Probability and Statistics
MA-202
Summary

Reference

- Rohatgi, V. K., & Saleh, A. M. E. (2015). An Introduction to Probability and Statistics. John Wiley & Sons.

Random Experiment: A random (or statistical) experiment is an experiment in which:

- All possible outcomes of the experiment are known in advance.
- Any performance of the experiment yields an outcome that is unpredictable beforehand.
- The experiment can be repeated under identical conditions.

Our goal is to study this uncertainty of a random experiment. For this purpose, we associate with each such experiment a set Ω, the set of all possible outcomes of the experiment. This collection Ω must be

- **mutually exclusive:** that is, if head comes in a coin tossing experiment then tail doesn't come and vice-versa (or there is a unique outcome).
- **collectively exhaustive:** no matter what happens in the experiment, we always obtain an outcome that has been included in Ω.

Recap

Example 1. *Consider tossing a coin two times. Let A be a subset that includes outcomes containing atleast one head in two throws and let B be a subset that including outcomes where both tosses results in tails. Find the probability of A and B.*

Solution The set of possible outcomes, Ω, is given by

$$\Omega = \{HH, HT, TH, TT\}$$

where H denotes a head and T denotes a tail. Clearly, we have

$$A = \{HH, HT, TH\}, \quad B = \{TT\}.$$

The required probabilities are given by

$$P(A) = \frac{3}{4}, \quad P(B) = \frac{1}{4}.$$

Example 2. *A die is rolled twice. Let A be the collection of outcomes that the first throw shows a number ≤ 2, and B be the collection of outcomes that the second throw shows at least 4. Find probability of A and B.*

Solution The set of possible outcomes, Ω, is given by

$$\Omega = \{(i, j)|1 \leq i, j \leq 6\}.$$

We have

$$\begin{aligned} A &= \{(i, j)|i = 1, 2, j = 1, 2, 3, 4, 5, 6\} \\ B &= \{(i, j)|i = 1, 2, 3, 4, 5, 6, j = 4, 5, 6\} \end{aligned}$$

The required probabilities are given by

$$P(A) = \frac{12}{36} = \frac{1}{3}, \quad P(B) = \frac{18}{36} = \frac{1}{2}.$$

Remark 1. *The random experiments in above examples have a finite number of possible outcomes and it is assumed that all the outcomes are equally likely. Thus, the probability of any subset is defined as the proportion of the number of favorable outcomes to the total number of all possible outcomes, i.e.,*

$$P(A) = \frac{n(A)}{n(\Omega)}$$

where $n(A)$ *denotes the cardinality of set* A*. This is the classical definition of the probability given by* ***Laplace*** *in his monumental work,* ***Theorie analytique des probabilites*** *(1812).*

Remark 2. *Next, suppose the* Ω *is countably infinite, say,* $\Omega = \{\omega_1, \omega_2, \ldots, \}$*. For example, we select a number at random from the set of natural numbers,* $\mathbb{N}$*. The classical definition is not applicable in such cases. Further, we cannot have uniform probability for each outcome.*

- *If* $P(\omega_i) = p > 0, \forall i$*, we can find sufficiently large* N *such that* $Np > 1$*, hence a contradiction.*

Extension of the classification definition

An extension of Laplace's classical definition was used to evaluate probabilities of sets with infinite outcomes. According to this extension, if Ω is some region with a well-defined measure (length, area, volume, etc.), the probability that a point chosen at random lies in a sub-region A of Ω is the ratio

$$P(A) = measure(A)/measure(\Omega).$$

Example 3. *Consider throwing a dart on the square target and viewing the point of impact as outcome. The outcome of a single through is going to be a point in the region denoted by* Ω *and given by*

$$\Omega = \{(x, y) | 0 \leq x, y \leq 1)\}.$$

Since there are infinitely many points, the sample space is of size infinite and is uncountable. Find the probability that dart will hit the region $\{(x, y) | x + y \leq 0.5\}$*.*

Solution Let A denotes the target region, i.e.,

$$A = \{(x, y) | x + y \leq 0.5\}.$$

We can find the probability of A as

$$P(A) = \frac{area \ \ of \ \ A}{area \ \ of \ \ \Omega} = \frac{\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2}}{1 \times 1} = \frac{1}{8}.$$

This extension proved instrumental in resolving numerous geometric probability problems. However, the challenge lies in the fact that one can arbitrarily define "at random" in various ways, and distinct definitions can yield different solutions.

Example 4. ***(Bertrand's Paradox)*** *A chord is drawn at random in the unit circle. What is the probability that the chord is longer than the side of the equilateral triangle inscribed in the circle?*

Solution 1 The length of a chord is uniquely characterized by the midpoint's distance from the circle's center. Because the circle is symmetrical, we assume the midpoint M is on a fixed line from the center O (see Figure 1). It is evident that the chord's length exceeds the side of the inscribed equilateral triangle when the length of OM is less than half of the radius. Consequently, the desired probability is 0.5.

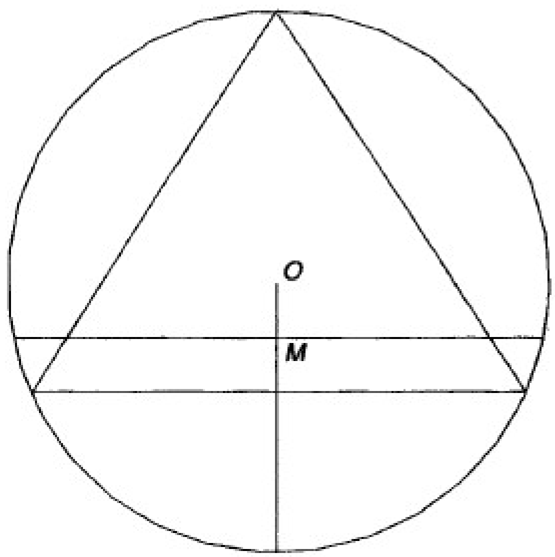

Figure 1: Solution 1

Solution 2 Let us pick a point, say V, on the circle. Now, draw a tangent to the circle at V. Draw a line at random angle Φ passing through V which intersects the circle at some point, thus forming a chord. Clearly, the length of the chord is greater than the side of the equilateral triangle, if $\Phi \in [\frac{\pi}{3}, \frac{2\pi}{3}]$. Therefore, the required probability is $\frac{\frac{2\pi}{3}-\frac{\pi}{3}}{\pi-0} = \frac{1}{3}$.

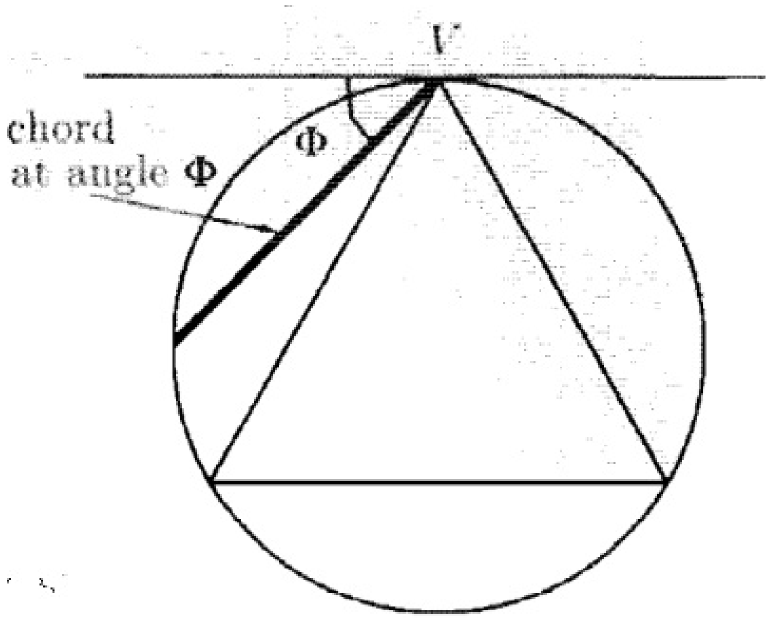

Figure 2: Solution 2

Axiomatic definition of probability

A. N. Kolmogorov, 1933.

For the reasons above, it is necessary to develop a consistent probability theory, and the idea is to define directly the probability of subsets of Ω. In other words, restrict our attention to "good subsets" of the sample space on which the probability is well-defined. The collection of such "good subsets" is known as $\sigma-$field, which is defined below.

Definition 1. ***(σ- Field)*** *Let $\mathcal{F}$ denote the collection of subsets of Ω which satisfies the following properties:*

a) $\emptyset \in \mathcal{F}$.

b) If $A \in \mathcal{F}$ then the complement set A, $A^c \in \mathcal{F}$.

c) If $A_i, i = 1, 2, \ldots, \in \mathcal{F}$ then $\cup_i A_i \in \mathcal{F}$.

Then, $\mathcal{F}$ is called as σ-field on Ω. The pair $(\Omega, \mathcal{F})$ is called the sample space. The default $\mathcal{F}$ is the collection of all possible subsets of Ω including the null set which is also called as the power set on Ω.

- The elements of Ω are called sample points.
- Any set $A \in \mathcal{F}$ is known as an event.

- We say that an event A occurs if the outcome of the experiment corresponds to a point in A.
- If the set Ω contains only a finite number of points, we say that $(\Omega, \mathcal{F})$ is a finite sample space.
- If Ω contains at most a countable number of points, we call $(\Omega, \mathcal{F})$ a discrete sample space.
- If Ω contains uncountably many points, we say that $(\Omega, \mathcal{F})$ is an uncountable sample space. For instance, if $\Omega = \mathbb{R}^p$ or some rectangle in $\mathbb{R}^p$, we call it a continuous sample space.

Definition 2. ***(Probability)*** *Define a real-valued function $\mathbb{P}$ on $\mathcal{F}$ satisfying the following conditions:*

a) $\mathbb{P}(A) \geq 0 \;\; \forall\; A \in \mathcal{F}$.

b) $\mathbb{P}(\Omega) = 1$.

c) Let, $A_1, A_2, \ldots$ *be mutually exclusive events in* $\mathcal{F}$*, i.e.,* $A_i \bigcap A_j = \emptyset \quad \forall\; i \neq j$*, then*

$$\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

The function P is called a probability function over sample space $(\Omega, \mathcal{F})$, and the triplet $(\Omega, \mathcal{F}, P)$ is called a probability space/model.

Remark 3. *Note that, different $\mathcal{F}$ and different P will give different probability model for a given Ω.*

Discussion

(1) **Ω is Finite**

If Ω contains n points $\{\omega_1, \ldots, \omega_n\}$, with $n < \infty$, it is sufficient to define P for ω_i's, say,

$$P(\omega_i) = p_i.$$

One can then consider $\mathcal{F}$ as the class of all subsets of Ω, and obtain probability of any event (subset of Ω) as

$$P(A) = \sum_{\{j \;\mid\; \omega_j \in A\}} P(\omega_j) = \sum_{\{j \;\mid\; \omega_j \in A\}} p_j$$

From point (a) in the definition of probability, we need $p_i \geq 0$, and from point (b), we need $\sum_{j=1}^{n} p_j = 1$.

Example 5. *One such function is the equally likely assignment (uniform probabilities). According to this assignment,* $P(\{\omega_j\}) = \frac{1}{n}$*,* $j = 1, 2, \ldots, n$*. Thus* $P(A) = \frac{m}{n}$ *if* A *contains* m *elementary events,* $1 < m < n$*. Note that this is the classical definition of probability.*

2. Ω **is countably infinite** As noted earlier, if Ω is discrete and contains a countable number of points, one cannot assign uniform probabilities. However, due to countable additivity of probability function, it suffices to make the assignment for each elementary event. Similar to the finite case, if Ω contains countably infinite points $\{\omega_1, \omega_2, \ldots\}$, it is sufficient to define P for ω_i's, say,

$$P(\omega_i) = p_i \quad \forall i$$

because using the countable additivity of probability, one can find probability of any event (subset of Ω) as

$$P(A) = \sum_{\{j \mid \omega_j \in A\}} P(\omega_j) = \sum_{\{j \mid \omega_j \in A\}} p_j$$

- From point (a) in the definition of probability, we need $p_i \geq 0$, and from point (b), we need $\sum_j p_j = 1$.
- In this case, $\mathcal{F}$ can be considered as the power set of Ω.

Example 6. *Let* $\mathbb{N} = \{1, 2, 3, \ldots\}$ *be the set of positive integers, and define* P *as follows:*

$$P(\{n\}) = \frac{1}{2^n}, \quad n = 1, 2, \ldots$$

Since $\frac{1}{2^n} \geq 0$ *and* $\sum_{i \in \mathbb{N}} P(\{i\}) = 1$*,* P *defines a probability.*

- $P(number \;\; is \;\; even) = P(\{2, 4, 6, \ldots\}) = P(\{2\}) + P(\{4\}) + \ldots = \frac{1}{2^2} + \frac{1}{2^4} + \ldots = \frac{1}{3}$.
- $P(number \;\; is \;\; multiple \;\; of \;\; 3) = P(\{3, 6, 9, \ldots\}) = P(\{3\}) + P(\{6\}) + \ldots = \frac{1}{2^3} + \frac{1}{2^6} + \ldots = \frac{1}{7}$.

Example 7. *Suppose* $\Omega = \{0, 1, 2, \ldots\}$ *and* $P(\{x\}) = (1 - \theta)\theta^x$*,* $x = 0, 1, 2, \ldots$*,* $0 < \theta < 1$*. One can verify that all the axioms are satisfied.*

Remark 4. *We observe from (1) and (2) above that if* Ω *contains countable number of points, it is sufficient to define probabilities of sample points* ω_i*'s and probability of any subset of* Ω *can be obtained using the countable (or finite) additivity. Thus, we need not bother about what* $\mathcal{F}$ *is* *(which, of course, is power set of Ω).*

3. Ω **is uncountable** Probabilistic models with continuous sample space differ from their discrete counterparts in that the probabilities of the single element events may not be sufficient to characterize the probability.

   - Clearly, one cannot make an equally likely assignment of probabilities. (See the case of countable Ω)

   - Indeed, one cannot assign positive probability to each elementary event without violating the axiom $P(\Omega) = 1$.

   Thus, in this case, one assigns probabilities directly to good subsets. In this course, we will mostly be working with intervals in $\mathbb{R}$ (or their higher dimensional counterparts).

   Example 8. *Let $\Omega = [0, \infty)$. Define P as follows: For each interval $I \subset \Omega$,*

$$P(I) = \int_I e^{-x} dx.$$

   Clearly, $P(I) > 0$, $P(\Omega) = 1$, and P is countably additive by properties of integrals. Thus, P is a probability function on Ω.

   Using the above P, we can find probabilities of different events. For example,

$$\begin{aligned}
P((0, \infty)) &= \int_0^\infty e^{-x} dx = 1 \\
P((1, \infty)) &= \int_1^\infty e^{-x} dx = e^{-1} \\
P([1, \infty)) &= \int_1^\infty e^{-x} dx = e^{-1} \\
P((0, 1)) &= \int_0^1 e^{-x} dx = e^{-0} - e^{-1}
\end{aligned}$$

   Remark 5. *The probability of elementary events, i.e., singletons is zero for a continuous sample space.*

   Remark 6. *Note that if $P(A) = 0$ for some event A, we call A an event with zero probability or a null event. However, it does not follow that $A = \emptyset$. For example, $P(\{1\}) = 0$ in the above example. Similarly, if $P(B) = 1$ for some event B, we call B a certain event, but it does not follow that $B = \Omega$. For example, $P((0, \infty)) = 1$ in the above example.*

Example 9. *Let* $\Omega = \{\omega : 0 \leq \omega \leq a\}$. *Define*

$$P\{[0, x]\} = \frac{x}{a}, \quad 0 < x \leq a.$$

Check whether all probability axioms are satisfied or not? Find $P\{[\frac{a}{4}, \frac{a}{2}]\}$.

Other properties

Theorem 1. *Given the events* A *and* B *such that* $A \subset B$, *show that*

(i) $P(A) \leq P(B)$ *(*P *is monotone)*

(ii) $P(B - A) = P(B) - P(A)$ *(*P *is subtractive)*

Remark 7. *It follows from Theorem 1 that* $P(A) \leq 1$ *for all events.*

Theorem 2. ***(Addition Rule)*** *For the events* A, B, *we have*

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

and hence $P(A^c) = 1 - P(A)$.

Conditional Probability and Bayes Theorem

Up to this point, our calculations of event probabilities have been based on the premise that we possess no additional information about the experiment beyond the sample space. Nevertheless, there are instances where we are aware that an event B has occurred. How can we incorporate this knowledge when making assertions about the outcome of another event A?

Definition 3. (Conditional Probability) *Let* (Ω, S, P) *be a probability space. Let* $B \in S$ *be any event with* $P(B) > 0$. *For any other* $A \in S$, *we have the conditional probability of* A *given* B *as*

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0$$

If $P(B) = 0$, *the conditional probability is not defined.*

The computation of conditional probability can be done by the above definition using the reduced sample space. The reduced sample space is the collection of outcomes in which $P(B) > 0$. The draw back of reduced sample space is that, when the number of elements of Ω is large, the collection is difficult.

Example 10. *Consider the experiment of tossing a coin twice. The sample space, Ω, of the event is $\{\{H,H\},\{H,T\},\{T,H\},\{T,T\}\}$. Let us find the probability of getting a tail given that a head has occurred on the first toss.*

Solution:
Let the event of getting a head on the first toss, and the event of getting a tail in the second toss be A and B respectively. Then,

$$P(A) = \frac{1}{2}$$

$$P(B) = \frac{1}{2}$$

$$P(A \cap B) = \frac{1}{4}$$

Now, according to our definition,

$$P(tail/head) = P(B/A) = \frac{P(A \cap B)}{P(B)}$$

$$= 1/2.$$

Notice how this is different from the probability of the event of getting a head before a tail in the two tosses.

Example 11. *The probability that an automobile being filled with gasoline will also need an oil change is 0.25, the probability that it needs a new oil filter is 0.40; and the probability that both the oil and filter need changing is 0.14.*

a) If the oil had to be changed, what is the probability that a new oil filter is needed?

b) If a new oil filter is needed, what is the probability that the oil has to be changed?

Solution:
Let A be the event that an automobile being filled with gasoline will also need an oil change and B be the event that it will need a new oil filter. Then,

$$P(A) = 0.25,$$

$$P(B) = 0.40,$$

$$P(A \cap B) = 0.14,$$

(i) *Probability that a new oil filter is needed given that oil had to be changed is given by*

$$\begin{aligned} P(B \mid A) &= \frac{P(A \cap B)}{P(A)} \\ &= \frac{0.14}{0.25} \\ &= \frac{14}{25}. \end{aligned}$$

(ii) *Probability that oil has to be changed given that a new filter is needed is given by*

$$\begin{aligned} P(A \mid B) &= \frac{P(A \cap B)}{P(B)} \\ &= \frac{0.14}{0.40} \\ &= \frac{7}{20}. \end{aligned}$$

Definition 4. (Independent events) *Two events A and B defined on a probability space (Ω, S, P) are said to be independent iff $P(A \cap B) = P(A)P(B)$.*

Remarks:

a) If $P(A) = 0$, then A is independent of any event $B \in S$.

b) Any event is always independent of the events Ω and $\emptyset$.

c) If A and B are independent and $P(A \cap B) = 0$, then either $P(A) = 0$ or $P(B) = 0$.

d) If $P(A) > 0$; $P(B) > 0$ and A, B are independent, then they are not mutually exclusive.

The reader should verify the above using the definition of independence.

Definition 5. (Pairwise independent) *Let U be a family of events from S. We say that the events in U are pairwise independent iff for every pair of distinct events $A, B \in U$, $P(A \cap B) = P(A)P(B)$.*

Definition 6. (Mutually Independent) *Let U be a family of events from S. The events in U are mutually independent iff for any finite sub collection $A_{i1}, A_{i2}, \ldots, A_{ik}$ of U, we have*

$$P\left(A_{i1} \cap A_{i2} \cap \ldots \cap A_{ik}\right) = \prod_k P(A_{ik}) \quad \forall \quad k.$$

Example 12. *X can solve 75% of the problems of a mathematics book while Y can solve 70% of the problems of the book. What is the chance that a problem selected at random will be solved when both X and Y try.*

Solution: *Let A and B be the events that X and Y can solve a problem respectively. Then, $P(A) = 0.75, P(B) = 0.70$. Since A and B are independent events, we have*

$$P(A \cap B) = P(A)P(B) = 0.75 \times 0.70.$$

Hence, the chance that problem selected at random will be solved when both X and Y try is obtained as

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.75 + 0.70 - 0.75 \times 0.70 = 0.925. \end{aligned}$$

Example 13. *Consider the random experiment of selecting a ball from an urn containing four balls numbered 1,2,3,4. Let $A = \{1, 2\}$, $B = \{1, 3\}$ and $C = \{1, 4\}$ be the events. Suppose that all the four outcomes are assumed equally likely. Prove that these events are pairwise independent but not mutually independent.*

Solution:

$A \cap B = \{1\} = A \cap C = B \cap C = A \cap B \cap C$

$P(A) = \frac{2}{4} = \frac{1}{2}, P(B) = \frac{2}{4} = \frac{1}{2}, P(C) = \frac{2}{4} = \frac{1}{2}$
$P(A \cap B) = \frac{1}{4}, P(A \cap C) = \frac{1}{4}, P(B \cap C) = \frac{1}{4}$
$P(A \cap B \cap C) = \frac{1}{4}$ *As we know, if*

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

then A and B are pairwise independent. So, we can see $A\&B, B\&C, A\&C$ are pairwise independent. To be mutually independent, we have to check

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

Here, $P(A)P(B)P(C) = \frac{1}{8} \neq P(A \cap B \cap C)$. Hence, A, B, C are pairwise independent but not mutually independent.

Total Probability Rule

Theorem 3. (Total probability theorem) *Let $B_1, B_2, \ldots, B_k$ be mutually disjoint events in the probability space (Ω, S, P) such that $\Omega = \bigcup_{i=1}^{k} B_i$ where $P(B_i) > 0$ for $i = 1, 2, \ldots, k$. Then for any $A \in S$ we have*

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

Remarks:

If $A, B \in S$, then $P(A) = P(A/B)P(B) + P(A/B^c)P(B^c)$ for any B∈S satisfying $0 < P(B) < 1$.

Theorem 4. (Multiplication Rule) *Let $A_1, A_2, \ldots, A_n$ be arbitrary events in the given Probability space (Ω,S,P) such that P($A_1 \cap A_2 \cap \ldots \cap A_{n-1}) > 0$, for any $n > 1$, then*

$$P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \ldots P(A_n/A_1 \cap A_2 \cap \ldots A_{n-1}). \qquad (1)$$

Example 14. *An urn contains 10 balls of which three are black and seven are white. At each trial, a ball is selected at random, its color is noted and it is replaced by two additional balls of the same color.*
(i) What is the probability that a white ball is selected in the second trial?
(ii) What is the probability that a black ball is selected in each of the first three trials?
Solution: *Let B_i denote the event that a black ball is selected in the i^{th} trial.*
(i) We need to find $P(B_1 \cap {B_2}^c) + P({B_1}^c \cap {B_2}^c)$.
(ii) We require to find $P(B_1 \cap B_2 \cap B_3)$. By multiplication rule (equation 1), we get

$$\begin{aligned} P(B_1 \cap B_2 \cap B_3) &= P(B_1)P(B_2/B_1)P(B_3/B_1 \cap B_2) \\ &= \frac{3}{10} \times \frac{(3+2)}{(10+2)} \times \frac{(3+2+2)}{(10+2+2)} \\ &= \frac{3}{10} \times \frac{5}{12} \times \frac{7}{14} = \frac{1}{16}. \end{aligned}$$

Similarly, the probability that a white ball is selected in each of the first three trial is

$$\frac{7}{10} \times \frac{9}{12} \times \frac{11}{14} = 0.4125.$$

Bayes' Theorem

[1] We state a very important result in conditional probability which has wide applications.

Theorem 5. Bayes' Rules or Bayes' Theorem
Let $B_1, B_2, \ldots$ be a collection of mutually disjoint events in the probability space (Ω, S, P) such that $\Omega = \cup_{i=1} B_i$ and $P(B_i) > 0$ for $i = 1, 2, \ldots$. Then for any $A \in S$ with $P(A) > 0$, we have

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_{l=1} P(A/B_l)P(B_l)}$$

Proof: *By the definition of conditional probability,*

$$\begin{aligned} P(B_i/A) &= \frac{P(B_i \cap A)}{P(A)} \\ P(A/B_i) &= \frac{P(B_i \cap A)}{P(B_i)} \end{aligned}$$

Combining above two equations, we get

$$P(B_i/A) = P(A/B_i)P(B_i)/P(A)$$

By the theorem of total probability

$$P(A) = \sum_{l=1} P(A/B_l)P(B_l)$$

Therefore,

$$P(B_i/A) = \frac{P(A/B_i)P(B_i)}{\sum_{l=1} P(A/B_l)P(B_l)}.$$

Example 15. *An insurance company insured 2000 scooter drivers, 4000 car drivers and 6000 truck drivers. The probabilites of an accident involving a scooter, car or truck are $0.01, 0.03, 0.15$ respectively. One of the insured persons meet with an accident. What is the probability that he is a scooterist.*

Solution: *Let A, B and C respectively be the events that an insured person owns a scooter, a car*

[1] Thomas Bayes, (1702-1761) was a British Mathematician known for having formulated a special case of Bayes' Theorem. Bayes' Theorem (also known as Bayes' rule or Bayes'law) is a result in probability theory, which relates the conditional and marginal probability of events. Bayes' theorem tells how to update or revise beliefs in light of new evidence: a posteriori.

or a truck; and E be the event that an accident has occurred. We have,

$$\begin{aligned} P(A) &= \frac{2000}{12000} = \frac{1}{6}, \\ P(B) &= \frac{4000}{12000} = \frac{1}{3}, \\ P(C) &= \frac{6000}{12000} = \frac{1}{2}, \\ \text{[Given]} \quad P(E \mid A) &= 0.01, \\ P(E \mid B) &= 0.03, \\ P(E \mid) &= 0.15, \end{aligned}$$

Thus the probability that the person who met with an accident is a scooter driver is

$$\begin{aligned} P(A \mid E) &= P(E \mid A)P(A)/((P(E \mid A)P(A)) + (P(E \mid B)P(B)) + (P(E \mid C)P(C)) \\ &= (.01 \times (\frac{1}{6}))/((.01 \times (\frac{1}{6})) + (.03 \times (\frac{1}{3})) + (.015 \times (\frac{1}{2}))) \\ &= \frac{2}{23}. \end{aligned}$$

Example 16. *Urn 1 contains one white and two black marbles, urn 2 contains one black and two white marbles, and urn 3 contains three black and three white marbles. A die is rolled. If a 1, 2, or 3 shows up, urn 1 is selected; if a 4 shows up, urn 2 is selected; and if a 5 or 6 shows up, urn 3 is selected. A marble is then drawn at random from the urn selected. Let* A *be the event that the marble drawn is white. If* U, V, W *respectively, denote the events that the urn selected is 1, 2, 3, then find* $P(A)$.

Solution: Observe that

$$A = (A \cap U) + (A \cap V) + (A \cap W),$$

Further, we have

$$\begin{aligned} P(A \cap U) &= P(U) \cdot P(A|U) = \frac{3}{6} \cdot \frac{1}{3} = \frac{1}{6}, \\ P(A \cap V) &= P(V) \cdot P(A|V) = \frac{1}{6} \cdot \frac{2}{3} = \frac{1}{9}, \\ P(A \cap W) &= P(W) \cdot P(A|W) = \frac{2}{6} \cdot \frac{3}{6} = \frac{1}{6}. \end{aligned}$$

It follows that

$$P(A) = P(A \cap U) + P(A \cap V) + P(A \cap W) = \frac{1}{6} + \frac{1}{9} + \frac{1}{6} = \frac{4}{9}.$$

Example 17. *In Example 16, let us compute the conditional probability* $P(V|A)$*.*

Solution: We have

$$P(V|A) = \frac{P(V \cap A)}{P(A)}.$$

Using the law of total probability,

$$P(V|A) = \frac{P(V \cap A)}{P(A)} = \frac{P(V) \cdot P(A|V)}{P(U) \cdot P(A|U) + P(V) \cdot P(A|V) + P(W) \cdot P(A|W)}.$$

Substituting the given probabilities, we get $\frac{1}{4}$.