Applied Statistics

Course Code: MAT1011

Dr. Sukanta Nayak

Department of Mathematics
School of Advanced Sciences
VIT-AP University, Amaravati
Andhra Pradesh

Module No. 2

Probability

Sample Space

- Statistician is often dealing with either numerical data, representing counts or measurements, or categorical data, which can be classified according to some criterion. We shall refer to any recording of information, whether it be numerical or categorical, as an **observation**.
- Statisticians use the word experiment to describe any process that generates a set of data. A simple example of a statistical experiment is the tossing of a coin. In this experiment, there are only two possible outcomes, heads or tails.
- Another experiment might be the launching of a missile and observing of its velocity at specified times.
- The opinions of voters concerning a new sales tax can also be considered as observations of an experiment.
- The set of all possible outcomes of a statistical experiment is called the **sample space** and is represented by the symbol S.

Sample Space

- Each outcome in a sample space is called an element or a member of the sample space, or simply a sample point. If the sample space has a finite number of elements, we may list the members separated by commas and enclosed in braces.
- The sample space S, of possible outcomes when a coin is flipped, may be written

$$S = \{H, T\},$$

where H and T correspond to heads and tails, respectively.

- Consider the experiment of tossing a die. If we are interested in the number that shows on the top face, the sample space is

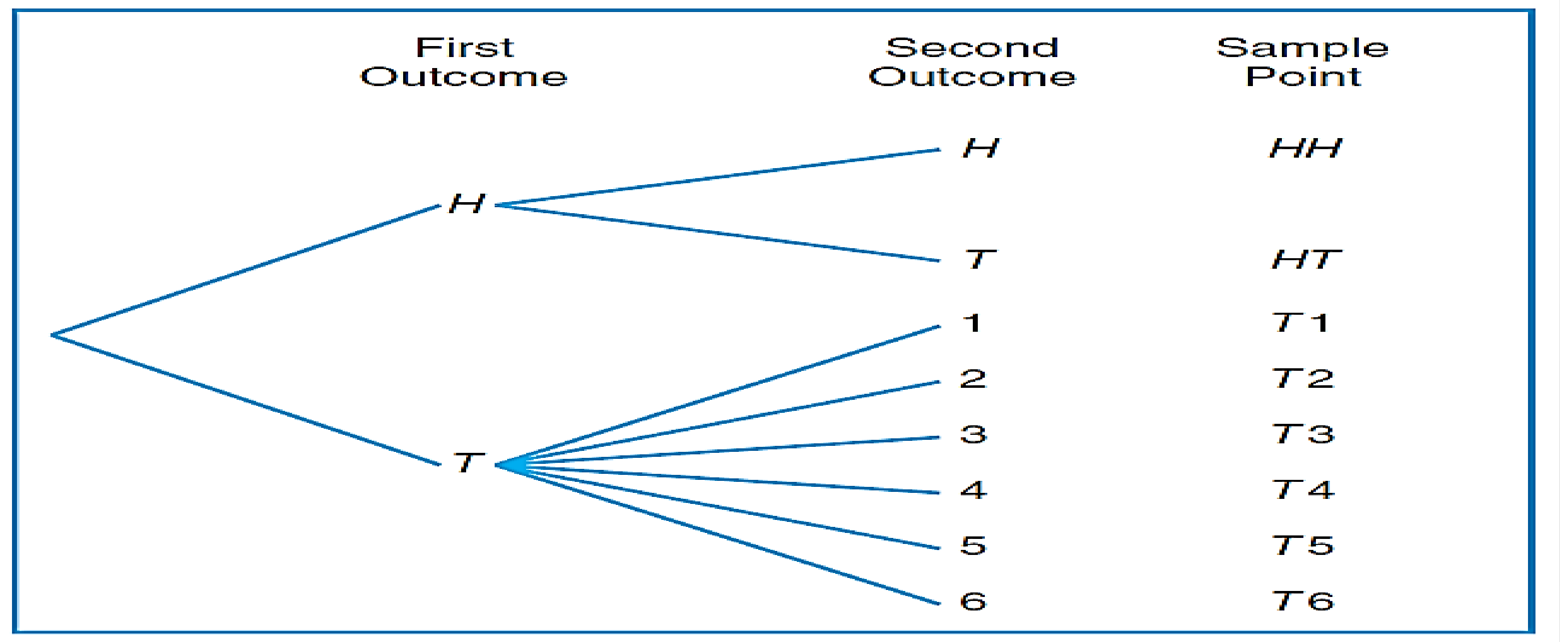
$$S_1 = \{1, 2, 3, 4, 5, 6\}.$$

- If we are interested only in whether the number is even or odd, the sample space is simply

$$S_2 = \{\text{even}, \text{odd}\}.$$

Sample Space

- In some experiments, it is helpful to list the elements of the sample space systematically by means of a tree diagram.
- An experiment consists of flipping a coin and then flipping it a second time if a head occurs. If a tail occurs on the first flip, then a die is tossed once.



Here the sample space is $S = \{HH, HT, T1, T2, T3, T4, T5, T6\}$.

Event

- An **event** is a subset of a sample space.
- Given the sample space $S = \{t|t > 0\}$, where t is the life in years of a certain electronic component, then the event A that the component fails before the end of the fifth year is the subset

$$\boldsymbol{A = \{t|0 \leq t \leq 5\}}.$$

- It is conceivable that an event may be a subset that includes the entire sample space S or a subset of S called the null set and denoted by the symbol ϕ, which contains no elements at all.
- The **complement of an event** $\boldsymbol{A}$ with respect to S is the subset of all elements of S that are not in A. We denote the complement of A by the symbol A'.
- Consider the sample space $S = \{\text{book, cell phone, mp3, paper, stationery, laptop}\}$.

Let $A = \{\text{book, stationery, laptop, paper}\}$. Then the complement of A is $A' = \{\text{cell phone, mp3}\}$.

Event

- The intersection of two events A and B, denoted by the symbol $A \cap B$, is the event containing all elements that are common to A and B.
- Two events A and B are mutually exclusive, or disjoint, if $A \cap B = \phi$, that is, if A and B have no elements in common.
- The union of the two events A and B, denoted by the symbol $A \cup B$, is the event containing all the elements that belong to A or B or both.

If $M = \{x \mid 3 < x < 9\}$ and $N = \{y \mid 5 < y < 12\}$, then

$$M \cup N = \{z \mid 3 < z < 12\}.$$

Counting Sample Points

- If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the two operations can be performed together in $n_1 n_2$ ways.
- If an operation can be performed in n_1 ways, and if for each of these a second operation can be performed in n_2 ways, and for each of the first two a third operation can be performed in n_3 ways, and so forth, then the sequence of k operations can be performed in $n_1 n_2 \cdots n_k$ ways.
- A permutation is an arrangement of all or part of a set of objects.
- The number of permutations of n objects is $n!$.
- The number of permutations of n distinct objects taken r at a time is $n_{P_r} = \frac{n!}{(n-r)!}$.
- The number of permutations of n objects arranged in a circle is $(n - 1)!$.

Counting Sample Points

- The number of distinct permutations of n things of which n_1 are of one kind, n_2 of a second kind, ..., n_k of a kth kind is $\dfrac{n!}{n_1!n_2!\cdots n_k!}$.

The number of ways of partitioning a set of n objects into r cells with n_1 elements in the first cell, n_2 elements in the second, and so forth, is

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!},$$

where $n_1 + n_2 + \cdots + n_r = n$.

The number of combinations of n distinct objects taken r at a time is

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}.$$

Probability of an Event

The **probability** of an event A is the sum of the weights of all sample points in A. Therefore,

$$0 \leq P(A) \leq 1, \quad P(\phi) = 0, \quad \text{and} \quad P(S) = 1.$$

Furthermore, if $A_1, A_2, A_3, \ldots$ is a sequence of mutually exclusive events, then

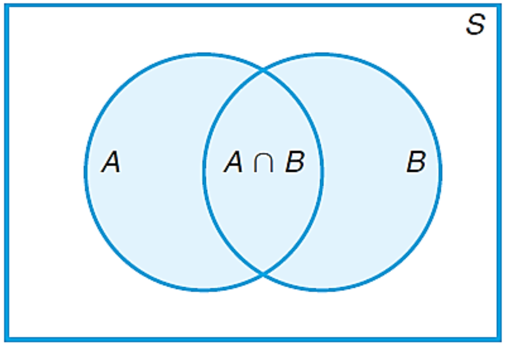
$$P(A_1 \cup A_2 \cup A_3 \cup \cdots) = P(A_1) + P(A_2) + P(A_3) + \cdots .$$

If an experiment can result in any one of N different equally likely outcomes, and if exactly n of these outcomes correspond to event A, then the probability of event A is

$$P(A) = \frac{n}{N}.$$

If A and B are two events, then

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



Probability of an Event

If A and B are mutually exclusive, then

$$P(A \cup B) = P(A) + P(B).$$

If $A_1,\ A_2, \ldots, A_n$ are mutually exclusive, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n).$$

If $A_1,\ A_2, \ldots, A_n$ is a partition of sample space S, then

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(A_1) + P(A_2) + \cdots + P(A_n) = P(S) = 1.$$

For three events A, B, and C,

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

If A and A' are complementary events, then

$$P(A) + P(A') = 1.$$

Conditional Probability, Independence, and the Product Rule

The probability of an event B occurring when it is known that some event A has occurred is called a conditional probability and is denoted by $P(B|A)$. The symbol $P(B|A)$ is usually read "the probability that B occurs given that A occurs" or simply "the probability of B, given A."

The conditional probability of B, given A, denoted by $P(B|A)$, is defined by

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad \text{provided} \quad P(A) > 0.$$

	Employed	Unemployed	Total
Male	460	40	500
Female	140	260	400
Total	600	300	900

M: a man is chosen,

E: the one chosen is employed.

Using the reduced sample space E, we find that

$$P(M|E) = \frac{460}{600} = \frac{23}{30}.$$

Let $n(A)$ denote the number of elements in any set A. Using this notation, since each adult has an equal chance of being selected, we can write

$$P(M|E) = \frac{n(E \cap M)}{n(E)} = \frac{n(E \cap M)/n(S)}{n(E)/n(S)} = \frac{P(E \cap M)}{P(E)},$$

where $P(E \cap M)$ and $P(E)$ are found from the original sample space S. To verify this result, note that

$$P(E) = \frac{600}{900} = \frac{2}{3} \quad \text{and} \quad P(E \cap M) = \frac{460}{900} = \frac{23}{45}.$$

Hence,

$$P(M|E) = \frac{23/45}{2/3} = \frac{23}{30},$$

as before.

Question

The probability that a regularly scheduled flight departs on time is $P(D) = 0.83$; the probability that it arrives on time is $P(A) = 0.82$; and the probability that it departs and arrives on time is $P(D \cap A) = 0.78$. Find the probability that a plane (a) arrives on time, given that it departed on time, and (b) departed on time, given that it has arrived on time.

Answer

(a) 0.94

(b) 0.95

Question

The concept of conditional probability has countless uses in both industrial and biomedical applications. Consider an industrial process in the textile industry in which strips of a particular type of cloth are being produced. These strips can be defective in two ways, length and nature of texture. For the case of the latter, the process of identification is very complicated. It is known from historical information on the process that 10% of strips fail the length test, 5% fail the texture test, and only 0.8% fail both tests. If a strip is selected randomly from the process and a quick measurement identifies it as failing the length test, what is the probability that it is texture defective?

Answer

Consider the events L: length defective, T: texture defective. Given that the strip is length defective, the probability that this strip is texture defective is given by

$$P(T|L) = \frac{P(T \cap L)}{P(L)} = \frac{0.008}{0.1} = 0.08.$$

In the die-tossing experiment discussed on page 62, we note that $P(B|A) = 2/5$ whereas $P(B) = 1/3$. That is, $P(B|A) \neq P(B)$, indicating that B depends on A. Now consider an experiment in which 2 cards are drawn in succession from an ordinary deck, with replacement. The events are defined as

A: the first card is an ace,

B: the second card is a spade.

Since the first card is replaced, our sample space for both the first and the second draw consists of 52 cards, containing 4 aces and 13 spades. Hence,

$$P(B|A) = \frac{13}{52} = \frac{1}{4} \quad \text{and} \quad P(B) = \frac{13}{52} = \frac{1}{4}.$$

That is, $P(B|A) = P(B)$. When this is true, the events A and B are said to be **independent**.

Two events A and B are **independent** if and only if

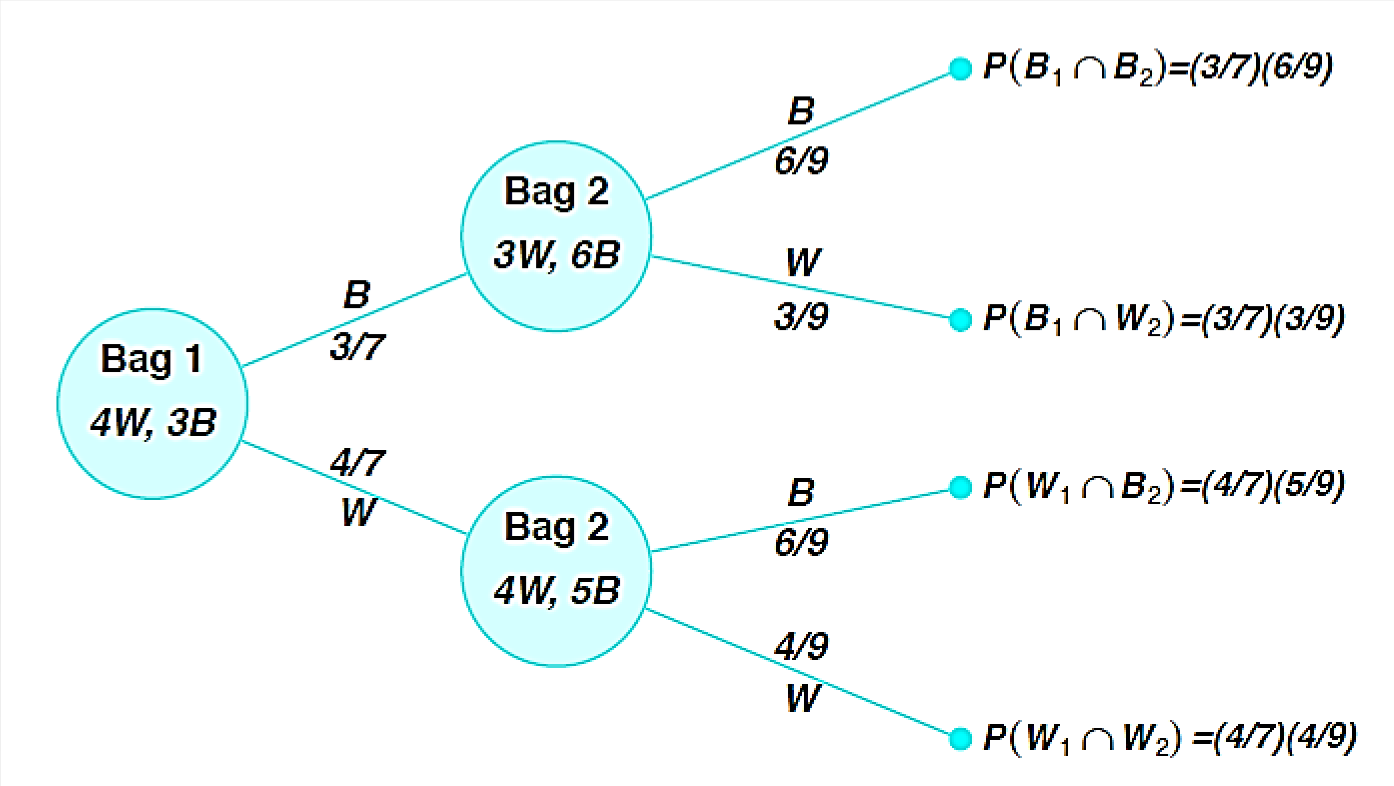
$$P(B|A) = P(B) \quad \text{or} \quad P(A|B) = P(A),$$

assuming the existences of the conditional probabilities. Otherwise, A and B are **dependent**.

If in an experiment the events A and B can both occur, then

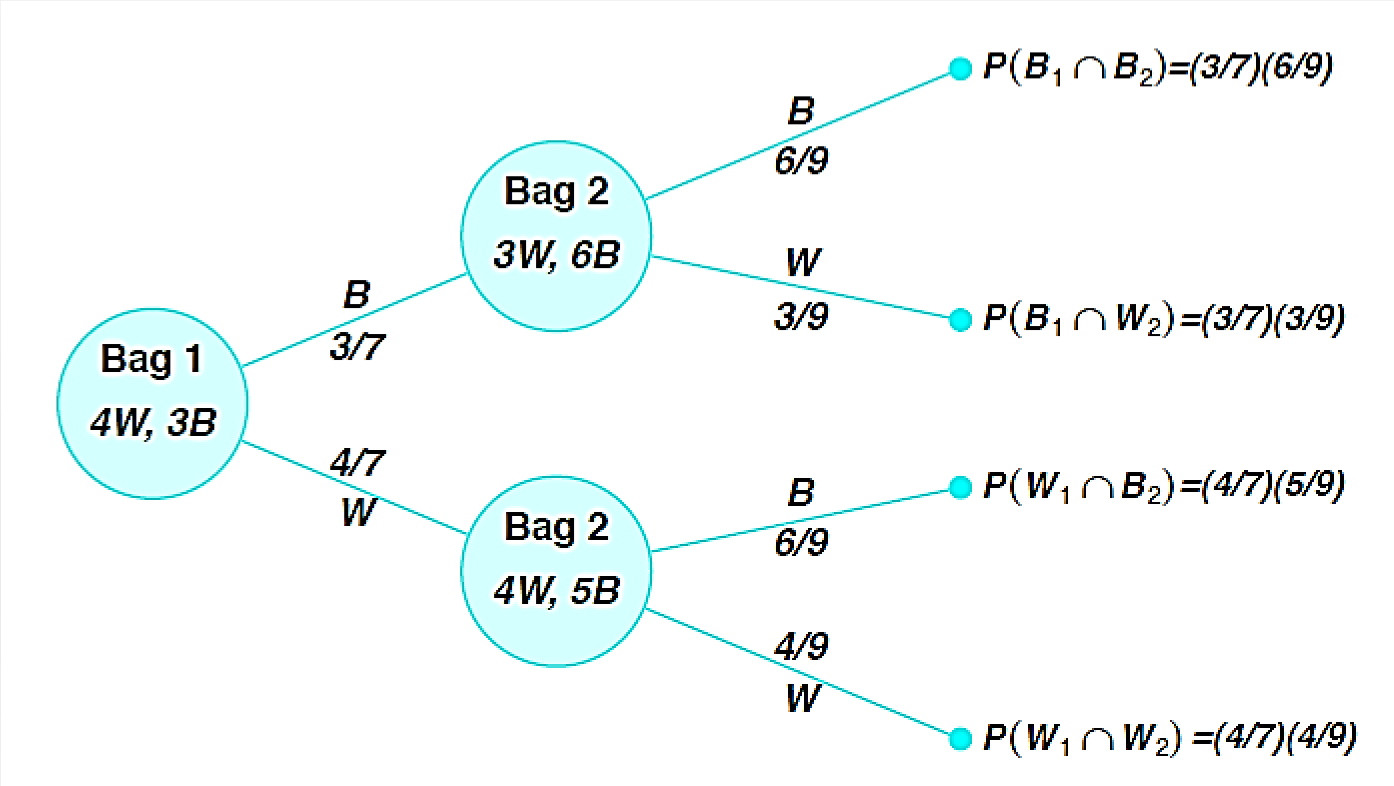
$$P(A \cap B) = P(A)P(B|A), \text{ provided } P(A) > 0.$$

One bag contains 4 white balls and 3 black balls, and a second bag contains 3 white balls and 5 black balls. One ball is drawn from the first bag and placed unseen in the second bag. What is the probability that a ball now drawn from the second bag is black?



Let B_1, B_2, and W_1 represent, respectively, the drawing of a black ball from bag 1, a black ball from bag 2, and a white ball from bag 1.

$$\begin{aligned} P[(B_1 \cap B_2) \text{ or } (W_1 \cap B_2)] &= P(B_1 \cap B_2) + P(W_1 \cap B_2) \\ &= P(B_1)P(B_2|B_1) + P(W_1)P(B_2|W_1) \\ &= \left(\frac{3}{7}\right)\left(\frac{6}{9}\right) + \left(\frac{4}{7}\right)\left(\frac{5}{9}\right) = \frac{38}{63}. \end{aligned}$$



Two events A and B are independent if and only if

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

Therefore, to obtain the probability that two independent events will both occur, we simply find the product of their individual probabilities.

If, in an experiment, the events $A_1, A_2, \ldots, A_k$ can occur, then

$$\begin{aligned} &P(A_1 \cap A_2 \cap \cdots \cap A_k) \\ &\qquad = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_k|A_1 \cap A_2 \cap \cdots \cap A_{k-1}). \end{aligned}$$

If the events $A_1, A_2, \ldots, A_k$ are independent, then

$$P(A_1 \cap A_2 \cap \cdots \cap A_k) = P(A_1)P(A_2) \cdots P(A_k).$$

A collection of events $\mathcal{A} = \{A_1, \ldots, A_n\}$ are mutually independent if for any subset of $\mathcal{A}$, $A_{i_1}, \ldots, A_{i_k}$, for $k \leq n$, we have

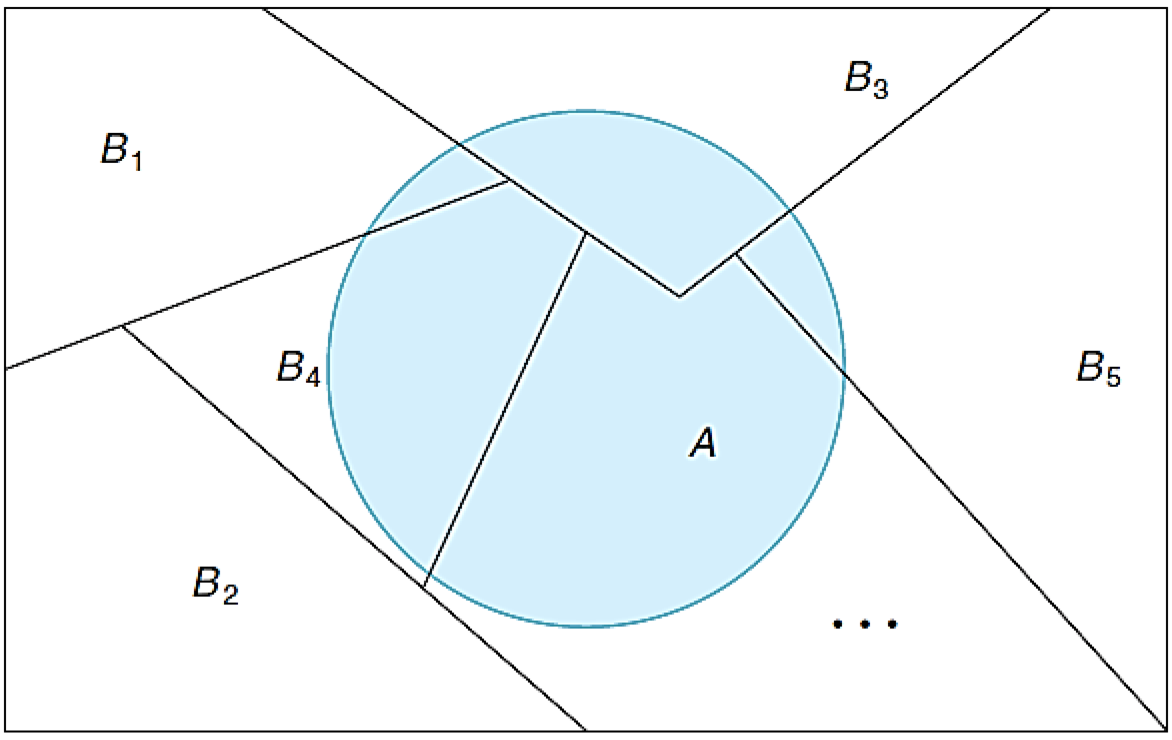
$$P(A_{i_1} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_k}).$$

Bayes' Rule

Theorem of total probability

If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A of S,

$$P(A) = \sum_{i=1}^{k} P(B_i \cap A) = \sum_{i=1}^{k} P(B_i)P(A|B_i).$$



Bayes' Rule

Question

In a certain assembly plant, three machines, B_1, B_2, and B_3, make 30%, 45%, and 25%, respectively, of the products. It is known from past experience that 2%, 3%, and 2% of the products made by each machine, respectively, are defective. Now, suppose that a finished product is randomly selected. What is the probability that it is defective?

Answer

Consider the following events:

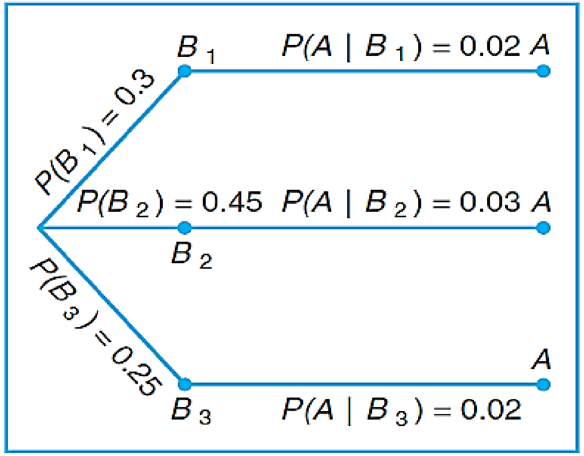
A: the product is defective,

B_1: the product is made by machine B1,

B_2: the product is made by machine B2,

B_3: the product is made by machine B3.

Applying the rule of elimination, we can write



$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + P(B_3)P(A|B_3).$$

$$P(B_1)P(A|B_1) = (0.3)(0.02) = 0.006,$$
$$P(B_2)P(A|B_2) = (0.45)(0.03) = 0.0135,$$
$$P(B_3)P(A|B_3) = (0.25)(0.02) = 0.005,$$

$$P(A) = 0.006 + 0.0135 + 0.005 = 0.0245.$$

Bayes' Rule

(Bayes' Rule) If the events $B_1, B_2, \ldots, B_k$ constitute a partition of the sample space S such that $P(B_i) \neq 0$ for $i = 1, 2, \ldots, k$, then for any event A in S such that $P(A) \neq 0$,

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_{i=1}^{k} P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_{i=1}^{k} P(B_i)P(A|B_i)} \quad \text{for } r = 1, 2, \ldots, k.$$

Question

A manufacturing firm employs three analytical plans for the design and development of a particular product. For cost reasons, all three are used at varying times. In fact, plans 1, 2, and 3 are used for 30%, 20%, and 50% of the products, respectively. The defect rate is different for the three procedures as follows:

$$P(D|P_1) = 0.01, \qquad P(D|P_2) = 0.03, \qquad P(D|P_3) = 0.02,$$

where $P(D|P_j)$ is the probability of a defective product, given plan j. If a random product was observed and found to be defective, which plan was most likely used and thus responsible?

Bayes' Rule

Answer

From the statement of the problem

$$P(P_1) = 0.30, \quad P(P_2) = 0.20, \quad \text{and} \quad P(P_3) = 0.50,$$

we must find $P(P_j|D)$ for $j = 1, 2, 3$. Bayes' rule (Theorem 2.14) shows

$$\begin{aligned} P(P_1|D) &= \frac{P(P_1)P(D|P_1)}{P(P_1)P(D|P_1) + P(P_2)P(D|P_2) + P(P_3)P(D|P_3)} \\ &= \frac{(0.30)(0.01)}{(0.3)(0.01) + (0.20)(0.03) + (0.50)(0.02)} = \frac{0.003}{0.019} = 0.158. \end{aligned}$$

Similarly,

$$P(P_2|D) = \frac{(0.03)(0.20)}{0.019} = 0.316 \text{ and } P(P_3|D) = \frac{(0.02)(0.50)}{0.019} = 0.526.$$

The conditional probability of a defect given plan 3 is the largest of the three; thus a defective for a random product is most likely the result of the use of plan 3.

References

1. Norean R. Sharpe, Richard D. De Veaux, Paul F. Velleman, Business Statistics, Fourth Edition, Pearson Education. 2019.
2. R.E. Walpole, R.H. Mayers, S.L. Mayers and K.Ye, Probability and Statistics for engineers and scientists, 9th Edition, Pearson Education, 2018.
3. Douglas A. Wolfe, Grant Schneider, Intuitive Introductory Statistics, Springer, 2017.
4. Miller & Freund's, Probability and statistics for engineers, 8th edition, Pearson publication, 2018.
5. Richard I. Levin, David S. Rubin, Masood H. Siddiqui, Sanjay Rastogi, Statistics for Management, 8th Edition, Pearson Publications, 2018.
6. S. C. Gupta and V. K. Kapoor, Fundamentals of Applied Statistics, S. Chand, 2006.

THE END