 7_{c_2} ways. So, *P* (A) = $\frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{7_{c_2}}{12_{c_2}}$.

(b) Let B = one set is Vivo, other is LG. B can occur in $7_{c_1} \times 5_{c_1}$ ways.

So, *P* (B) = $\frac{\mathrm{n(B)}}{\mathrm{n(S)}} = \frac{7_{c_1} \times 5_{c_1}}{12_{c_2}}$.

Problem 3.13: There are 7 Vivo and 5 LG mobile sets in a box. Two sets are drawn one by one with replacement. Find the probability that (a) both are Vivo sets, and (b) one set is Vivo, other is LG.

Solution: Let us first construct the sample space, S.

First draw	Second draw		Sample space
	Vivo (V)	LG (L)	
Vivo (V)	VV	VL	S = {VV, VL, LV, LL}
LG (L)	LV	LL	

(a) *P* (both sets are Vivo) = *P*(VV) = *P*(V) × *P*(V) = $\frac{7}{12} \times \frac{7}{12} = \frac{49}{144}$.

(b) *P* (one set is Vivo, other is LG) = *P* (VL ∪ LV) = *P*(VL) + *P*(LV)

$= P(\mathrm{V}) \times P(\mathrm{L}) + P(\mathrm{L}) \times P(\mathrm{V}) = \frac{7}{12} \times \frac{5}{12} + \frac{5}{12} \times \frac{7}{12} = \frac{35}{144} + \frac{35}{144} = \frac{70}{144}$.

Problem 3.14: Find the probabilities specified in Problem 3.13 if sets are drawn without replacement.

Solution: (a) *P* (both sets are Vivo) = *P*(VV) = *P*(V) × *P*(V) = $\frac{7}{12} \times \frac{6}{11} = \frac{42}{132}$.

(b) *P* (one set is Vivo, other is LG) = *P* (VL ∪ LV) = *P*(VL) + *P*(LV)

$= P(\mathrm{V}) \times P(\mathrm{L}) + P(\mathrm{L}) \times P(\mathrm{V}) = \frac{7}{12} \times \frac{5}{11} + \frac{5}{12} \times \frac{7}{11} = \frac{35}{132} + \frac{35}{132} = \frac{70}{132}$.

Problem 3.15: In a box there are four balls numbered as 1, 2, 3, and 4. Two balls are drawn one by one with replacement. Find the probability that (a) sum of the numbers is 5 or first drawn ball has the number 3. (b) sum of the numbers is 5 given that second drawn ball bears the number 3.

Solution: Let us first construct the sample space.

First draw	Second draw			
	1	2	3	4
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)

Let A = sum of the numbers is 5 = {(1, 4), (2, 3), (3, 2), (4, 1)}. Then, $P(\mathrm{A}) = \frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{4}{16}$.

B = first ball has the number 3 = {(3, 1), (3, 2), (3, 3), (3, 4)}. Then, $P(\mathrm{B}) = \frac{\mathrm{n(B)}}{\mathrm{n(S)}} = \frac{4}{16}$.

C = second ball has the number 3 = {(1, 3), (2, 3), (3, 3), (4, 3)}. Then, $P(\mathrm{C}) = \frac{\mathrm{n(C)}}{\mathrm{n(S)}} = \frac{4}{16}$.

Now, A∩B = {(3, 2)}. So, $P(\mathrm{A\cap B}) = \frac{\mathrm{n(A\cap B)}}{\mathrm{n(S)}} = \frac{1}{16}$.

Again, A∩C = {(2, 3)}. So, $P(\mathrm{A\cap C}) = \frac{\mathrm{n(A\cap C)}}{\mathrm{n(S)}} = \frac{1}{16}$.

(a) $P(\mathrm{A\cup B}) = P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{A\cap B}) = \frac{4}{16} + \frac{4}{16} - \frac{1}{16} = \frac{7}{16}$.

(b) $P(\mathrm{A|C}) = \frac{P(\mathrm{A\cap C})}{P(\mathrm{C})} = \frac{\frac{1}{16}}{\frac{4}{16}} = \frac{1}{4}$.

Problem 3.16: Find the probabilities in Problem 3.15 if balls are drawn without replacement.

Solution: In case of without replacement, outcomes as (1, 1), (2, 2), (3, 3), and (4, 4) will not be counted. Hence, our total number of possible outcomes will be 16 – 4 = 12. Then,

A = {(1, 4), (2, 3), (3, 2), (4, 1)}. Then, $P(\mathrm{A}) = \frac{\mathrm{n(A)}}{\mathrm{n(S)}} = \frac{4}{12}$.

B = {(3, 1), (3, 2), (3, 4)}. Then, $P(\mathrm{B}) = \frac{\mathrm{n(B)}}{\mathrm{n(S)}} = \frac{3}{12}$.

C = {(1, 3), (2, 3), (4, 3)}. Then, $P(\mathrm{C}) = \frac{\mathrm{n(C)}}{\mathrm{n(S)}} = \frac{3}{12}$.

Now, A∩B = {(3, 2)}. So, $P(\mathrm{A\cap B}) = \frac{\mathrm{n(A\cap B)}}{\mathrm{n(S)}} = \frac{1}{12}$.

Again, A∩C = {(2, 3)}. So, $P(\mathrm{A\cap C}) = \frac{\mathrm{n(A\cap C)}}{\mathrm{n(S)}} = \frac{1}{12}$.

(a) $P(\mathrm{A\cup B}) = P(\mathrm{A}) + P(\mathrm{B}) - P(\mathrm{A\cap B}) = \frac{4}{12} + \frac{3}{12} - \frac{1}{12} = \frac{6}{12}$.

(b) $P(\mathrm{A|C}) = \frac{P(\mathrm{A\cap C})}{P(\mathrm{C})} = \frac{\frac{1}{12}}{\frac{3}{12}} = \frac{1}{3}$.

Problem 3.17: Suppose there are 30 students, out of which 12 are from EEE department and 18 are from CSE department. The CGPA of 8 EEE and 12 CSE students are found to be good. One student is selected randomly. Find the probability that the selected student is (a) from EEE dept. given that his CGPA is not good, and (b) from CSE dept. or his CGPA is good.

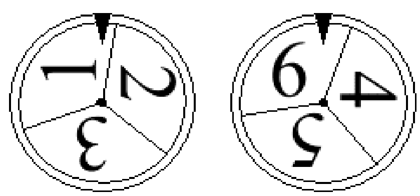
Solution: Let us first construct the sample space, S.

Department	CGPA		Total
	Good (G)	Not good ($\overline{\mathrm{G}}$)	
EEE (E)	8	4	12
CSE (C)	12	6	18
Total	20	10	30

(a) $P\,(\mathrm{E}\,|\,\overline{\mathrm{G}}) = \frac{P(\mathrm{E}\cap\overline{\mathrm{G}})}{P(\overline{\mathrm{G}})} = \frac{4/30}{10/30} = \frac{4}{10}$.

(b) $P\,(\mathrm{C}\cup\mathrm{G}) = P\,(\mathrm{C}) + P\,(\mathrm{G}) - P\,(\mathrm{C}\cap\mathrm{G}) = \frac{18}{30} + \frac{20}{30} - \frac{12}{30} = \frac{26}{30}$.

Problem 3.18: Julia spins 2 spinners. One is labeled as 1, 2 and 3. The other is labeled 4, 5 and 6.

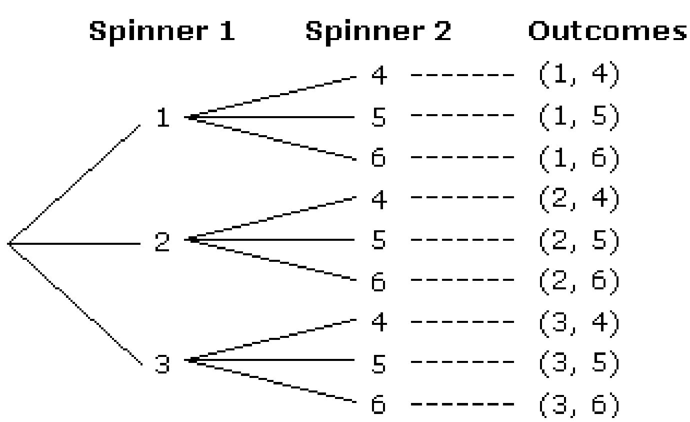


a) Draw a tree diagram for the experiment.

b) What is the probability that the spinners stop at “3” and “4”?

c) Find the probability that the spinners do not stop at “3” and “4”.

Solution: (a) Let us first draw the tree diagram.

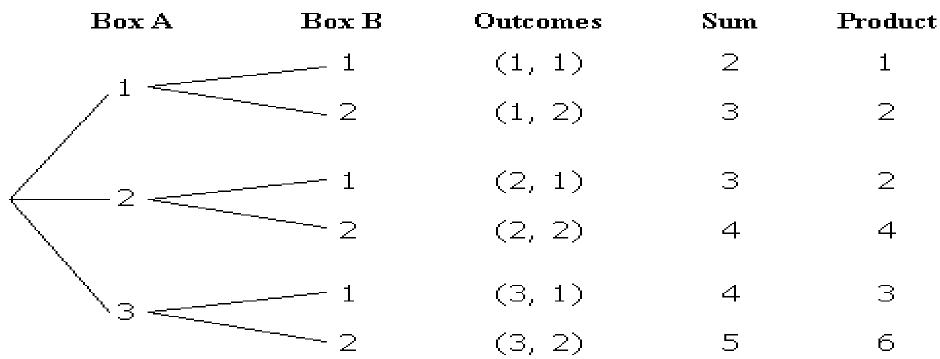


(b) P (the spinners stop at “3” and “4”) = P (3, 4) = $\frac{1}{9}$.

(c) P (the spinners do not stop at "3" and "4") = $1 - P(3, 4) = 1 - \frac{1}{9} = \frac{8}{9}$.

Problem 3.19: Box A contains three cards numbered as 1, 2 and 3. Box B contains 2 cards numbered as 1 and 2. One card is drawn randomly from each box. Draw a tree diagram to list all the possible outcomes. Find the probability that (a) sum of the numbers is 4, and (b) sum is equal to the product.

Solution: Let us first draw the tree diagram.



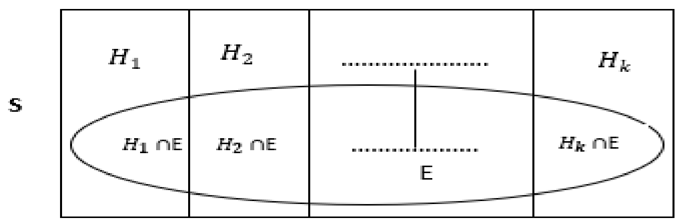
(a) P (sum of the numbers is 4) = $\frac{2}{6}$.

(b) P (sum is equal to the Product) = $\frac{1}{6}$.

Bayes' theorem and data science:

Are you planning to become a data scientist? If yes, you must know Bayes' Theorem. No data scientist can work without a complete understanding of conditional probability and Bayesian inference. So, today, we will discuss the same with the help of examples and applications. Bayes' Theorem is most widely used in Machine Learning as a classifier that makes use of Naive Bayes' Classifier. It has also emerged as an advanced algorithm for the development of Bayesian Neural Networks. The applications of Bayes' Theorem are everywhere in the field of Data Science.

Bayes' theorem: Let, S is the sample space having n equally likely outcomes. With some of outcomes let us define an event E. With some of outcomes of E we can define, separate mutually exclusive events $H_{1,} H_{2,} \ldots, H_k$. Then-



We have, E = $\mathrm{H_1} \cap$ E + $\mathrm{H_2} \cap$ E + … + $\mathrm{H_k} \cap$ E

$\therefore P$ (E) = P ($\mathrm{H_1} \cap$ E) + P ($\mathrm{H_2} \cap$ E) + … + P ($\mathrm{H_k} \cap$ E)

Again, P (E/H$_i$) = $\frac{\mathrm{P(H_i \cap E)}}{\mathrm{P(H_i)}}$; i = 1, 2, ……., k $\quad \therefore$ P (H$_i$ $\cap$ E) = P(H$_i$) P (E/H$_i$)

Now, Bayes' theorem states that,

$$P\,(\mathrm{H}_i/E) = \frac{P(\mathrm{H}_i)\,P(\mathrm{E}/\mathrm{H}_i)}{\sum P(\mathrm{H}_i)\,P(\mathrm{E}/\mathrm{H}_i)}\ ;\ i = 1, 2, \ldots\ldots, k$$

P (H$_i$/E) $-$ This is the posterior probability. Posteriori basically means deriving theory out of given evidence. It denotes the conditional probability of H (hypothesis), given the evidence E.

P (E/H$_i$) $-$ It is the conditional probability of the occurrence of the evidence, given the hypothesis.

P (H) $-$ It is the prior probability which is without the involvement of the data or the evidence.

P (E) − This is the probability of the occurrence of evidence regardless of the hypothesis.

Problem 3.20: You are planning a picnic today, but the morning is cloudy. Oh no! 50% of all rainy days start off cloudy! But cloudy mornings are common (40% of days start cloudy). This is a dry month (only 3 of 30 days tend to be rainy, or 10%). What is the chance of rain during the day?

Solution: P (Rain$|$Cloud) = $\frac{\mathrm{P(Cloud|Rain)\ P(Rain)}}{\mathrm{P(Cloud)}} = \frac{0.50 \text{ x } 0.10}{0.40}$ = 0.125.

=> **A 12.5% chance of rain. Not too bad, let's have a picnic!!!!!!!!!!!!!!!!!!** <=

Problem 3.21: In a box there are 70% mathematics books and 30% electrical engineering books. Among mathematics books 40% are foreign books and among electrical engineering books 50% are foreign books. A foreign book is selected. What is the probability that the selected one is an electrical engineering book?

Solution: Let, E = Foreign book, H_1= Mathematics book, H_2= Electrical engineering book.

$P(H_1) = 0.70$, $P(E|H_1) = 0.40$. So, $P(E|H_1)\,P(H_1) = 0.40 \times 0.70 = 0.28$

$P(H_2) = 0.30$, $P(E|H_2) = 0.50$. So, $P(E|H_2)\,P(H_2) = 0.50 \times 0.30 = 0.15$

$\sum P(E|H_i)\,P(H_i) = P(E|H_1)\,P(H_1) + P(E|H_2)\,P(H_2) = 0.28 + 0.15 = 0.43$

So, $P(H_2|E) = \frac{P(E|H_2)P(H_2)}{\sum P(E|H_i)P(H_i)} = \frac{0.15}{0.43} = \frac{15}{43}$.

Decision Tree Analysis: Decision tree analysis is a powerful decision-making tool which initiates a structured nonparametric approach for problem-solving. It facilitates the evaluation and comparison of the various options and their results, as shown in a decision tree. It helps to choose the most competitive alternative. Decision Trees are excellent tools for helping you to choose between several courses of action.

They provide a highly effective structure within which you can lay out options and investigate the possible outcomes of choosing those options. They also help you to form a balanced picture of the risks and rewards associated with each possible course of action. It is a widely used technique for taking crucial decisions like project selection, cost management, operation management, and production method and to deal with various other strategic issues in an organization.

Drawing a Decision Tree: A decision tree is the graphical depiction of all the possibilities or outcomes to solve a specific issue or avail a potential opportunity. You start a Decision Tree with a decision that you need to make. Draw a small square to represent this towards the left of a large piece of paper.

From this box draw out lines towards the right for each possible solution, and write that solution along the line. Keep the lines apart as far as possible so that you can expand your thoughts. At the end of each line, consider the results. If the result of taking that decision is uncertain, draw a small circle. If the result is another decision that you need to make, draw another square. Squares represent decisions, and circles represent uncertain outcomes. Write the decision or factor above the square or circle. If you have completed the solution at the end of the line, just leave it blank.

Starting from the new decision squares on your diagram, draw out lines representing the options that you could select. From the circles draw lines representing possible outcomes. Again make a brief note on the line saying what it means.

An example of the sort of thing you will end up with is shown in figure 1:

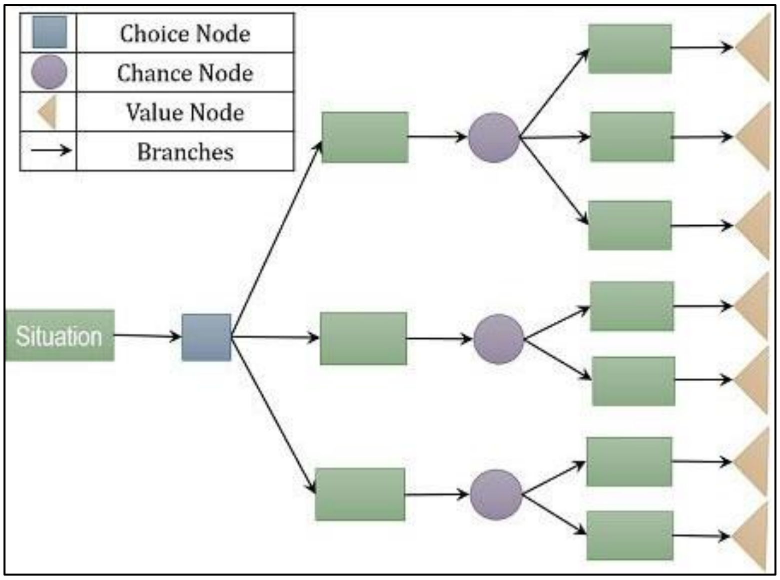


Figure 1: Decision Tree Analysis

Keep on doing this until you have drawn out as many of the possible outcomes and decisions as you can see leading on from the original decisions. Once you have done this, review your tree diagram. Challenge each square and circle to see if there are any solutions or outcomes you have not considered. If there are, draw them in. If necessary, redraft your tree if parts of it are too congested or untidy. You should now have a good understanding of the range of possible outcomes of your decisions.

By calculating the expected utility or value of each choice in the tree, you can minimize risk and maximize the likelihood of reaching a desirable outcome.

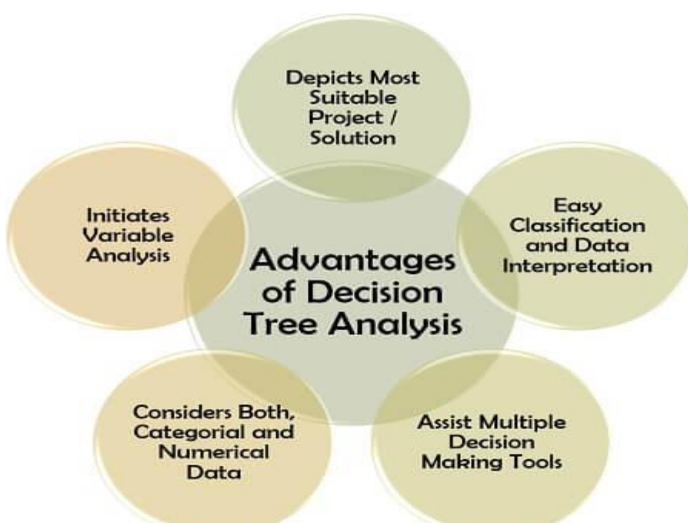
To calculate the expected utility of a choice, just subtract the cost of that decision from the expected benefits. The expected benefits are equal to the total value of all the outcomes that could result from that choice, with each value multiplied by the likelihood that it will occur.

When identifying which outcome is the most desirable, it is important to take the decision maker’s utility preferences into account. For instance, some may prefer low-risk options while others are willing to take risks for a larger benefit.

Steps in Decision Tree Analysis: Following steps simplify the interpretation process of a decision tree:

1. The first step is to understand and specify the problem area for which decision making is required.
2. The second step is interpreting and chalking out all possible solutions to the particular issue as well as their consequences.
3. The third step is presenting the variables on a decision tree along with its respective probability values
4. The fourth step is finding out the outcomes of all the variables and specifying it in the decision tree.
5. The final step is highly crucial and backs the overall analysis if this process

Benefits of using Decision Tree Analysis:

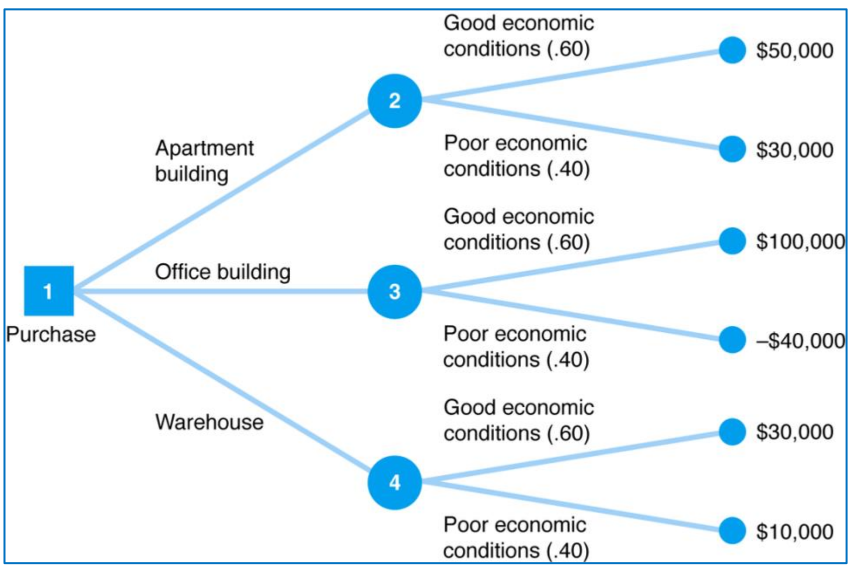


Disadvantages of using Decision Tree analysis:

1. Inappropriate for excessive data.
2. Difficult to handle numerous outcomes.
3. Chances of classification errors.
4. Not suitable for continuous variable
5. Sensitive towards biasness

Example:

A company decided to purchase a building either an apartment building, office building, or warehouse. After rigorous research, management came up with the following decision tree. Evaluate of each of these alternatives.



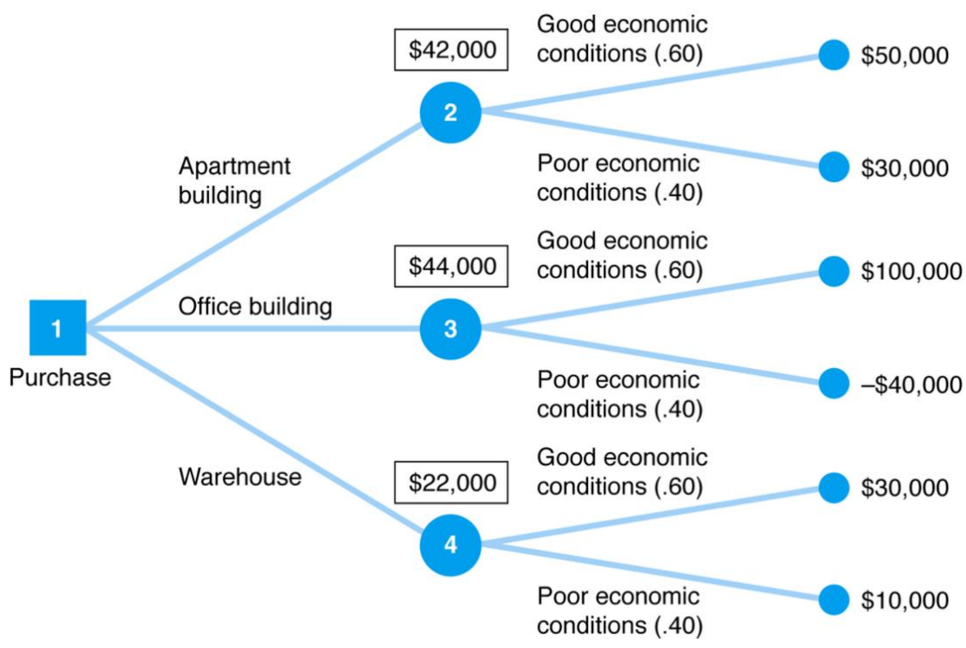
Solution: This decision tree illustrates the decision to purchase either an apartment building, office building, or warehouse. Since this is the decision being made, it is represented with a square and the branches coming off of that decision represent 3 different choices to be made. Circles 2, 3, and 4 represent probabilities in which there is uncertainty involved. The branches coming off of the circles show 2 states of nature that are possible: a) good economic conditions and b) poor economic conditions. It states here that there is a 60% chance that there will be good economic conditions and a 40% chance that there will be poor economic conditions. To further explain, at node 2, the payoff for a 60% chance of good economic conditions is $50,000. To determine whether to purchase the apartment building, office building, or warehouse, the decision maker must compute the expected value at each probability (or circle) node.

To compute the expected value at each node, the decision maker will work backward:

Expected value = (Probability of good economic conditions × Payoff associated with that probability) + (Probability of poor economic conditions × Payoff associated with that probability)

Expected value (at node 2): 0.60($50,000) + 0.40($30,000) = $42,000
Expected value (at node 3): 0.60($100,000) + 0.40(-$40,000) = $44,000
Expected value (at node 4): 0.60($30,000) + 0.40($10,000) = $22,000

These expected values are then written over top their corresponding nodes in a square box for easy-access and understanding:



The purchase we choose is whichever node has the expected value that results in the highest payoff. In this case, it is node 3, with an expected payoff of $44,000 (3).

Exercise 3

3.1. Tickets are numbered as 1 to 20, mixed up and then one is drawn randomly. Find the probability that the ticket drawn has a number which is a multiple of 3 or 5.

3.2. A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

3.3. There are 8 red, 7 blue and 6 green balls in a box. One ball is picked up randomly. What is the probability that it is neither red nor green?

3.4. What is the probability of getting a sum 9 from two throws of a die?

3.5. Three unbiased coins are tossed. What is the probability of getting at most two heads?

3.6. Two die are thrown at once. Find the probability of getting two numbers whose product is even.

3.7. There are 15 boys and 10 girls in a class. Three students are selected at random. Find the probability that 1 girl and 2 boys are selected.

3.8. In a lottery, there are 10 prizes and 25 blanks. A lottery is drawn at random. What is the probability of getting a prize?

3.9. A bag contains 4 white, 5 red and 6 blue balls. Three balls are drawn at random from the bag. Find the probability that all of them are red.

3.10. In an office there are 5 computers identified by serial number 1, 2, 3, 4 and 5. Two computers are selected by two persons who work at (i) different working hours [with replacement], (ii) same working hour [without replacement]. Find the probability that sum of the numbers of the selected computers is (a) 8 or first selected computer has the number 3, and (b) 6 given that second selected computer has the number 4.

3.11. There are 5 electronic engineers and 6 computers engineers in a mobile operator's office. A committee of 4 is to be formed to perform a duty. Find the probability that the committee will consist of (a) all electronic engineers, and (b) 2 electronic engineers and 2 computer engineers.

3.12. In a packet there are 7 Samsung and 5 Nokia mobile phone sets. Two sets are drawn one after another with replacement. Find the probability that (a) both are Samsung sets, and (b) one set is Samsung and another one is Nokia.

3.13. Find the probabilities specified in problem 3.12 if sets are drawn without replacement.

3.14. 75% signals sent from a server reach to its goal properly. Once 3 signals are checked randomly. Find the probability that out of 3, (a) at least 1 reaches properly, and (b) at best 2 reach properly.

3.15. There are 50 computers in an office. Of them, 20 are ACER and 30 are Dell. The computers are investigated and found that 12 ACER and 25 Dell computers are good. One computer is selected at random. Find the probability that the selected computer is (a) either Dell or good, and (b) ACER given that it is good.

3.16. In a communication system, signals are sent, where some of the signals are Faded (F) and some reached to the destination Properly (P). If power of the signals is not appropriate, then 50% signals are faded. An inspection team observed 3 consecutive signals. Find the probability that (a) at least one signal is faded, and (b) at best one signal is faded.

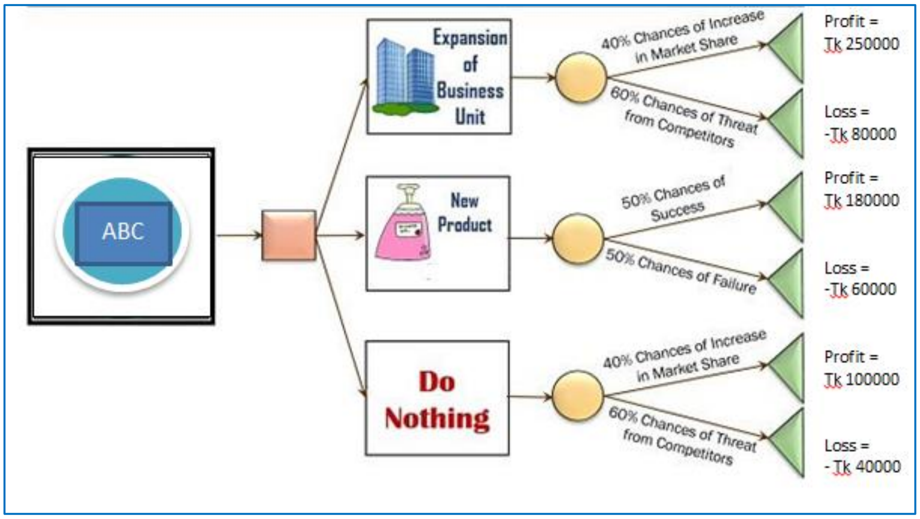
3.17. Find the probabilities specified in problem 3.16 if 10% signals are faded.

3.18. Find the probability of a number that is odd or less than 5 when a fair die is rolled.

3.19. While watching a game of Champions League football in a cafe, you observe someone who is clearly supporting Manchester United in the game. What is the probability that they were actually born within 25 miles of Manchester? Assume that the probability that a randomly selected person in a typical local bar environment is born within 25 miles of Manchester is 1/20, the chance that a person born within 25 miles of Manchester actually supports United is 7/10, and the probability that a person not born within 25 miles of Manchester supports United with probability 1/10.

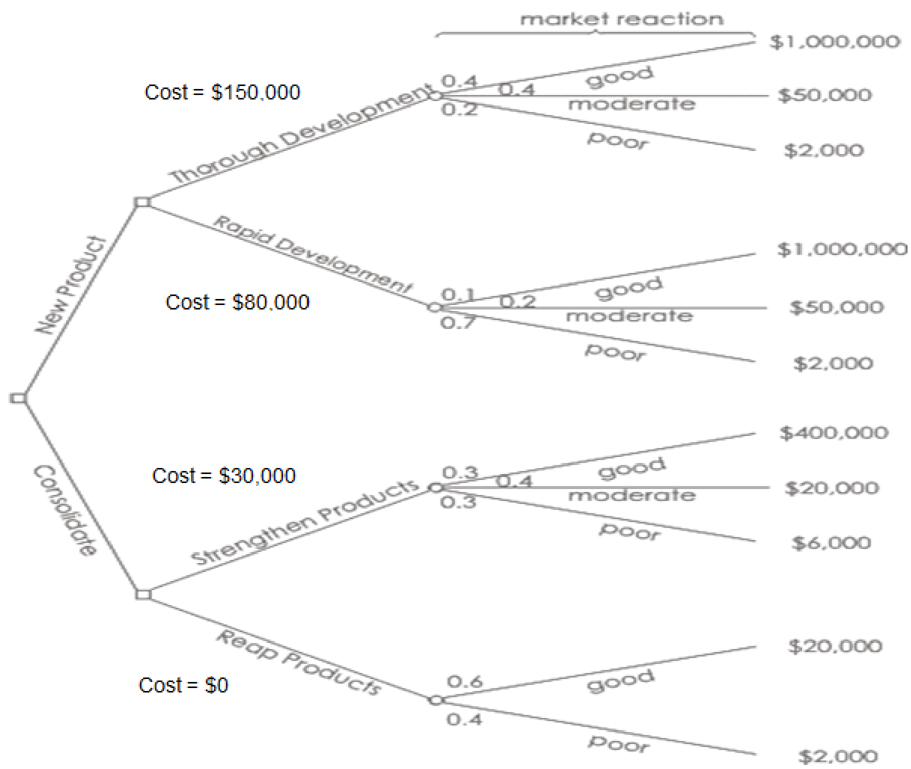
3.20. Two dies are thrown at once. Find the probability of getting two numbers whose product is even.

3.21 *ABC* Ltd. Company manufacturing household products. It was found the business is at the maturity stage, demanding some change. After arduous research, management came up with the following decision tree:



In the above decision tree, we can easily make out that the company can expand its existing unit or innovate a new product or make no changes. Evaluate each of these alternatives.

3.22 Decision tree: Should we develop a new product or consolidate?



Sample MCQs

1. In a company there are six robots numbered 1, 2, 3, 4, 5, and 6 to serve the employees. Two robots are selected one by one without replacement, Find the probability that both robots are of same number.

a) $\frac{1}{36}$ b) $\frac{0}{30}$ c) $\frac{5}{30}$ d) $\frac{1}{5}$

2. Two digits are randomly from the digits 1, 2, 3, 4, 5. Find the probability that sum of the digits will be even.

a) $\frac{1}{10}$ b) $\frac{3}{10}$ c) $\frac{4}{10}$ d) $\frac{1}{5}$

3. A group of students are trained to write a program. They are successful in 80% cases. They are asked to write three programs. Find the probability that at least one reaches properly.

a) 0.9661 b) 0.9920 c) 0.0008 d) None of them

4. Out of 20 Robots 12 are installed by Company A and 8 are installed by Company B. Eight Robots of A and 6 installations of B doing well. One installation is chosen at random to observe its performance. Find the probability that the selected Robot is of Company A under the condition that its performance is good.

a) $\frac{4}{7}$ b) $\frac{4}{10}$ c) $\frac{6}{20}$ d) $\frac{2}{6}$

5. In an office there are 15 Philips and 5 Samsung computers. If 5 computers are selected at random, what is the probability that 2 are Philips and 3 are Samsung?

a) 0.2681 b) 0.0930 c) 0.0677 d) None of them

6. Signals are sent from Station - 1 and Station - 2. Fifty signals from Station- 1 and 30 signals from Station - 2 are sent. It is known that 20% sent from Station - 1 and 30% sent from Station - 2 do not reach properly. One day on random investigation it is found that one signal is not reached properly. Find the probability that the signal is sent from Station - 2.

a) $\frac{4}{7}$ b) $\frac{9}{10}$ c) $\frac{3}{20}$ d) $\frac{9}{19}$

7. Two unbiased dice are thrown once. Find the probability that sum of the upper faces of the dice is 8 or first selected dice bears number 3.

a) $\frac{1}{36}$ b) $\frac{10}{36}$ c) $\frac{5}{30}$ d) $\frac{1}{5}$

8. In a packet there are 6 green circuits and 4 precise circuits. Two circuits are drawn one by one with replacement. Find the probability that one circuit is green and another one is precise.

a) 0.0681 b) 0.3930 c) 0.5277 d) 0.4800

9. A coin is biased so that a head is twice as likely to occur as a tail. If the coin is tossed twice, what is the probability of getting exactly 2 heads?

a) 0.4444 b) 0.3930 c) 0.5277 d) 0.4800

Reference Book: *Statistics and Probability for Engineering Applications*, D. J. Decoursey, Elsevier science, 2003.