Probability Theory (MA 590)

Class Notes
January – May, 2023

Instructor
Ayon Ganguly
Department of Mathematics
IIT Guwahati

Contents

Chapter 1

Probability

1.1 Probability

1.1.1 Classical Probability

As we know that the probability of an event A, denoted by $P(A)$, is defined by

$$P(A) = \frac{\text{Favourable number of cases to } A}{\text{Total number of cases}} = \frac{\#A}{\#S},$$

where S is the set of all possible outcomes. This definition is known as classical definition of probability. Note that this definition is only meaningful if number of elements in S is finite. As $A \subseteq S$, A is finite if S is finite. Let us consider the following examples.

Example 1.1. Let a fair die is rolled. What is the probability of getting three on upper face? It is easy to see that the required probability is 1/6. ||

Example 1.2. Consider a target comprising of three concentric circles of radii 1/3, 1, and $\sqrt{3}$ feet. Consider the event that a shooter hits inside the inner circle and its' probability. Let A be an event that the shooter hits inside the inner circle. Then

$$S = \mathbb{R}^2 \quad \text{and} \quad A = \left\{ x \in \mathbb{R}^2 : |x| \leq \frac{1}{3} \right\}.$$

In this case both A and S are infinite and therefore probability of A can not be found using classical definition of probability. ||

Here we will try to provided a general definition of probability such that we can apply the new definition for larger class of problems, like Example 1.2. Note that probability can be viewed as a function where the argument of the function is a set and output is a real number. To give the new definition of probability, we will use three basic properties (will be discussed) of the classical definition of probability, and we will say that a function which satisfy these three properties is called a probability or a probability function. Of course, we need to define the domain of the function properly.

Definition 1.1 (Set Function)**.** *A function which takes a set as its' argument is called a set function.*

1.1.2 Countable and Uncountable Sets

For further discussion, we need the concepts of countable and uncountable sets. The definitions and some properties of countable and uncountable sets are given in this subsection. You must have read these concepts in analysis course and therefore it is a recapitulation.

Definition 1.2. *We say that two sets A and B are equivalent if there exists a bijection from A to B. We denote it by $A \sim B$.*

Definition 1.3. *For any positive integer n, let $J_n = \{1, 2, \ldots, n\}$ and $\mathbb{N}$ be the set of all positive integers (natural numbers). For any set A, we say:*

(a) A is finite if $A = \phi$ or $A \sim J_n$ for some $n \in \mathbb{N}$. n is said to be the cardinality of A or number of elements in A.

(b) A is infinite if A is not finite.

(c) A is countable if $A \sim \mathbb{N}$.

(d) A is atmost countable if A is finite or countable.

(e) A is uncountable if A is neither finite nor countable.

Example 1.3. The set of all integers, $\mathbb{Z}$, is countable. Consider the function $f : \mathbb{N} \to \mathbb{Z}$ given by

$$f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ even} \\ -\frac{n-1}{2} & \text{if } n \text{ odd.} \end{cases}$$

It is easy to see that $f(\cdot)$ is a bijection from $\mathbb{N}$ to $\mathbb{Z}$. Therefore, $\mathbb{Z}$ is countable. ||

Remark 1.1. A finite set cannot be equivalent to any of its proper subset. However, this is possible for an infinite set. For example, consider the bijection $f : \mathbb{N} \to 2\mathbb{N}$ defined by

$$f(n) = 2n.$$

Here, $2\mathbb{N}$ is a proper subset of $\mathbb{N}$ and $f(\cdot)$ is a bijection from $\mathbb{N}$ to $2\mathbb{N}$. Therefore, $\mathbb{N}$ and $2\mathbb{N}$ are equivalent. †

Remark 1.2. If a set is countable, then it can be written as a sequence $\{x_n\}_{n\geq 1}$ of distinct terms. †

Theorem 1.1. *Every infinite subset of a countable set A is countable.*

Theorem 1.2. *Let $\{E_n\}_{n\geq 1}$ be a sequence of atmost countable sets and $S = \bigcup_{n=1}^{\infty} E_n$. Then S is atmost countable.*

Theorem 1.3. *Let $A_1, A_2, \ldots, A_n$ be atmost countable sets. Then $B_n = A_1 \times A_2 \times \ldots \times A_n$ is atmost countable.*

Corollary 1.1. *The set of rationals, $\mathbb{Q}$, is countable.*

Theorem 1.4. *The set, A, of all binary sequences is uncountable.*

Corollary 1.2. $[0, 1]$ *is uncountable.*

Corollary 1.3. $\mathbb{R}$ *is uncountable.*

Corollary 1.4. $\mathbb{Q}^c$ *is uncountable.*

Corollary 1.5. *Any interval is uncountable.*

1.1.3 Axiomatic Probability

To define the probability under more general framework, we need the concepts of random experiment, sample space, σ-field. These concepts are needed to define the domain of the probability function adequately.

Definition 1.4 (Random Experiment)**.** *An experiment is called a random experiment if it satisfies the following three properties:*

1. *All the outcomes of the experiment is known in advance.*
2. *The outcome of a particular performance of the experiment is not known in advance.*
3. *The experiment can be repeated under identical conditions.*

Note that according to the definition of a random experiment, we know all possible outcomes before hand, and hence we can make a list of all possible outcomes. This list is called sample space. The third condition in the definition of random sample is some what hypothetical in the sense that we will in general assume that the third condition is satisfied (if not very absurd to assume).

Definition 1.5 (Sample Space)**.** *The collection of all possible outcomes of a random experiment is called the sample space of the random experiment. It will be denoted by $\mathcal{S}$.*

Example 1.4. A toss of a coin is a random experiment as all the conditions of the definition of random experiment hold true. In this case, the sample space is $\mathcal{S} = \{H, T\}$. The sample space is finite in this example. ||

Example 1.5. Tossing a coin until the first head appears is also a random experiment with sample space $\mathcal{S} = \{H, TH, TTH, \ldots\}$. In this case, the sample space is countably infinite. ||

Example 1.6. The experiment of measuring the height of a student is a random experiment with sample space $\mathcal{S} = (0, \infty)$. Here the sample space is uncountable. ||

Definition 1.6 (σ-algebra or σ-field)**.** *A collection, $\mathcal{F}$, of subsets of $\mathcal{S}$ is called a σ-algebra or a σ-field if it satisfy the following properties:*

1. $\mathcal{S} \in \mathcal{F}$.
2. $A \in \mathcal{F}$ *implies* $A^c \in \mathcal{F}$.
3. $A_1, A_2, \ldots \in \mathcal{F}$ *implies* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

The first condition in the definition of σ-field implies that a σ-field is always non-empty. The second condition is called closed under complementation. Note that here $A^c = \mathcal{S} - A$, *i.e.*, the complementation is with respect to the sample space. The third condition of the definition of σ-field is called closed under countable union. Thus, by definition, a σ-field is closed under complementation and countable union.

Definition 1.7 (Event)**.** *A set E is said to be an event with respect to a σ-field $\mathcal{F}$ if $E \in \mathcal{F}$. We will say "the event E occurs" if the outcomes of a performance of the random experiment is in E.*

Example 1.7 (Continuation of Example 1.4)**.** Consider following three classes of subsets of $\mathcal{S}$. $\mathcal{F}_1 = \{\emptyset, \mathcal{S}, \{H\}, \{T\}\}$, $\mathcal{F}_2 = \{\emptyset, \mathcal{S}\}$, and $\mathcal{F}_3 = \{\emptyset, \mathcal{S}, \{H\}\}$. Here we will show that $\mathcal{F}_1$ and $\mathcal{F}_2$ are σ fields, but $\mathcal{F}_3$ is not a σ-field.

Note that $\mathcal{S} \in \mathcal{F}_1$ and for any $A \in \mathcal{F}_1$, $A^c \in \mathcal{F}_1$. Hence, it is easy to see that the first two conditions of the definition of σ-field are hold true. For the third condition, let $A_1, A_2, \ldots \in \mathcal{F}_1$.
Case I: If $A_i = \mathcal{S}$ for at least one $i \in \mathbb{N}$, $\bigcup_{i=1}^{\infty} A_i = \mathcal{S} \in \mathcal{F}_1$.
Case II: If $A_i = \emptyset$ for all $i \in \mathbb{N}$, $\bigcup_{i=1}^{\infty} A_i = \emptyset \in \mathcal{F}_1$.
Case III: If $A_i = \{H\}$ for at least one $i \in \mathbb{N}$ and rest of $A_i = \emptyset$, $\bigcup_{i=1}^{\infty} A_i = \{H\} \in \mathcal{F}_1$.
Case IV: If $A_i = \{T\}$ for at least one $i \in \mathbb{N}$ and rest of $A_i = \emptyset$, $\bigcup_{i=1}^{\infty} A_i = \{T\} \in \mathcal{F}_1$.
Case V: If $A_i = \{H\}$ for at least one $i \in \mathbb{N}$ and $A_j = \{T\}$ for at least one $j \in \mathbb{N}$, $\bigcup_{i=1}^{\infty} A_i = \mathcal{S} \in \mathcal{F}_1$.
These are the exhaustive cases and in all the cases, $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}_1$. Hence, $\mathcal{F}_1$ is a σ-field on the subsets of $\mathcal{S}$. Note that $\mathcal{F}_1$ is power set of $\mathcal{S}$.

It is easy to see that $\mathcal{F}_2$ is σ-field and therefore left as a practice problem. To show that $\mathcal{F}_3$ is not a σ-field, we need to show that at least one of the three conditions is not true. It is very easy to check that the second condition is not true as $\{H\} \in \mathcal{F}_3$, but $\{H\}^c = \{T\} \notin \mathcal{F}_3$. ||

Example 1.8 (Continuation of Example 1.5)**.** Consider $\mathcal{F} = \mathcal{P}(\mathcal{S})$, the power set of $\mathcal{S}$. Clearly, $\mathcal{S} \in \mathcal{F}$. For any $A \in \mathcal{F}$, A^c is a subset of $\mathcal{S}$ and hence belongs to $\mathcal{F}$. For any countable collection of sets $A_1, A_2, \ldots \in \mathcal{F}$, $\bigcup_{i=1}^{\infty} A_i$ is a subset of $\mathcal{S}$ and belongs to $\mathcal{F}$. Hence, $\mathcal{F}$ is a σ-field on the subsets of $\mathcal{S}$. ||

Example 1.9 (Continuation of Example 1.6)**.** $\mathcal{F} = \{\phi, \mathcal{S}, (4, 5), (4, 5)^c\}$ is a σ-field. ||

Remark 1.3. Note that there could be multiple σ-field on subsets of a sample space. Power set of sample space is always a σ-field and it is the largest σ-field. On the other hand $\{\mathcal{S}, \emptyset\}$ is also a σ-field and it is the smallest σ-field. †

Definition 1.8 (Measurable Space)**.** *Let $\mathcal{S}$ be a sample space of a random experiment and $\mathcal{F}$ is a σ-field on subsets of $\mathcal{S}$. Then the ordered pair $(\mathcal{S}, \mathcal{F})$ is called a measurable space.*

Definition 1.9 (Probability)**.** *Let $(\mathcal{S}, \mathcal{F})$ be a measurable space. A set function $P : \mathcal{F} \to \mathbb{R}$ is called a probability if*

1. *$P(E) \geq 0$ for all $E \in \mathcal{F}$.*
2. *$P(\mathcal{S}) = 1$.*
3. *(Countable Additivity) Let $E_1, E_2, \ldots \in \mathcal{F}$ be a sequence of disjoint events (i.e., $E_i \cap E_j = \emptyset$ for all $i \neq j \in \mathbb{N}$) then*

$$P\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} P(E_i).$$

The idea of probability germinates to predict the outcomes of gambling, where the classical definition of probability was used. When the people tried to give the axiomatic definition of the probability, it was observed that these three properties (mentioned in Definition 1.9) of the classical definition of probability are working fine. Hence, these three properties are used. Note that the first and the third axioms (mentioned in Definition 1.9) have good

intuitions from the concepts of area of a region or volume of a shape. The area or volume is always non-negative and if we have several disjoint reasons or shapes, then the combined area or volume is the sum of the individual areas or volumes, respectively.

Definition 1.10 (Probability Space)**.** *Let $\mathcal{S}$ be a sample space and $\mathcal{F}$ be a σ-field on the subsets of $\mathcal{S}$. Let P be a probability defined on $\mathcal{F}$. The triplet $(\mathcal{S}, \mathcal{F}, P)$ is called a probability space.*

Example 1.10 (Continuation of Example 1.4)**.** Consider the random experiment of tossing of a coin, where sample space is $\mathcal{S} = \{H, T\}$ and $\mathcal{F} = \mathcal{P}(\mathcal{S})$, the power set of $\mathcal{S}$. Consider a function $P : \mathcal{F} \to \mathbb{R}$ defined by

$$P(\mathcal{S}) = 1,\ P(\{H\}) = 0.6,\ P(\{T\}) = 0.4, \text{ and } P(\emptyset) = 0.$$

Here it is very easy to see that the first two axioms of Definition 1.9 are hold true. To check if the third axiom hold or not, let us consider the following cases. Note that here we have to choose E_i's such that E_i are disjoint.
Case I: $E_i = \emptyset$ for all $i \in \mathbb{N}$. Then $P(E_i) = 0$ for all $i \in \mathbb{N}$ implies $\sum_{i=1}^{\infty} P(E_i) = 0$. On the other hand, $P(\cup_{i=1}^{\infty} E_i) = P(\emptyset) = 0$.
Case II: $E_i = \mathcal{S}$ if $i = i_0$ for some $i_0 \in \mathbb{N}$ and $E_i = \emptyset$ for $i \neq i_0$. In this case $\sum_{i=1}^{\infty} P(E_i) = 1$ and $P(\cup_{i=1}^{\infty} E_i) = P(\mathcal{S}) = 1$.
Case III: $E_i = \{H\}$ if $i = i_0$ for some $i_0 \in \mathbb{N}$ and $E_i = \emptyset$ for $i \neq i_0$. In this case $\sum_{i=1}^{\infty} P(E_i) = P(\{H\}) = 0.6$ and $P(\cup_{i=1}^{\infty} E_i) = P(\{H\}) = 0.6$.
Case IV: $E_i = \{T\}$ if $i = i_0$ for some $i_0 \in \mathbb{N}$ and $E_i = \emptyset$ for $i \neq i_0$. In this case $\sum_{i=1}^{\infty} P(E_i) = P(\{T\}) = 0.4$ and $P(\cup_{i=1}^{\infty} E_i) = P(\{T\}) = 0.4$.
Case V: $E_i = \{T\}$ if $i = i_1$, $E_i = \{H\}$ if $i = i_2$ for some $i_1 \neq i_2 \in \mathbb{N}$, and $E_i = \emptyset$ for $i \neq i_1, i_2$. In this case $\sum_{i=1}^{\infty} P(E_i) = P(\{H\}) + P(\{T\}) = 1$ and $P(\cup_{i=1}^{\infty} E_i) = P(\mathcal{S}) = 1$. ||

Example 1.11. Consider a roll of a die. The sample space $\mathcal{S} = \{1, 2, \ldots, 6\}$ and take $\mathcal{F} = \mathcal{P}(\mathcal{S})$. Let $P(\emptyset) = 0$ and $P(i) = 1/6$ for $i \in \mathcal{S}$. Note that in this case the function $P(\cdot)$ have not defined for all the members in $\mathcal{F}$. However, if we assume that $P(\cdot)$ is a probability defined on the σ-field $\mathcal{F}$, we can uniquely extend $P(\cdot)$ for all other members of $\mathcal{F}$. Let $E \in \mathcal{F}$. As $\mathcal{S}$ is a finite set, so is E. Let the cardinality of E is n and the elements of E be $x_1 < x_2 < \ldots < x_n$. Define $E_i = \{x_i\}$ for $i = 1, 2, \ldots, n$ and $E_i = \emptyset$ for $i > n$. Clearly, E_i's are disjoint and $E = \cup_{i=1}^{\infty} E_i$. Now, if $P(\cdot)$ is a probability, using the third axiom of Definition 1.9, $P(E) = P(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i) = n/6$. ||

Example 1.12. Consider a roll of a die. The sample space $\mathcal{S} = \{1, 2, \ldots, 6\}$ and take $\mathcal{F} = \mathcal{P}(\mathcal{S})$. Let $P(\emptyset) = 0$ and $P(i) = i/21$ for $i \in \mathcal{S}$. As before, in this case also we can extend the function $P(\cdot)$ on $\mathcal{F}$ such that it becomes a probability on $\mathcal{F}$. ||

We have already pointed out that $\mathcal{P}(\mathcal{S})$ is a σ-field. Now, a natural question is that if $\mathcal{P}(\mathcal{S})$ is σ-field, then why do we define σ-field? Why should not we work with power set of sample space, always, and define probability on the power set of the sample space? We will try to answer these questions using next two examples. We will show that the choice of σ-field is an important issue.

Example 1.13. Let $\mathcal{S} = \{1, 2, \ldots, 60\}$ and $\mathcal{F} = \mathcal{P}(\mathcal{S})$. Let us define $P(E) = \frac{\#E}{\#\mathcal{S}}$ for all $E \in \mathcal{F}$. Note that as $\mathcal{S}$ is a finite set, $P(\cdot)$ satisfies all the axioms of probability. ||

Example 1.14. Now, consider a different problem where $\mathcal{S} = \mathbb{N}$ and $\mathcal{F} = \mathcal{P}(\mathbb{N})$. Can we extend the definition of probability in the previous example to define a probability for this example? A natural extension is

$$P(E) = \limsup_{n\to\infty} \frac{N_n(E)}{n}, \tag{1.1}$$

where $E \in \mathcal{F}$ and $N_n(E)$ is the number of times E occurs in the first n natural numbers. Here we have used lim sup instead of lim to overcome the issue of existence of limit of $\frac{N_n(E)}{n}$. Before answering if $P(\cdot)$ defined above is a probability, let us see the values of $P(\cdot)$ evaluated on some specified subsets of $\mathcal{S}$. Let us consider $A = \{\omega \in \mathbb{N} : \omega \text{ is a multiple of } 3\}$ and we want to calculate $P(A)$. Note that $N_n(A)$ is the number of multiple of three in the set $J_n = \{1, 2, \ldots, n\}$. Thus,

$$\frac{N_n(A)}{n} = \begin{cases} \frac{m}{3m} & \text{if } n = 3m \\ \frac{m}{3m+1} & \text{if } n = 3m+1 \\ \frac{m}{3m+2} & \text{if } n = 3m+2. \end{cases}$$

Hence, for all $n \in \mathbb{N}$, $\dfrac{1}{3 + \frac{6}{n-2}} \leq \dfrac{N_n(A)}{n} \leq \dfrac{1}{3}$ which implies $P(A) = \frac{1}{3}$. Similarly, $P(B) = \dfrac{1}{4}$ *(why?)* for $B = \{\omega \in \mathbb{N} : \omega \text{ is a multiple of } 4\}$. Now, assume that $C = \{2\}$. Then

$$\frac{N_n(C)}{n} = \begin{cases} 0 & \text{if } n = 1 \\ \frac{1}{n} & \text{if } n \geq 2. \end{cases}$$

Hence, $P(C) = 0$. Similarly, $P(D) = 0$ for any singleton set D. However, $\mathcal{S} = \mathbb{N} = \cup_{i\in\mathbb{N}}\{i\}$. Hence, if P satisfies the third axiom then $P(\mathcal{S}) = \sum_{i=1}^{\infty} P(\{i\}) = 0 \neq 1$, which contradicts the second axiom. Though $P(\cdot)$ as defined in (1.1) gives meaningful values for some sets like A and B, it does not satisfy all the three axioms, when it is defined on the power set of $\mathcal{S}$. ||

Note that we can always define a probability on the power set of a sample space. For example, let $\omega_0 \in \mathcal{S}$ be a fixed element. Define $P : \mathcal{P}(\mathcal{S}) \to \mathbb{R}$ by

$$P(A) = \begin{cases} 1 & \text{if } \omega_0 \in A \\ 0 & \text{if } \omega_0 \notin A. \end{cases}$$

It is easy to see that $P(\cdot)$ is a probability. However, in practice, a probability is used to model a practical situation, where the probability may need to satisfy extra conditions other then three conditions mentioned in Definition 1.9. The previous example suggests, depending on our objective we may need to choose from the set of all subsets of $\mathcal{S}$, certain subsets (not all) of $\mathcal{S}$ on which to define a probability P. For example, $P(\cdot)$ defined in (1.1) becomes a probability on the σ-fields $\mathcal{F}_1 = \{\mathcal{S}, \emptyset, A, A^c\}$ or $\mathcal{F}_2 = \{\mathcal{S}, \emptyset, A, B, A^c, B^c, A \cap B, A^c \cap B, A \cap B^c, A^c \cap B^c, A \cup B, A \cup B^c, A^c \cup B, A^c \cup B^c, (A \cap B) \cap (A^c \cap B^c), (A^c \cap B) \cup (A \cap B^c)\}$, where A and B are as defined in the previous example.

Next we will see some of the properties of probability. Let us assume that $(\mathcal{S}, \mathcal{F}, P)$ be a probability space.

Theorem 1.5. $P(\emptyset) = 0$.

Proof: Consider $E_i = \emptyset$ for all $i \in \mathbb{N}$. Clearly, E_i's are disjoint and $\cup_{i=1}^{\infty} E_i = \emptyset$. Using the third axiom of Definition 1.9,

$$P(\emptyset) + P(\emptyset) + \ldots = P(\emptyset) \implies P(\emptyset) = 0,$$

as using first axiom $P(\emptyset) \geq 0$. □

Theorem 1.6 (Finite Additivity). *If $E_1, E_2, \ldots, E_n$ are n disjoint events, then*

$$P\left(\bigcup_{i=1}^{n} E_i\right) = \sum_{i=1}^{n} P(E_i).$$

Proof: Take $E_i = \emptyset$ for $i > n$ in the third axiom, to get the required result. □

Theorem 1.7 (Mononomicity). *$P(\cdot)$ is monotone, i.e., for $E_1, E_2 \in \mathcal{F}$ and $E_1 \subset E_2$, $P(E_1) \leq P(E_2)$.*

Proof: Note that $E_2 = (E_2 \cap E_1) \cup (E_2 \cap E_1^c)$ with $(E_2 \cap E_1) \cap (E_2 \cap E_1^c) = \emptyset$. Hence, using finite additivity, $P(E_2) = P(E_2 \cap E_1) + P(E_2 \cap E_1^c) = P(E_1) + P(E_2 \cap E_1) \geq P(E_1)$. Here the second equality is true as $E_1 \subset E_2$ and the last inequality is true as $P(\cdot) \geq 0$. □

The first and second terms in the right hand side of the decomposition $E_2 = (E_2 \cap E_1) \cup (E_2 \cap E_2^c)$ can be interpreted as E_2 occurring with E_1 and E_2 occurring without E_1, respectively. This decomposition is quite useful. We can use it to solve several problems in this course.

Theorem 1.8. *Let $A, B \in \mathcal{F}$ such that $P(B) = 0$. Then $P(A \cap B) = 0$.*

Proof: Note that $0 = P(B) \geq P(A \cap B) \geq 0$. Hence $P(A \cap B) = 0$. □

Corollary 1.6. *Let $A, B \in \mathcal{F}$ with $P(B) = 1$. Then $P(A \cap B) = P(A)$.*

Proof: As $P(B^c) = 0$, $P(A \cap B^c) = 0 \implies P(A \cap B) = P(A)$. □

Theorem 1.9 (Subtractive Property). *$P(\cdot)$ is subtractive, i.e., for $E_1, E_2 \in \mathcal{F}$ and $E_1 \subset E_2$, $P(E_2 \setminus E_1) = P(E_2) - P(E_1)$.*

Proof: As in the proof of Theorem 1.7,

$$P(E_2) = P(E_1) + P(E_2 \cap E_1^c) \implies P(E_2 \setminus E_2) = P(E_2) - P(E_1).$$

□

Theorem 1.10. $0 \leq P(E) \leq 1$.

Proof: For any $E \in \mathcal{F}$, $\emptyset \subset E \subset \mathcal{S} \implies 0 \leq P(E) \leq 1$, using Theorem 1.7. □

Theorem 1.11. *If $E_1, E_2 \in \mathcal{F}$, then $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$.*

Proof: Note that $E_1 \cup E_2 = E_2 \cup (E_1 \setminus E_2)$. Also, $E_2 \cap (E_1 \setminus E_2) = \emptyset$. Hence, $P(E_1 \cup E_2) = P(E_2) + P(E_1 \setminus E_2)$. We have already seen that $P(E_1) = P(E_1 \cap E_2) + P(E_1 \setminus E_2)$. Combining, we get the required result. □

Theorem 1.12. *If $E_1, E_2 \in \mathcal{F}$, then $P(E_1 \cup E_2) \leq P(E_1) + P(E_2)$.*

Proof: Trivial from the last theorem. □

Theorem 1.13. *If $E \in \mathcal{F}$, then $P(E^c) = 1 - P(E)$.*

Proof: This is trivial as $\mathcal{S} = E \cup E^c$. □

Definition 1.11 (Elementary Event)**.** *A single-ton event is called an elementary event.*

If $\mathcal{S}$ is finite, and $\mathcal{F} = \mathcal{P}(\mathcal{S})$, it is sufficient to assign probability to each elementary event in the sense that for any subset E of $\mathcal{S}$, we can calculate $P(E)$. For any $E \in \mathcal{F}$, E can be written as the union of elementary events that are in E, *i.e.*, $E = \cup_{\omega \in E} \{\omega\}$. Note that as E is finite (being a subset of $\mathcal{S}$, which is finite), there are finite number of elementary events in the expression. Also, elementary events are disjoint. Hence, $P(E) = \sum_{\omega \in E} P(\{\omega\})$.

Let $\mathcal{S}$ be finite, $\mathcal{F} = \mathcal{P}(\mathcal{S})$ and the elementary events be equally likely (*i.e.*, all the elementary events have same probability). Let the cardinality of the set $\mathcal{S}$ is n. Then $\mathcal{S}$ can be written as $\{\omega_1, \omega_2, \ldots, \omega_n\}$. Let $P(\{\omega_i\}) = c$ for all $i = 1, 2, \ldots, n$. Note that $\mathcal{S} = \cup_{i=1}^{n} \{\omega_i\}$ implies that $c = 1/n$. Now, for any event E, $P(E) = \frac{\#E}{n}$, which is classical probability. Hence, classical definition of probability is a particular case of the axiomatic definition of probability.

If $\mathcal{S}$ is countably infinite, and $\mathcal{F} = \mathcal{P}(\mathcal{S})$, it is still sufficient to assign probability to each elementary event. For any $E \in \mathcal{F}$, E is atmost countable, which means that E can be expressed as countable union of elementary events. Therefore, $P(E) = \sum_{\omega \in E} P(\{\omega\})$. However, in this case one cannot assign equal probability to each elementary event without violating the second axiom in Definition 1.9. Revisit Examples 1.11 and 1.12.

If $\mathcal{S}$ is uncountable, and $\mathcal{F} = \mathcal{P}(\mathcal{S})$, one can not make an equally likely assignment of positive probabilities to each elementary event. We can prove this statement by contradiction. If possible, suppose that an equally likely assignment of positive probability can be done, *i.e.*, $P(\{\omega\}) = c > 0$ for all $\omega \in \mathcal{S}$. $\mathcal{S}$ can be written as union of elementary events. However there are uncountable elementary events, and hence it is an uncountable union of sets. Therefore, the third axiom of probability can not be used directly to conclude that $P(\mathcal{S}) > 1$. Now, note that there exists a countable subset E of $\mathcal{S}$. Clearly, $P(\mathcal{S}) \geq P(E) = \infty$. This is a contradiction to the second axiom of Definition 1.9. Hence, our assumption is wrong.

Indeed, for uncountable $\mathcal{S}$ and $\mathcal{F} = \mathcal{P}(\mathcal{S})$, one can not assign positive probability to each elementary event without violating the axiom $P(\mathcal{S}) = 1$. This statement can be proved, again, by contradiction. If possible, suppose that $P(\{\omega\}) > 0$ for all $\omega \in \mathcal{S}$. Let us define the sets $A_n = \left\{\omega \in \mathcal{S} : P(\{\omega\}) > \frac{1}{n}\right\}$ for $n = 1, 2, \ldots$. The claim is that A_n is finite set for all $n = 1, 2, \ldots$. If not, then A_n is either countably infinite or uncountable. In both the cases, $P(A_n)$ is infinite, which is a contradiction. Hence, A_n is finite. Now, note that $\mathcal{S} = \cup_{n=1}^{\infty} A_n$. As A_n are finite, $\mathcal{S}$ is atmost countable, which is a contradiction and therefore our assumption that $P(\{\omega\}) > 0$ for all $\omega \in \mathcal{S}$ is wrong.

1.1.4 Continuity of Probability

Note that a function $f : \mathbb{R} \to \mathbb{R}$ is said to be continuous at x_0 if for every real sequence $\{x_n\}_{n \geq 1}$ converging to x_0, the sequence $\{f(x_n)\}_{n \geq 1}$ converges to $f(x_0)$. If we want to extend this definition of continuity of a function $f : \mathbb{R} \to \mathbb{R}$ to probability, first the convergence of sequence of events need to be defined, as the argument of $P(\cdot)$ is an event. Here we will consider the limits of increasing and decreasing sequences of events.

Definition 1.12 (Increasing Sequence of Events)**.** *A sequence, $\{E_n\}_{n \geq 1}$, of events are said to be increasing if $E_n \subseteq E_{n+1}$ for all $n = 1, 2, \ldots$.*

Definition 1.13 (Decreasing Sequence of Events). *A sequence, $\{E_n\}_{n\geq 1}$, of events are said to be decreasing if $E_{n+1} \subseteq E_n$ for all $n = 1, 2, \ldots$.*

Definition 1.14 (Limit of Increasing Sequence of Events). *For an increasing sequence, $\{E_n\}_{n\geq 1}$, of events, the limit is defined by $\lim_{n\to\infty} E_n = \bigcup_{n=1}^{\infty} E_n$.*

Definition 1.15 (Limit of Decreasing Sequence of Events). *For a decreasing sequence, $\{E_n\}_{n\geq 1}$, of events, the limit is defined by $\lim_{n\to\infty} E_n = \bigcap_{n=1}^{\infty} E_n$.*

Theorem 1.14 (Continuity from below). *Let $\{E_n\}_{n\geq 1}$ be an increasing sequence of events, then*

$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n).$$

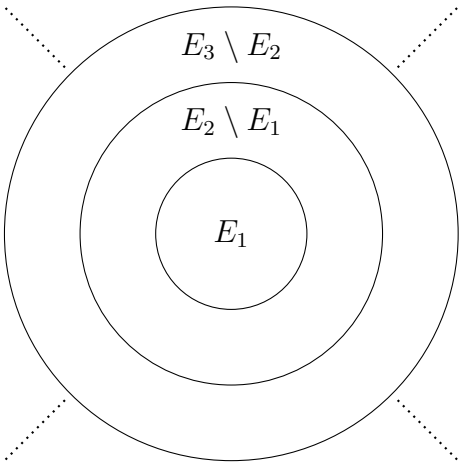


Figure 1.1: Sequence of events $\{A_n\}_{n\geq 1}$.

Proof: Let us define the following sequence, $\{A_n\}_{n\geq 1}$, of events as

$$A_1 = E_1 \text{ and } A_n = E_n \setminus E_{n-1} \text{ for } n = 2, 3, \ldots.$$

Please see the Figure 1.1. Clearly, A_n's are disjoint and $\bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} E_n$. Also,

$$P(A_n) = \begin{cases} P(E_1) & \text{if } n = 1 \\ P(E_n) - P(E_{n-1}) & \text{if } n = 2, 3, \ldots, \end{cases}$$

as $\{E_n\}_{n\geq 1}$ is an increasing sequence of events. Now,

$$P\left(\lim_{n\to\infty} E_n\right) = P\left(\bigcup_{n=1}^{\infty} E_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_i) = \lim_{N\to\infty} \sum_{n=1}^{N} P(A_n) = \lim_{n\to\infty} P(E_n).$$

Here the third equality is true for the third axiom of Definition 1.9, the fourth equality is true from the definition of a convergence of series, and the last equality is true for telescopic series. □

Theorem 1.15 (Continuity from above)**.** *Let* $\{E_n\}_{n\geq 1}$ *be a decreasing sequence of events, then*

$$P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n).$$

Proof: Let $A_n = E_n^c$ for all $n = 1, 2, \ldots$. Clearly, $\{A_n\}_{n\geq 1}$ is an increasing sequence of events. Hence,

$$\begin{aligned}
&P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \\
\implies &P\left(\bigcup_{n=1}^{\infty} E_n^c\right) = \lim_{n\to\infty} P\left(E_n^c\right) \\
\implies &P\left(\left(\bigcap_{n=1}^{\infty} E_n\right)^c\right) = \lim_{n\to\infty} \left(1 - P\left(E_n\right)\right) \\
\implies &P\left(\lim_{n\to\infty} E_n\right) = \lim_{n\to\infty} P(E_n).
\end{aligned}$$

□

1.2 Conditional Probability

We use conditional probability when we have some information about the outcome of a random experiment. Let us consider the following example.

Example 1.15. Let a die is thrown twice. Suppose that we are interested in the probability of the event that the sum of the outcomes of the rolls is six. Clearly, the sample space has 36 points and

$$\mathcal{S} = \{(n, m) : n, m = 1, 2, \ldots, 6\}.$$

As the sample space is finite, let us use the classical definition of probability. The required probability is 5/36.

Now, assume that you have observed that the first throw results in a 4. We are interested in the probability of the same event as before, but now we have extra information that the outcome of the first roll is 4. Note that when we know that the first roll results in a 4, the sample space changes and the new sample space is

$$\mathcal{S}_1 = \{(4, m) : m = 1, 2, \ldots, 6\},$$

which is the event that the first throw is a 4. We need to find the probability of the event that the sum is 6 in the sample space $\mathcal{S}_1$. There is only one case (4, 2) (in $\mathcal{S}_1$) which is favorable to the event of interest and hence the required probability is

$$\frac{1}{6} = \frac{1/36}{6/36} = \frac{P(A \cap H)}{P(H)},$$

where A and H are the events that sum is 6 and first roll results in 4, respectively. ||

Once you are given some information or you observe something, the sample space changes. Conditional probability is a probability on the changed sample space. Motivated by the above example, the definition of conditional probability is given as follows.

Definition 1.16 (Conditional Probability)**.** *Let H be an event with $P(H) > 0$. For any arbitrary event A, the conditional probability of A given H is denoted by $P(A|H)$ and defined by*

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

Note that to define the conditional probability, the probability of the conditioning event has to be positive. The probability of the intersection of two events can be expressed in terms of the conditional probability and the relationship is given below.

$$P(A \cap B) = \begin{cases} P(A)P(B|A) & \text{if } P(A) > 0 \\ P(B)P(A|B) & \text{if } P(B) > 0. \end{cases}$$

Definition 1.17 (Mutually Exclusive Events)**.** *A collection of events $\{E_1, E_2 \ldots\}$ is said to be mutually exclusive if $E_i \cap E_j = \emptyset$ for all $i \neq j$.*

Definition 1.18 (Exhaustive Events)**.** *A collection of events $\{E_1, E_2 \ldots\}$ is said to be exhaustive if $P\left(\cup_{i=1}^{\infty} E_i\right) = 1$.*

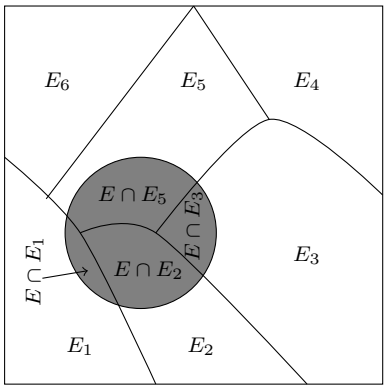
Thus if a collection of events $\{E_n\}_{n\geq 1}$ is such that $\cup_{n=1}^{\infty} E_n = \mathcal{S}$, then the collection is exhaustive.

Theorem 1.16 (Theorem of Total Probability)**.** *Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0$ for all $i = 1, 2, \ldots$. Then for any event E,*

$$P(E) = \sum_{i=1}^{\infty} P(E|E_i)P(E_i)\,.$$

Proof: Let us denote $\tilde{E}_i = E_i \cap E$ for all $i = 1, 2, \ldots$. Then $\left\{\tilde{E}_1, \tilde{E}_2, \ldots\right\}$ is mutually exclusive. Now, as E_i's are exhaustive using Corollary 1.6

$$\begin{aligned} P(E) &= P\left(E \cap \left(\bigcup_{i=1}^{\infty} E_i\right)\right) \\ &= P\left(\bigcup_{i=1}^{\infty} (E \cap E_i)\right) \\ &= \sum_{i=1}^{\infty} P(E \cap E_i)\,, \text{ as } \tilde{E}_i \text{ are mutually exclusive} \\ &= \sum_{i=1}^{\infty} P(E|E_i)P(E_i). \text{ as } P(E_i) > 0 \text{ for all } i \in \mathbb{N}. \end{aligned}$$



□

The theorem of total probability tells that probability of an event can be computed by computing the probability of several partitions of the event. See the above figure, where the square and the shaded region indicates the sample space and the event E, respectively. In the figure $\{E_1, E_2, E_3, E_4, E_5, E_6\}$ is mutually exclusive and exhaustive. The event E can be partitioned into $E \cap E_1$, $E \cap E_2$, $E \cap E_3$, and $E \cap E_5$ and hence the probability of E can be computed by computing the probability of several partitions and then adding these probabilities.

Theorem 1.17 (Bayes' Theorem)**.** *Let $\{E_1, E_2 \ldots\}$ be a collection of mutually exclusive and exhaustive events with $P(E_i) > 0$ for all $i = 1, 2, \ldots$. Let E be any event with $P(E) > 0$. Then*

$$P(E_i|E) = \frac{P(E|E_i)P(E_i)}{\sum_{j=1}^{\infty} P(E|E_j)P(E_j)} \quad \text{for } i = 1, 2, \ldots.$$

Proof: Using the definition of conditional probability and the theorem of total probability, the proof is straight forward. □

In the theorem of total probability and Bayes' theorem, we have considered a countable collection of events $\{E_1, E_2, \ldots\}$. However, the theorems hold true even if we have a finite collection of mutually exclusive and exhaustive events *(Why?)*.

Example 1.16. There are 3 boxes. Box 1 containing 1 white, 4 black balls. Box 2 containing 2 white, 1 black ball. Box 3 containing 3 white, 3 black balls. First you throw a fair die. If the outcomes are 1, 2 or 3 then box 1 is chosen, if the outcome is 4 then box 2 is chosen and if the outcome is 5 or 6 then box 3 is chosen. Finally, you draw a ball at random from the chosen box. Let W denote the event that the drawn ball is white. Also, assume that B_i, $i = 1, 2, 3$, denotes the event that ith box is selected after the roll of the die. Using Bayes' theorem, the (conditional) probability that the ball is from box 1 given the chosen ball is white is

$$P(B_1|W) = \frac{P(W|B_1)P(B_1)}{\sum_{i=1}^{3} P(W|B_i)P(B_i)} = \frac{9}{34}.$$

Similarly given the fact that the drawn ball is white, the probability that the ball is from box 2 is $P(B_2|W) = 5/17$. ||

1.3 Independence

Observe in the previous example that $P(B_1|W) = 9/34 < 1/2 = P(B_1)$, whereas $P(B_2|W) = 5/17 > 1/6 = P(B_2)$. Thus the occurrence of one event can make the occurrence of a second event more or less likely. Also, occurrence of an event may not change the probability of the occurrence of a second event. For example, let a coin is tossed two times. Then the probability of a head in the second toss does not change if the result of the first toss is a tail.

When occurrence of one event, say A, reduce the probability of the occurrence of another event, say B, we say that the events are negatively associated. That means A and B are negatively associated if $P(B|A) < P(B)$. For the conditional probability $P(B|A)$, $P(A)$ must be strictly greater than zero. Now, note that $P(B|A) < P(B)$ can be equivalently written as $P(A \cap B) < P(A)P(B)$, where we do not need the restriction $P(A) > 0$. Motivated by this discussion, we have the following definition.

Definition 1.19. *Let A and B be two events. They are said to be*

1. *negatively associated if $P(A \cap B) < P(A)P(B)$.*
2. *positively associated if $P(A \cap B) > P(A)P(B)$.*
3. *independent if $P(A \cap B) = P(A)P(B)$.*

Theorem 1.18. *If A and B are independent, so are A and B^c.*

Proof: As $A = (A \cap B) \cup (A \cap B^c)$, where $A \cap B$ and $A \cap B^c$ are disjoint. Hence

$$\begin{aligned} P(A \cap B^c) &= P(A) - P(A \cap B) \\ &= P(A) - P(A)\,P(B), \text{ as } A \text{ and } B \text{ are independent events} \\ &= P(A)P(B^c). \end{aligned}$$

Hence, A and B^c are independent events. □

Corollary 1.7. *If A and B are independent events, then*

1. *A^c and B are independent events.*
2. *A^c and B^c are independent events.*

Proof: This proof is simple using the previous theorem, and hence left as an exercise. □

Example 1.17. Let $P(B) = 0$. For any event A, $0 \leq P(A \cap B) \leq P(B) = 0$. Hence, $P(A \cap B) = 0$. On the other hand, $P(A)P(B) = 0$. Therefore, A and B are independent.

Now, assume that $P(B) = 1$. Then for any event A, A and B^c are independent as $P(B^c) = 0$. Using the previous theorem, A and B are independent events. In particular any event A is independent of $\mathcal{S}$ and $\emptyset$. ||

We have talked about independence of two events. A natural question is: Is the concept of independence be extended for more than two events? The answer is yes. However, there are two types of independence that are of interest for more than two events. We will discuss these concepts now.

Definition 1.20 (Pairwise Independent)**.** *A countable collection of events $E_1, E_2, \ldots$ are said to be pairwise independent if E_i and E_j are independent for all $i \neq j$.*

Definition 1.21 (Independent for Finite Collection of Events)**.** *A finite collection of events E_1, E_2, …, E_n are said to be independent (or mutually independent) if for any sub-collection E_{n_1}, …, E_{n_k} of $E_1, E_2, \ldots, E_n$,*

$$P\left(\bigcap_{i=1}^{k} E_{n_k}\right) = \prod_{i=1}^{k} P(E_{n_i}).$$

Definition 1.22 (Independent for Countable Collection of Events)**.** *A countable collection of events $E_1, E_2, \ldots$ are said to be independent (or mutually independent) if any finite sub-collection is independent.*

Suppose that we have three events E_1, E_2, and E_3 and we want to check if they are pairwise independent or not. We need to verify three conditions, *viz.*,

$$\begin{aligned} P(E_1 \cap E_2) &= P(E_1)P(E_2), \\ P(E_1 \cap E_3) &= P(E_1)P(E_3), \\ P(E_2 \cap E_3) &= P(E_2)P(E_3). \end{aligned}$$

However, to check if they are independent or not, we need to verify four conditions, *viz.*,

$$P(E_1 \cap E_2) = P(E_1)P(E_2),$$

$$P(E_1 \cap E_3) = P(E_1)P(E_3),$$
$$P(E_2 \cap E_3) = P(E_2)P(E_3),$$
$$P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3).$$

That means to check if three events are independent or not, we need to check one extra condition over the conditions that need to verify for pairwise independence.

In general, one needs to verify $\binom{n}{2}$ conditions to check if a collection of n events are pairwise independent or not. To check if a collection of n events are independent or not, $2^n - n - 1$ conditions need to be verified. Clearly, if a collection of events are independent, then they are pairwise independent. However, in general, the converse is not true as illustrated by the following example.

Example 1.18. Suppose that a coin is tossed twice. The sample space has four points and is given by $\mathcal{S} = \{HH, HT, TH, TT\}$. Suppose that all elementary events are equally likely. That is $P(HH) = P(HT) = P(TH) = P(TT) = 1/4$. Let $E_1 = \{HH, HT\}$, $E_2 = \{HH, TH\}$ and $E_3 = \{HH, TT\}$. Clearly, E_1, E_2, and E_3 are the events that the first toss results in a heads, second toss results in heads, and both the tosses have same outcomes, respectively. It is easy to see that $P(E_1) = P(E_2) = P(E_3) = 1/2$. Also

$$P(E_1 \cap E_2) = P(HH) = \frac{1}{4}.$$
$$P(E_1 \cap E_3) = P(HH) = \frac{1}{4}.$$
$$P(E_2 \cap E_3) = P(HH) = \frac{1}{4}.$$
$$P(E_1 \cap E_2 \cap E_3) = P(HH) = \frac{1}{4}.$$

Thus $P(E_1 \cap E_2) = 1/4 = P(E_1)P(E_2)$, $P(E_1 \cap E_3) = 1/4 = P(E_1)P(E_3)$, and $P(E_2 \cap E_3) = 1/4 = P(E_2)P(E_3)$. This shows that the events E_1, E_2, and E_3 are pairwise independent. However, $P(E_1 \cap E_2 \cap E_3) = 1/4 \neq 1/8 = P(E_1)P(E_2)P(E_3)$. Hence, E_1, E_2, and E_3 are not independent. ||

Example 1.19. Let a die be rolled twice. The sample space is given by

$$\mathcal{S} = \{(i, j) : i = 1, \ldots, 6, j = 1, \ldots, 6\}.$$

Suppose all elementary events are equally likely, *i.e.*, $P(\omega) = 1/36$ for all $\omega \in \mathcal{S}$. Let us consider following events

$$E_1 = \text{1st roll is 1, 2 or 3},$$
$$E_2 = \text{1st roll is 3, 4 or 5},$$
$$E_3 = \text{Sum of the rolls is 9}.$$

Clearly, $P(E_1) = 1/2$, $P(E_2) = 1/2$, and $P(E_3) = 1/9$. Also, $P(E_1 \cap E_2 \cap E_3) = 1/36 = P(E_1)P(E_2)P(E_3)$. However, E_1 and E_2 are not independent as $P(E_1 \cap E_2) = 1/6 \neq 1/4 = P(E_1)P(E_2)$. Thus E_1, E_2, and E_3 are not independent, not even pairwise independent. This example shows that verifying the condition $P(E_1 \cap E_2 \cap E_3) = P(E_1)P(E_2)P(E_3)$ is not enough to check if three events are independent or not. ||

Definition 1.23 (Conditional Independent)**.** *Given an event C two events A and B are said to be conditionally independent if $P(A \cap B|C) = P(A|C)P(B|C)$.*

Example 1.20. A box contains two coins: a fair regular coin and one fake two-headed coin (*i.e.*, $P(H) = 1$). The regular coin is called Coin 1 and the other is called Coin 2. You choose a coin at random and toss it twice. Define the following events.

$$A = \text{First coin toss results in a } H.$$
$$B = \text{Second coin toss results in a } H.$$
$$C = \text{Coin 1 (regular) has been selected.}$$

Here $P(A|C) = 1/2 = P(B|C)$, $P(A \cap B|C) = 1/4$. Hence, A and B are conditionally independent given C. As $P(A) = 3/4 = P(B)$ and $P(A \cap B) = 5/8$, A and B are not independent. Thus, the conditional independence does not imply independence in general. ||