

Friday Lecture Note

Friday Lecture: 29/07/2011

Probability: Mathematical Basis

By

Professor Umesh Singh

Department of Statistics and Coordinator DST-CIMS

Banaras Hindu University (BHU), Varanasi-221005, INDIA.

E-mail:

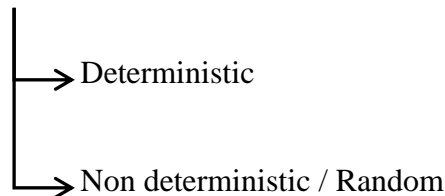
PART - I

Section 1

Where / when we need probability? \Rightarrow The experiments results cannot be predicted in advanced.

Experiment: An experiment is any process of trial and observation.

Experiment



Random Experiment: An experiment whose outcome is uncertain before it is performed is called a random experiment non deterministic. A random experiment / phenomenon are one that, under repeated observation, yields different outcomes that are not deterministically predictable. However, these outcomes obey certain conditions of statistical regularity whereby the relative frequency of occurrence of the possible outcomes is approximately predictable.

- (i). Suppose that the experiment is that you jump off a cliff to see whether you fly out in the North, East, West, South, up or down directions. In this experiment we know the physical laws governing the outcome. There are not 6 possible directions in which you fly but for sure there is only one direction, that is, going

down. The outcome is predetermined and the experiment is not a random experiment.

- (ii). If you put a 5-rupee coin into a bucket of water to see whether it sinks or not. There are not two possibilities that either it sinks or it does not sink. From the physical laws we know that the coin sinks in water and it is a sure event. The outcome is predetermined and the experiment is not a random experiment.
- (iii). If you throw a coin upward to see whether one side, call it head, or the other side, call it tail, will turn up when the coin falls to the floor, assuming that the coin will not stay on its edge when it falls to the floor, then the outcome is not predetermined. You do not know for sure whether head = $\{H\}$ will turn up or tail = $\{T\}$ will turn up. These are the only possible outcomes here, head or tail and these outcomes are not predetermined. This is a random experiment.
- (iv). A child played with a pair of scissors and cut a string of length 20 cm. If the experiment is to see whether the child cut the string then it is not a random experiment because the child has already cut the string. If the experiment is to see at which point on the string the cut is made then it is a random experiment because the point is not determined beforehand.

Sample Space: When we perform a random experiment, the collection of possible elementary outcomes is called the **sample space** of the experiment.

Sample Point: The outcomes defined as elementary outcomes, in which exactly one of the outcomes occurs when the experiment is performed. The elementary outcomes of an experiment are called the **sample points** of the sample space.

Notations:

E : Experiment; ω : Outcome; Ω : All possible outcome (Sample space)

A : Event,

: Statement about outcome of E ,

: Set of elementary outcomes / Set of sample point of sample space,

: Subset of Ω ,

$$A = \{\omega : \omega \in \Omega\}$$

$\mathcal{A} = \{A : A \subseteq \Omega\}$ is called event space. \mathcal{A} is a set of sets.

Let A be an event in the sample space Ω then $A \subset \Omega$, that is, A is a subset of Ω or all elements in A are also elements of Ω .

If A and B are two subsets of Ω , that is,

$$A \subset \Omega \text{ and } B \subset \Omega \text{ then } A \cup B \subset \Omega$$

is called the event of occurrence of either A or B (or both) or the occurrence of at least one of A and B .

$A \cap B$: is called the simultaneous occurrence of A and B .

In summary,

$$A \cup B = \text{occurrence of at least } A \text{ or } B$$

$$A \cap B = \text{simultaneous occurrence of } A \text{ and } B.$$

Example 1.1: Non deterministic

i. E_1 : Tossing of a coin twice.

$$\Omega_1 = \{(H, H), (H, T), (T, H), (T, T)\}; H, T \text{ denotes Head and Tail respectively.}$$

ii. E_2 : Tossing of a coin till Head is obtained.

$$\Omega_2 = \{H, TH, TTH, TTTH, \dots\}; H, T \text{ denotes head and tail respectively.}$$

iii. E_3 : Life of an electric bulb.

$$\Omega_3 = \{x : x \in \mathbb{R}^+\}; \mathbb{R}^+ \text{ denotes the set of positive real numbers?}$$

Event space, \mathcal{A}_1 , corresponding experiment E_1 :

$$\mathcal{A}_1 = \left\{ \begin{aligned} &\phi, \{(H, H)\}, \{(H, T)\}, \{(T, H)\}, \{(T, T)\}, \{(H, H), (H, T)\}, \{(H, H), (T, H)\}, \\ &\{(H, H), (T, T)\}, \{(H, T), (T, H)\}, \{(H, T), (T, T)\}, \{(T, H), (T, T)\}, \{(H, H), \\ &(H, T), (T, H)\}, \{(H, H), (H, T), (T, T)\}, \{(H, T), (T, H), (T, T)\}, \\ &\{(T, H), (T, T), (H, H)\}, \{(H, H), (H, T), (T, H), (T, T)\} \end{aligned} \right\}$$

Example 1.2: In a random experiment of rolling a die twice, construct

- i. the sample space S , and identify the events;
- ii. A = event of rolling 8 (sum of the face numbers is 8);
- iii. B = event of getting the sum greater than 10.

Solution 1.2.: Here in the first trial one of the 6 numbers can come and in the second trial also one of the six numbers can come. Hence the sample space consists of all ordered pairs of numbers from 1 to 6. That is,

$$\Omega = \left\{ \begin{array}{l} (1,1) (1,2) (1,3) (1,4) (1,5) (1,6) \\ (2,1) (2,2) (2,3) (2,4) (2,5) (2,6) \\ (3,1) (3,2) (3,3) (3,4) (3,5) (3,6) \\ (4,1) (4,2) (4,3) (4,4) (4,5) (4,6) \\ (5,1) (5,2) (5,3) (5,4) (5,5) (5,6) \\ (6,1) (6,2) (6,3) (6,4) (6,5) (6,6) \end{array} \right\}$$

There are 36 points in Ω .

A is the event of rolling 8. This can come by

$$A = \{(2,6), (3,5), (4,4), (5,3), (6,2)\}$$

The event of getting the sum greater than 10 means the sum is 11 or 12. Therefore

$$B = \{(5,6), (6,5), (6,6)\}.$$

Example 1.3: In the experiment of throwing a coin twice construct the events of getting:

- i. exactly two heads;
- ii. at least one head;
- iii. at least one tail, and interpret their unions and intersections.

Solution 1.3: Let A = event of getting exactly two heads, B = event of getting at least one head, and C = event of getting at least one tail. Here the sample space

$\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$; H, T denotes head and tail respectively. The events are

$$A = \text{exactly two heads} = \{(H,H)\}$$

$$B = \text{at least one head} = \{(H,T), (H,T), (H,H)\}$$

$$C = \text{at least one tail} = \{(T,H), (H,T), (T,T)\}$$

At least one head means exactly one head or exactly two heads (one or more heads), and similar interpretation for C also. [The phrase “**at most**” means that number or less, that is, at most 1 head means 1 head or zero head].

- Clearly, $A \cup B = B$, since $A \subset B$ (A is contained in B). Thus occurrence of A or B or both here means the occurrence of B itself because **occurrence of exactly 2 heads** or **at least one head** implies the **occurrence of at least one head**.

- $A \cup C = \{(H, H), (H, T), (T, H), (T, T)\} \equiv$ Occurrence of exactly 2 heads or at least one tail (or both), \Rightarrow covers the whole sample space or which is sure to occur and hence Ω can be called the sure event $= \Omega$.
- $A \cap B = \{(H, H)\} \equiv$ The simultaneous occurrence of exactly 2 heads and at least one head is the same as saying the occurrence of exactly 2 heads $= A$.
- $A \cap C = \text{null set} = \phi$. There is no element common to A and C . Also observe that it is impossible to have the simultaneous occurrence of exactly 2 heads and at least one tail because there is nothing common here. **Thus the null set, ϕ , can be interpreted as the impossible event.**
- $B \cup C = B \text{ or } C \text{ (or both)} = \{(H, T), (T, H), (H, H), (T, T)\} = \Omega$. This event is sure to happen. Because the event of getting at least one head or at least one tail (or both) will cover the whole sample space.
- $B \cap C = \{(H, T), (T, H)\} \equiv$ Event of getting exactly one head \equiv Event of getting exactly one tail.
- Also singleton elements are called *elementary event* in a *sample space*. Thus $\{(H, H)\}, \{(H, T)\}, \{(T, H)\}, \{(T, T)\}$ are four elementary events.
- Non-occurrence of an event A is denoted A^c or \bar{A} . For example

$$A \equiv \text{The event of getting exactly 2 head} = \{(H, H)\}$$

The event of non-occurrence of A is

$$\begin{aligned} A^c &\equiv \text{Occurrence of exactly one head or zero heads (exactly 2 tails)} \\ &= \{(H, T), (T, H), (T, T)\} \end{aligned}$$

- If A and B are two events in the event space, \mathcal{A}_1 , corresponding sample space, Ω , then $A \cap B = \phi$ means that event A and B cannot occur simultaneously. That is,

the occurrence of A excludes the occurrence of B and vice versa. In this case event A and B will be called *mutually exclusive events*.

In summary, from above discussion we have following results and interpretations for the events in the sample space:

- (i). Ω = Sample space = Sure event.
- (ii). ϕ = null space = Impossible event.
- (iii). $A \cup B$ = Occurrence of at least A or B or both.
- (iv). $A \cap B$ = Simultaneous occurrence of A and B .
- (v). $A \cap B = \phi$ Means A and B are mutually exclusive.
- (vi). $A \cup B = \Omega$ Means that A and B are totally exhaustive events.
- (vii). A^c = complement of A in Ω = non-occurrence of the event A

Note 1: (About randomness) In the example of throwing a coin once, suppose that a physicist is capable of computing the position of the coin and the amount of pressure applied when it was thrown, all the forces acting on the coin while it is in the air etc then the physicist may be able to tell exactly whether that throw will result in a head or tail for sure. In that case the outcome is predetermined. Hence one can argue that an experiment becomes **random only due to our lack of knowledge about the various factors affecting the outcome**.

Also note that we do not have to really throw a coin for the description of our random experiment to hold. We are only looking at the **possible outcomes** “if” a coin is thrown. After it is thrown we already know the outcome.

In olden days, a farmer used to watch the **nature of the cloud formation**, the **coolness in the wind**, the **direction of the wind** etc to predict whether a rain is going to come on that day. His prediction might be wrong in 70 percent of the times. Now a days a meteorologist can predict, **at least in the temperate zones**, the arrival of rain including the **exact time and the amount of rainfall**, very accurately at least one to two days beforehand. The meteorologist may be wrong only in less than one percent of the time. Thus, as we know more and more about the factors affecting an event we are

able to predict its occurrence more and more accurately and eventually possibly exactly.

In the light of the above details, is there anything called a random experiment?

Note 2 (About Impossible / Sure Event): The impossible event \emptyset is often misinterpreted. Suppose that a *monkey is given a computer to play with*. Assume that it does not know any typing but only playing with the keyboard with English alphabets. Consider the **event** that the monkey's final creation is **one of the speeches of President Bush word by word**.

This is not a logically impossible event and we will not denote this event by \emptyset . We use \emptyset for logically impossible events. The event that the monkey created one of the speeches is *almost surely impossible*. Such events are called *almost surely impossible events*.

Consider the event of the **best student in this class passing the next test**. We are almost sure that she will pass but only due to some unpredicted mishap she may not pass. This is *almost surely a sure event* but **not logically a sure event** and hence this cannot be denoted by our symbol Ω for sure event.

Assigning the numerical value for chance of occurrence of a random event in a event space for well-defined sample space:

Now consider again the **Example 1.1. (i):**

E_1 : Tossing of a coin twice.

$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$; H, T denotes head and tail respectively.

And corresponding event space is:

$$\mathcal{A}_1 = \left\{ \begin{array}{l} \emptyset, \{(H, H)\}, \{(H, T)\}, \{(T, H)\}, \{(T, T)\}, \{(H, H), (H, T)\}, \{(H, H), (T, H)\}, \\ \{(H, H), (T, T)\}, \{(H, T), (T, H)\}, \{(H, T), (T, T)\}, \{(T, H), (T, T)\}, \{(H, H), \\ (H, T), (T, H)\}, \{(H, H), (H, T), (T, T)\}, \{(H, T), (T, H), (T, T)\}, \\ \{(T, H), (T, T), (H, H)\}, \{(H, H), (H, T), (T, H), (T, T)\} \end{array} \right\}$$

Following properties are evident: For any $A, B \in \mathcal{A}_1$

- (i). $A \in \mathcal{A}_1 \implies A^c \in \mathcal{A}_1$; A^c is the complement of A . i.e. \mathcal{A}_1 is closed w.r.t. set complement operation.
- (ii). $A, B \in \mathcal{A}_1 \implies A \cap B \in \mathcal{A}_1$; i.e. \mathcal{A}_1 is closed w.r.t. set intersection operation.
- (iii). Properties (i) and (ii) imply $A \cup B \in \mathcal{A}_1$.

Note: 1.1. Add finite, infinite, countable, uncountable, and discreteness Concept.

Definition 1.1: If a collection of a set, i.e., set of sets, \mathcal{A} is closed under complement and intersection operation then it called a field.

Definition 1.2.: If \mathcal{A} is closed under countable union and intersection then it is called σ -field.

Thus \mathcal{A}_1 is field. Since \mathcal{A}_1 is countable, hence it is σ -field. Similarly we can show that event space corresponding and experiment E_2 also filed as well as σ -field.

Till now we transform the experiment into mathematical objects / notions filed / σ -field with the help of set Ω and primitive set operations (complement, intersection, union).

$$E \implies \Omega \implies (\mathcal{A}, \cap, \cup) \equiv (\mathcal{A}, \cap, ^c)$$

Till now we do not have an idea about the chance for occurrence a particular event in the event space for a trail. Now we shall try to assign a numerical value to each event, i.e., each element in event space, \mathcal{A} . This numeric value should gives the idea of chance of occurrence of particular event in an event space, \mathcal{A} , defined over sample space Ω ; and it is consistent with our common sense understanding of chance in everyday life.

Notation: Let $P(A)$ denotes the numeric value assign to the even $A \in \mathcal{A}$.

Let's try to give meaning to following types of statements that we use in regular routine of our life:

- i. There is 95% chance to of rain today. $\Rightarrow P(A) = 0.95$; where A is the event of having rain.
- ii. The chance of winning certain lottery is one in million. $\Rightarrow P(A) = 1/10000$.
- iii. The chance of flood on campus today is nearly zero. $\Rightarrow P(A) \approx 0$.

From above example, it is clear that the in every day life the chance / possibility of occurrence of particular event is associated a numeric value in between 0 and 1. The rule for associating numeric value to the event / phenomenon is such that as the possibility of occurrence of an event increases, the numeric value assign to event will increases and approaches to 1 and 1 for certain event. And in reverse case it decrease and approaches 0 and 0 for impossible event.

Thus, numeric association rule is a function from event space, \mathcal{A} , to $[0, 1]$.

$$P: \mathcal{A} \longrightarrow [0, 1](\subseteq \mathbb{R})$$

The value associates with an event $A \in \mathcal{A}$ is denoted as $P(A)$. We define this numeric association rule, i.e., function P , for every event, A , in event space, \mathcal{A} , such that it satisfy the some postulates or axioms. Postulates and axioms are logically consistent and non-overlapping types of basic assumptions that we make to define something that occurring in everyday life. Usually such postulates are taken by taking into consideration plausible properties that we would like to have for the item to be defined.

The following three axioms / postulates will be used to define $P(A)$, the probability or chance of occurrence of the event A :

- i. $P(A) \geq 0$ or the probability of an event is nonnegative.
- ii. $P(\Omega) = 1$ or the probability of the sure event is 1.
- iii. $P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$, whenever $A_i \cap A_j = \phi, i \neq j$ (i.e. events A_i 's are mutually exclusive). The events may be finite or countably infinite in number.

Thus $P(.)$ coming out of the above three axioms will be called the probability of the event $(.)$.

Thus we transform the phenomenon / experiment into triplet (Ω, \mathcal{A}, P) :

$$E \implies \Omega \implies (\mathcal{A}, \cap, \cup) \equiv (\mathcal{A}, \cap, ^c) \implies (\Omega, \mathcal{A}, P)$$

Triplet (Ω, \mathcal{A}, P) is called as probability space.

Note: 1.2: Using above three axioms one can derive that the upper bound of $P(.)$ is 1. i.e., for any event $A \in \mathcal{A}$ $P(A) \leq 1$. Most of the literature explicitly write the lower and upper bound of $P(.)$ in axiom (i) as $0 \leq P(A) \leq 1$

- 1.3. A function satisfying axiom (i) and (iii) is called measure.
- 1.4. The additive property given in axiom (iii) is called as σ -additive.
- 1.5. Axiom (ii) is called as property of normness.

Section 2:

How to Assign Probabilities to Individual Events?

Since elements of any event are the some of the element of the sample space (Ω). Thus, a event $A^k \in \mathcal{A}$ can be expressed as follows:

$$A^k = \bigcup_i \{\omega_{ki}\}, \text{ where } \omega_{ki} = \omega_j \in \Omega, \text{ for some } j$$

And $\{\omega_{ki}\} \cap \{\omega_{kj}\} = \phi$, for $i \neq j$. Now using axiom (iii), we get following:

$$\begin{aligned} P(A^k) &= P\left(\bigcup_i \{\omega_{ki}\}\right) \\ &= \sum_i P(\{\omega_{ki}\}) \end{aligned}$$

It is obvious from above relation if we know the probability of an event having single element, $P(\{\omega_{ki}\})$, we can calculate the $P(A^k)$.

How to assign probability to the events in event space having single out come ($\{\omega_i\}$), i.e., individual outcome, ω_i , in sample space ?

$$\{\omega_i\} \subseteq \Omega, \quad \bigcup_i \{\omega_i\} = \Omega, \quad \{\omega_i\} \cap \{\omega_j\} = \phi \text{ for } i \neq j$$

Now apply the axioms (i)-(iii), we have following:

$$\begin{aligned} P\left(\bigcup_i \{\omega_i\}\right) &= P(\Omega) \\ \Rightarrow \sum_i P(\{\omega_i\}) &= 1 \\ \Rightarrow \sum_i p_i &= 1, \text{ where } P(\{\omega_i\}) = p_i \geq 0 \end{aligned}$$

Thus, if we have a way to assign positive real numbers p_i to each events $\{\omega_i\}$ such that the $\sum_i p_i = 1$, then we can assign a real number to each event in event space.

Let see what will be that positive real numbers in the following simple example:

Example 2.1: Compute the probability of getting a head when a coin is tossed once.

Solution 2.1: The sample space for this experiment is as follows:

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

Let A be the event of getting a head, $A = \{H\}$, and B be the event of getting a tail,

$B=\{T\}$. Clearly event A and B are mutually exclusive events, i.e., $A \cap B = \phi$. In a particular trial, one of the events A or B must occur, i.e., either a head or a tail will turn up because we have ruled out the possibility that the coin will fall on its edge.

Thus, we have following facts:

$$A \cap B = \phi, \quad A \cup B = \Omega.$$

Using axiom (ii) we get

$$P(A \cup B) = P(\Omega) = 1$$

Since $A \cap B = \phi$, therefore on use of axiom (iii) we get

$$\begin{aligned} P(A \cup B) &= 1 \\ \Rightarrow P(A) + P(B) &= 1 \\ \Rightarrow P(A) &= 1 - P(B) \end{aligned}$$

With the help of axioms we can only come up to this point and cannot proceed further. The above statement does not imply that only $P(A) = 1/2$, and $P(B) = 1/2$ only satisfy the equation $P(A) + P(B) = 1$. There are infinitely many values in $[0, 1]$ for $P(A)$ and $P(B)$ such that the relation $P(A) + P(B) = 1$ is satisfied. In fact, by fixing any one of the $P(A)$ or $P(B)$ we can get other using above relation.

Thus, by using the axioms we cannot compute $P(A)$ even for this simple example of throwing a coin once.

Example 2.2: Compute the probability of an event having single outcome when a coin is tossed twice.

Solution 2.2: The sample space for this experiment is as follows:

$$\Omega = \{(H, H), (H, T), (T, H), (T, T)\}.$$

The events having single outcome are as follows:

$$A_1 = \{(H, H)\}, A_2 = \{(H, T)\}, A_3 = \{(T, H)\}, A_4 = \{(T, T)\}.$$

Clearly $A_i \cap A_j = \phi$ for $i \neq j$, i.e., all these events are mutually exclusive and

$$\bigcup_{i=1}^4 A_i = \Omega.$$

On use of axiom (iii), we get

$$\begin{aligned} P\left(\bigcup_{i=1}^4 A_i\right) &= P(\Omega), \\ \Rightarrow \sum_{i=1}^4 P(A_i) &= P(\Omega). \end{aligned}$$

Using axiom (ii) we get

$$\sum_{i=1}^4 P(A_i) = 1$$

$$\Rightarrow p_1 + p_2 + p_3 + p_4 = 1, \text{ where } P(A_i) = p_i$$

Consider the following solution of $p_1 + p_2 + p_3 + p_4 = 1$

$$p_1 = p^2, p_2 = pq, p_3 = pq, p_4 = q^2; \text{ where } 0 \leq p \leq 1, \text{ and } q = 1 - p.$$

Thus we have infinitely many solution, and hence infinitely many possibility to assign probability to event A_i^s , not only $1/4$ to each A_i^s . Again same bottleneck what we experienced in example of coin is tossed once.

Example 2.3: Compute the probability of getting sum greater than 10 when a die is rolled twice.

Solution 2.3: The total number of outcome in a sample space, Ω , for this experiment is 36.

Let A be the event of getting sum greater than 10, which means a sum 11 or 12. Then the event is given as follows:

$$A = \{(5, 6), (6, 5), (6, 6)\}$$

Let B be the event of getting a sum less than or equal to 10. Clearly this event is compliment of event A , i.e., $B = A^c$. But for any event A , $A \cup A^c = \Omega$ and $A \cap A^c = \phi$. Thus events A and B are mutually exclusive and totally exhaustive events.

Now on use of axioms (ii) and (iii) we get following:

$$P(A) + P(B) = 1$$

Number of elements in event A is 3 of the total 36 outcomes in sample space. But we cannot jump the conclusion that $P(A) = 3/36$, because from the definition of probability, as given by the postulates, does not depend upon the number of sample points or elementary events favorable to the event. If someone makes such a conclusion and write the probability as $3/36$ in the above case it will be wrong, which may be seen from the following considerations:

(i).A housewife is drying clothes on the terrace of high-rise apartment building. What is the probability that she will jump of the building?

There are only two possibilities that she will jump of the building, either she jumps off or she does not jump off. Therefore if you say that the probability is $1/2$ obviously you wrong. The chance is practically nil that she will jump off the building.

(ii).At this place, tomorrow can be sunny day, or a cloudy day or a mixed sunny and cloudy day, or a rainy day. What is the probability that tomorrow will be rainy

day?

If you say that the probability is $1/4$ since we have identified four possibilities you can be obviously wrong because these four possibilities need not have the same probabilities. Since today is sunny and since it is not a rainy season most probably tomorrow will also be a sunny day.

(iii).A child cuts a string of 40 cm length into two pieces while playing with a pair of scissors. Let one end of string is marked as 0 and the other end as 40. What is the probability that the point of cut is in the sector from 0 to 8 cm?

Here the number of sample points is infinite, not even countable. Hence by the misuse of the idea that the probability may be the number of sample points favorable to the event to the total number of sample points, we can not come up with an answer, even though wrong, as in previous example. The total number as well as the number of points favorable to the event cannot be counted. Then how do we calculate this probability?

(iv).A person is throwing a dart at a square board of 100cm length and width. What is the probability that the dart will hit in a particular 10cm×10cm region on the board?

Again we can count the number of sample points even if to misuse the numbers to come up with an answer. Then how do we compute this probability?

(v).A floor is paved with square tiles of length m units. A circular coin of diameter d units is thrown upward, where $d < m$. What is the probability that the coin will fall clean within a tile, not cutting its edges or corners?

This is the famous Baffon's, "clean tile problem" from where the theory of probability has its beginning.

How do we answer this type of questions, definitely not from counting the number of outcomes / sample points? From the above examples it is clear that the probability of an event does not depend upon the number of outcomes in sample space and number of outcome favorable to that event. It depends upon many factors.

Also we have seen that our definition of probability, through the three axioms, does not help us to assign probability to a given event. In other words the theory is useless, when it comes to the problem of computing the probability of a given event or theory is not applicable to practical situation, unless we introduce more assumptions or

extraneous considerations about the physical characteristics and the factor affecting the outcomes of the experiments. The knowledge about extraneous considerations can be gain through common sense, experiment modeling, or from observations. Moreover, this Information / knowledge about the experiment are mostly subjective in nature rather than objective.

Reference: All above 5 examples are from A. C. Mathai, Module 6, Page 25 – 26

Section 3 (This section is based on A. C. Mathai, Module 6, Page 32 – 38)

Some Guideline to Assign Probabilities to Individual Events

As we illustrated in section 2 through different examples that the meaning full assignment of probability to the out comes of an experiment is not only done by axioms and knowing all possible outcomes of an experiment. For assigning meaning full probability we need some additional information about the experiment.

There is no hard and fast rule for assigning probability to the outcomes of an experiment. Here we introduce a number of rules those together with the axiomatic definition helps us to assign / compute the probabilities for many cases.

Rule 1.1.: Symmetry in outcomes

For finite sample space with distinct number of elements / outcomes, if physical characteristics of experiment is such that - with respect to all factors which may affect the possible outcomes - there is no way to favoring one outcome over other then, the rule say to assign equal probabilities to the all elements in sample space.

In order to apply this rule one has to have a sample space consisting of a finite number of elements, situations such as:

- tossing a coin once or a number of times, rolling a die a number of times
- predicting successful completion of a job when there are only a fixed number of alternatives, predicting rainfall etc.

The rule ***does not apply*** to situations such as:

- the cutting a string where the sample space consists of a continuum of points,
- throwing a dart at a target where the sample space consists of regions,

- Buffon's clean tile problem where the sample space consists of a room paved with tiles or anite planar region, and so on.

The implication of **Rule 1.1** is the following:

Rule 1.1a: *When there is symmetry in the outcomes of a random experiment and when here are k elementary events in the sample space S , k being finite and if m of the sample points (elementary events) are favorable to an event A then the probability of the event A will be taken as:*

$$P(A) = \frac{\text{number of sample points favorable to } A}{\text{total number of sample points}} = \frac{m}{k}$$

- Let us take the example of tossing a coin twice. The sample space is $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. If the physical characteristics of the coin is such that there is no way of preferring one side to the other (in such a case we call the coin *an unbiased coin* or not loaded towards one side), the throwing of the coin is such that there is no advantage for one side over the other or, in short, with respect to all factors which may affect the outcomes, there is no advantage for one side over the other, then in this case we assign equal probabilities of $1/4$.

Note 1: “Events are equally likely” Events have equal probabilities. The statement is circumlocutionary in the sense of using “probability” to define probability. Symmetry has nothing to do with the chances for the individual outcomes. We have the axioms defining probabilities and we have seen that the axioms are not sufficient to compute the probabilities in specific situations and hence it is meaningless to say “**equally likely events**” when trying to compute the probabilities of events. *Symmetry is concerned about the physical characteristics of the experiments and the factors affecting the outcomes and not about the chances of occurrence of the events.*

Note 2: The phrase used to describe symmetry is “**unbiased coin**” in the case of coins, “**balanced die**” in the case of rolling a die and in other cases we say “when there is symmetry in the experiment or symmetry in the outcomes”.

- **Thus, in the example of tossing a coin if we ask:** what is the probability of getting a head when an **unbiased coin** is tossed once, then **the answer** is $1/2$. *This value is assigned by us by taking into account of symmetry in the experiment, and not coming from the axioms or deduced from somewhere.*

- **What is the probability of getting exactly one head when an unbiased coin is tossed twice?**

Answer: Let A be the event of getting exactly one head. Let A_1 be the event of getting the sequence (H, T) and A_2 be the event of getting the sequence (T, H) . Then

$$A = A_1 \cup A_2 \text{ and } A_1 \cap A_2 = \phi$$

Then from **axiom (iii)** we have:

$$P(A) = P(A_1) + P(A_2)$$

\Rightarrow Using **addition assumption of symmetry property** we can assign equal probability to events A_1 and A_2 :

$$P(A_1) = \frac{1}{4} \text{ and } P(A_2) = \frac{1}{4}$$

Therefore,
$$P(A) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Example: An unbiased coin is tossed (a) three times and (b) four times. What are the probabilities of getting the sequences (i): HHT , (ii): THT in (a) and the sequences (iii): $HHTT$, (iv): $HHHH$ or $HTTT$ in (b)?

Solution: In (a) the sample space consists of all possible sequences of H and T and there are 8 such elementary events. They are available by looking at the problem of filling three positions by using H , and T . The first position can be filled in two ways either H or T , for each such choice the second position can be filled in two ways, for each such choice for the first and second positions the third can be filled in two ways so that the number of possible outcomes is $2 \times 2 \times 2 = 8$. They are the following:

$HHH, HHT, HTH, HTT, THH, THT, TTH, TTT$

Assuming the symmetry, all these 8 points are assigned probabilities $1/8$. Hence answer to the (i) and (ii) are:

$$P(\{HHT\}) = \frac{1}{8}, \text{ and } P(\{THT\}) = \frac{1}{8}$$

(b) When the coin is tossed four times the sample space consists of $2 \times 2 \times 2 \times 2 = 16$ elementary events. Due to symmetry we assign probabilities $\frac{1}{16}$ to each of these points. Hence

$$P\{(HHHH)\} = \frac{1}{16}$$

In (iv) the event of getting the sequences $HHHH$ or $HTTT$ means the **union of two mutually exclusive events** and by the third axiom, the probability is the sum of the probabilities. We have assigned probabilities $1/16$ each and hence

$$\begin{aligned} P\{(HHHH \text{ or } HTTT)\} &= \{(HHHH) \cup (HTTT)\} \\ &= P\{(HHHH)\} + P\{(HTTT)\} \\ &= \frac{1}{16} + \frac{1}{16} \\ &= \frac{1}{8} \end{aligned}$$

Example: A balanced die is rolled two times. What is the probability of (i): rolling 9, (ii): getting a sum greater than or equal to 10?

Solution: When we say “**balanced**” it means that we are assuming symmetry in the experiment and we are assigning equal probabilities to all elementary events. Here there are 36 elementary events and each point will get probability of $\frac{1}{36}$ each.

(i): Rolling 9 means the sum of the face numbers is 9. The possible elementary events in this event are (3, 6), (4, 5), (5, 4), (6, 3). These are mutually exclusive because, for example, when the sequence (3, 6) comes at the same time another sequence cannot come. Let A be the event of rolling 9, let A_1 to A_4 denote the events of getting the sequences (3, 6), ..., (6, 3) respectively.

$$\begin{aligned} A_1 &= \{(3,6)\}, \quad A_2 = \{(4,5)\}, \quad A_3 = \{(5,4)\} \quad A_4 = \{(6,3)\} \\ A &= A_1 \cup A_2 \cup A_3 \cup A_4 \end{aligned}$$

and $A_i \cap A_j = \emptyset$; for $i \neq j$. That is, A_1, A_2, A_3 and A_4 are mutually exclusive. Hence by the third axiom in the definition of probability:

$$P(A) = P(A_1) + P(A_2) + P(A_3) + P(A_4).$$

We have assigned equal probabilities to elementary events. Hence

$$P(A) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{4}{36} = \frac{1}{9}$$

(ii): Let event B = be the event getting sum greater than equal to 10 = sum is 10 or 11 or 12. Thus elementary event / point in sample space favorable to this event are (6,4), (4,6), (5,5), (5,6), (6,5), (6,6). Thus event B :

$$B = \{(6,4), (4,6), (5,5), (5,6), (6,5), (6,6)\}$$

Thus, using third axiom of probability we can get the probability of event B :

$$P(B) = \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{9}$$

Rule 1.2: Assign probability 0 for almost surely impossible events and probability 1 almost surely sure events.

By assigning probability 0 we are not saying that the corresponding event is logically impossible. If an event is logically impossible then its probability is zero as a consequence of the axioms defining probability.

When we assign 1 to almost surely a sure event we are not saying that the event is a sure event. For a logically sure event the probability is 1 by the second axiom defining probability. But an *assigned probability 1 does not mean that the event is a sure event.*

Rule 1.3: If the sample space consists of a continuum of points giving a line segment (or segments) of finite length (or lengths), such as a piece of string of length 50cm, and if the experiment is to take a point from this line segment (or segments), such as a cut on this string, and if there is no preference of any sort in selecting this point then assign probabilities proportional to the lengths, taking the total length as unity.

When a point is selected from a line segment of finite length by using the rule of assigning probabilities proportional to the lengths then we use the phrase: ***a point is selected at random*** from the line segment or we have a “**random point**” from this line segment or if a string is cut by using the above rule we say that we have a ***random cut*** of the string or we say that the ***point of cut is uniformly distributed over the line segment***. These are all standard phrases used in this situation.

If an event A is that the random point lies on a segment of length m units out of a total length of n units, $n \geq m$, then rule says:

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

Example: A random cut is made on a string of 30cm in length. Marking one end of the string as zero and the other end as 30 what is the probability that **(i):** the cut is between 10 and 11.7, **(ii):** the cut is between 10 and 10.001, **(iii):** the cut is at 10, **(iv):** the smaller piece is less than or equal to 10cm?

Solution:

Since we use the phrase “random cut” we are assigning probabilities proportional to the lengths. Let x be the distance from the end marked 0 to the point of cut. **Let A be the event that $A = \{x | 10 \leq x \leq 11.7\}$, [this notation means: all values of x such that x is between 10 and 11.7, both the end points are included], B be the event that $B = \{x | 10 \leq x \leq 10.001\}$, let C be the event that $C = \{x | x = 10\}$ and let D be the event that the smaller piece is less than or equal to 10 cm.**

The length of the interval in A is $11.7 - 10.0 = 1.7$. Since we are assigning probabilities proportional to the lengths we have:

$$P(A) = \frac{11.7 - 10.0}{30} = \frac{1.7}{30} = \frac{17}{300}$$

$$P(B) = \frac{10.001 - 10.0}{30} = \frac{0.001}{30} = \frac{1}{30000}$$

and

$$P(C) = \frac{10 - 10}{30} = \frac{0}{30} = 0$$

Since we are assigning probabilities proportional to the lengths and since a *point does not have any length by definition*, then according to this rule the probability that the cut is at a specific point, in a continuum of points, is zero. *By assigning this value zero to this probability we are not saying that it is impossible to cut the string at that point.* As per our rule of assigning probabilities proportional to lengths, then since a *point does not have length the point will be assigned probability zero as per this rule.*

For the **event D** the smaller piece of *length less than or equal to 10cm* in the following two situations:

$$D_1 = \{x | 0 \leq x \leq 10\} \quad \text{and} \quad D_2 = \{x | 20 \leq x \leq 30\}$$

Therefore,

$$D = D_1 \cup D_2, \text{ where } D_1 \cap D_2 = \phi$$

Thus D_1 and D_2 are mutually exclusive. Thus from Axiom (iii) of probability:

$$P(D) = P(D_1) + P(D_2) = \frac{10-0}{30} + \frac{30-20}{30} = \frac{10}{30} + \frac{10}{30} = \frac{20}{30} = \frac{2}{3}$$

Note 1: This variable x can be said to be ***uniformly distributed*** over the line segment $[0, 30]$.

Note 2: The above rule cannot be applied if the string is of infinite length such as a beam of light or laser beam or sound wave etc. How do we compute probabilities in such situations of strings of infinite length?

Rule 1.4. *When a point is selected at random from a planar region of infinite area assign probabilities proportional to the area and when a point is selected at random from a higher dimensional space of infinite hyper-volume, then assign probabilities proportional to the volume. According to this rule if the **total area is** α and out of this, if $\mu(\alpha)$ **of the area is favorable** to an event **A** then the probability of **A** is assumed as:*

$$P(A) = \frac{\mu(\alpha)}{\alpha}$$

*Similarly, if v is the **total volume** (or hyper-volume) of the space under consideration and if $\mu(v)$ the **fraction of v is favorable to an event A** then, as per the above rule, the probability of A is taken as:*

$$P(A) = \frac{\mu(v)}{v}$$

Several items here need explanations. When a point is taken at random from a planar region of finite area α , such as the point of hit of an arrow when the arrow is shot onto a wall of length 10 meters and width 2 meters (area = $\alpha = 10 \times 2 = 20$ sq. meters), here “**at random**” means that there is no preference of any sort for the point to be found anywhere on the planar region. Then we assign probabilities $\frac{\alpha_1}{\alpha}$ to every possible sub region of area α_1 .

Similar interpretation for higher dimensional situations. The standard terminology for length, area, volume etc is the following: **length (one dimensional)**, **area (two-dimensional)**, **volume (3-dimensional)**, **hyper-volume (4 or higher dimensional)**. For simplicity we say “**volume**” for **3 or higher dimensional cases**, instead of saying “**hyper-volume**”.

Example: A person trying dart throwing for the first time throws a dart at random to a circular board of radius 2 meters. Assuming that the dart hit the board, what is the probability that (1): it hit within the central region of radius 1 meter; (2): it hit along a horizontal line passing through the center; (3): it hit exactly at the center of the board?

Solution: (1) Assuming that the point of hit is a random point on the board we may assign probabilities proportional to the area. The total area of the board is the area of a circle of radius 2 meters.

$$\text{Total area} = \pi r^2 = \pi(2)^2 = 4\pi \text{ m}^2$$

where the standard notation m^2 means square meters. (1): The area of the central region of radius one meter $= \pi(1)^2 = \pi \text{ m}^2$. Hence the required probability, denoted by $P(A)$, is

$$P(A) = \frac{\pi \text{ m}^2}{4\pi \text{ m}^2} = \frac{1}{4}$$

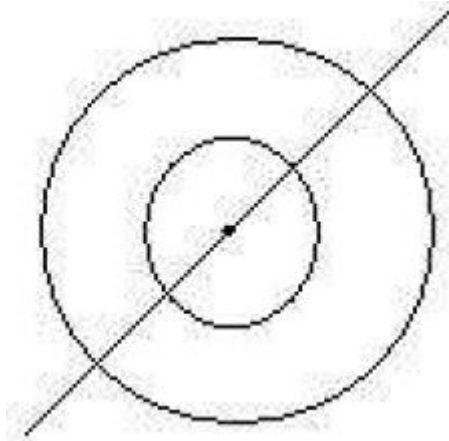


Figure : Circular board and circular, line, point targets.

(2) We have to look at the area along a line passing through the center. But, by definition, a line has no area and hence the area here is zero. Thus the required probability is $\frac{0}{4\pi} = 0$. In (3) also a point has no area by definition, and hence the probability is zero.

Note 1: Note that the numerator and denominator here are in terms of square meters, but probability is a pure number and has no unit of measurement or does not depend on any unit of measurement.

Note 2: Also when assigning probabilities proportional to the area remember that lines and points have no area, a point has no length or area but a line has length but

no area. Similarly when assigning probabilities proportional to the volume, remember that a planar region has no volume but it has area, a line has no volume or area but has length and a point has no length, area or volume.

Example: Solve Buffon's clean tile problem. That is, a circular coin of **diameter d** is thrown upward. When it falls on the floor paved with square tiles of **length m** with $d < m$ what is the probability that the coin will fall clean, which means that the coin will not cut any of the edges and corners of the tiles?

Solution: In Figure below a typical square tile is marked. Since the coin is tossed upward we assume that the center of the coin could be anywhere on the tile if the coin has fallen on that tile. In other words we are assuming that *the center of the coin is a random point on the square tile or uniformly distributed over that square tile*. In Figure below an inner square is drawn $d / 2$ distance away from the walls of the outer square. If the center of the coin is any where on the walls of the inner square or in the region between the walls of the two squares then the coin can touch or cut the walls of the outer square.

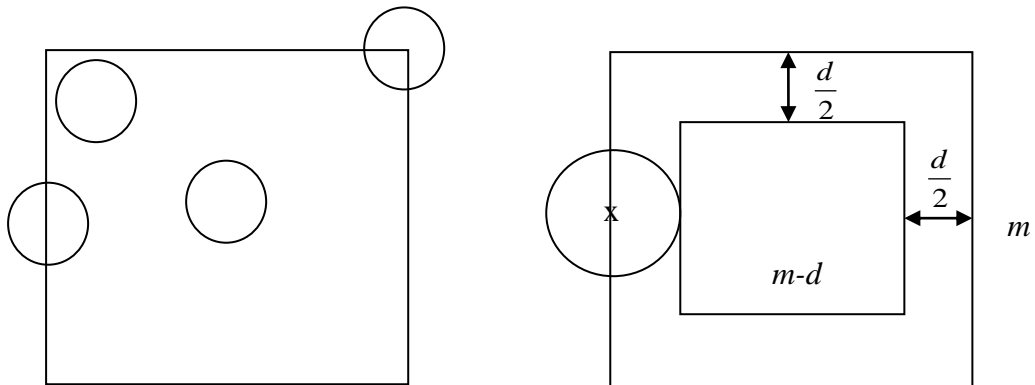


Figure : Square tile and circular coin

If the center of the coin is strictly within the inner square then the coin will fall clean. Therefore the probability of the event, A = the event that the coin falls clean, is given by:

$$P(A) = \frac{\text{area of inner square}}{\text{area outer square}} = \frac{\left(m - \frac{d}{2} - \frac{d}{2}\right)^2}{m^2} = \frac{(m-d)^2}{m^2}$$

Note 1: This problem is generalized by looking at a floor paved with rectangular tiles of length m units, width n units and a circular coin of diameter d units where $d < m$; $d < n$. This problem can be done in a similar way by looking at the center of the coin and assuming that the center is uniformly distributed over the rectangle. The floor can be paved with any symmetrical objects such as rhombuses or general polygons, and a circular coin is tossed. The problem is to compute the probability that the coin will fall clean.

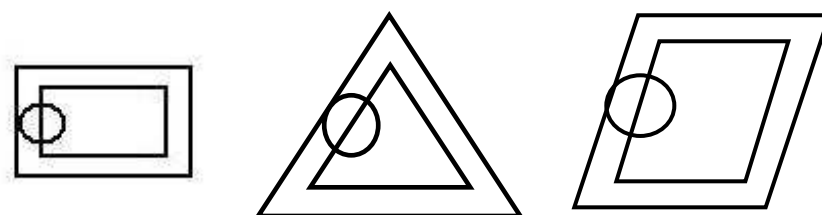


Figure: Tiles of various shapes

Note 2: A three-dimensional generalization of the problem is to consider a prism with square, rectangular, parallelogram, or general polygonal base and a ball or sphere of radius r is randomly placed inside the prism. What is the probability that the ball will not touch any of the sides or base or top of the prism? When we move from one-dimensional case to two or higher dimension then more axioms such as “**invariance**” is needed to define probability measures.

Example: Solve Buffon's needle problem: A floor is paved with parallel lines, m units apart. A headless needle (or a line segment) of length d is tossed up. What is the probability that the needle will touch or cut any of the parallel lines when the needle falls to the floor? There are several situations of interest here.

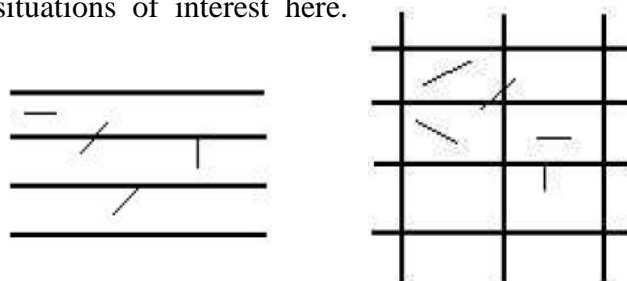


Figure: Buffon's needle problem

- i. Short needle where the length of the needle, d , is less than m .
- ii. When $d < 2m$ and $d > m$.

- iii. Another case is a long needle which can cut a number of parallel lines. Remember that however long the needle may be there is a possibility that the needle need not cut any of the lines, for example, the needle can fall parallel to the lines.
- iv. Another needle problem is when the floor has horizontal and vertical lines making rectangular grids of length m units and width n units and a needle of length d is tossed. A generalization of this problem is the case when the needle can be of any shape, need not be straight.

For dealing with Buffon's needle problem we need the concepts of random variables and independence of random variables. Hence we will not do examples here.

Event Algebra and Some Results From Axiomatic Definition of Probability

Event Algebra: Since any event is subset of sample space. Collection of all subset of sample space is called as event space. Thus, here event space of an experiment works as universal set. Almost all the identities related to the set, with respect to the set operation (intersection, union, complementation, and difference) holds for event also. Here we discuss some important result borrowed from set algebra with reference to event.

Venn Diagrams Representation of Set: In a Venn diagrammatic representation of a **set**, a set is represented by a closed curve, usually a rectangle or a circle or ellipse, and subsets are represented by **closed curves** within the set or **by points within the set or by region** within the set. Examples are as given in bellow Figure.

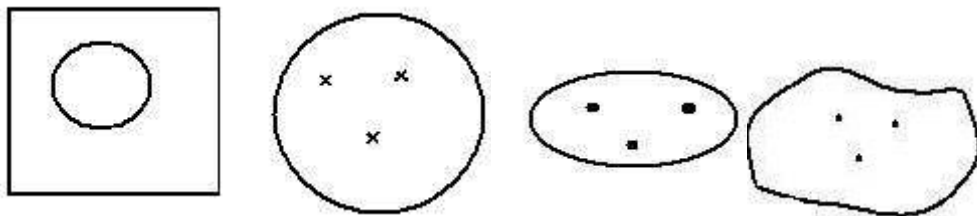


Figure: Venn diagram for sample space

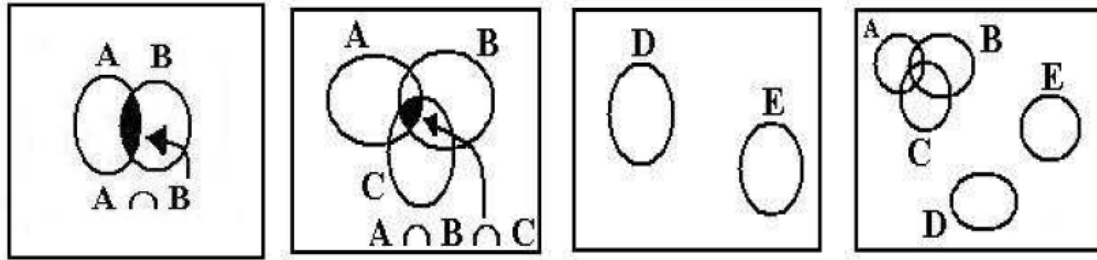


Figure: Representation of events

Let $A \subset \Omega$, $B \subset \Omega$, $C \subset \Omega$, $D \subset \Omega$, and $E \subset \Omega$ are events in the sample space Ω . In the Venn diagram in the above Figure, **A and B intersect, A and C intersect, B and C intersect, A , B , C all intersect, D and E do not intersect. $A \cap B$ is the shaded region, also $A \cap B \cap C$ is the shaded region.**

Clearly, $D \cap E = \phi$, hence they are mutually exclusive. Following properties are evident from the Venn diagram that $A \cap B^c$, $A \cap B$, and $A^c \cap B$ are mutually exclusive

$$\begin{aligned} A \cup B &= (A \cap B^c) \cup (A \cap B) \cup (A^c \cap B) \\ &= A \cup (A^c \cap B) \\ &= B \cup (B^c \cap A) \end{aligned}$$

where,

$$\begin{aligned} (A \cap B^c) \cap (A \cap B) &= \phi, \\ (A \cap B^c) \cap (A^c \cap B) &= \phi, \\ (A \cap B) \cap (A^c \cap B) &= \phi, \\ B \cap (B^c \cap A) &= \phi. \end{aligned}$$

or they are mutually exclusive events. Also for any event A ,

$$A \cup \phi = A, \quad A \cap \phi = \phi$$

This means that event A and impossible event ϕ are mutually exclusive, sure event Ω and the impossible event ϕ are mutually exclusive

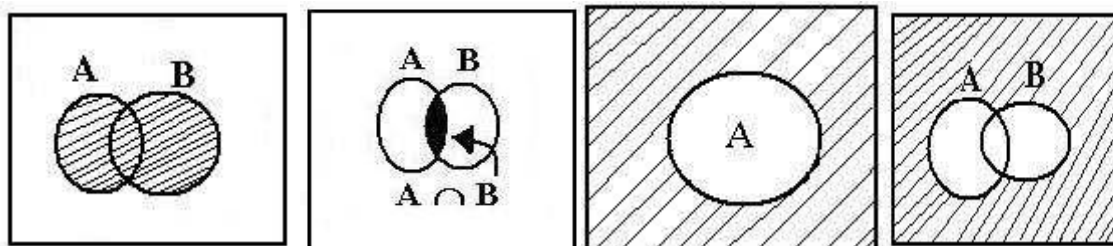


Figure: Union, Intersection, Complementation of Events

Partition of a Sample Space:

Consider the set of events A_1, \dots, A_k in the same sample space Ω , that is, $A_i \subset \Omega$, $j = 1, 2, 3, \dots, k$ where k could be infinity also or there may be countably infinite number of events. Let these events be mutually exclusive and totally exhaustive. That is:

$$A_1 \cap A_2 = \phi, A_1 \cap A_3 = \phi, \dots, A_1 \cap A_k = \phi$$
$$A_2 \cap A_3 = \phi, \dots, A_2 \cap A_k = \phi, \dots, A_{k-1} \cap A_k = \phi$$

and

$$S = A_1 \cup A_2 \cup A_3 \cup \dots \cup A_k.$$

This can also be written as follows:

$$A_i \cap A_j = \phi, \text{ for } i \neq j, i, j = 1, 2, \dots, k, A_1 \cup A_2 \cup \dots \cup A_k = \Omega$$

Then we say that the sample space Ω is partitioned into k mutually exclusive and totally exhaustive events. This may be represented following Venn diagram.

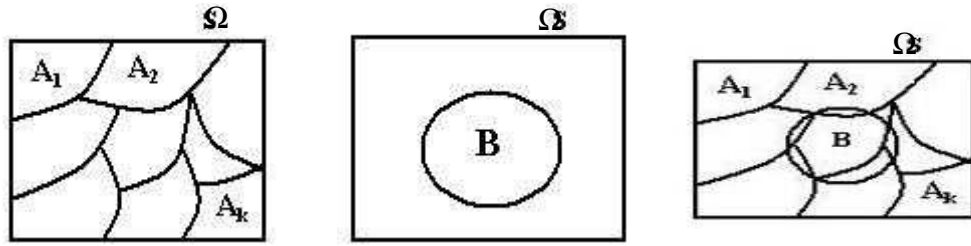


Figure: Partition of a sample space

Results from Axiomatic Definition of Probability

Result 1: Probability of an impossible event is zero, i.e., $P(\phi) = 0$

Proof: Consider the sure event Ω and impossible event ϕ . Now from the definitions $\Omega \cup \phi = \Omega$ and $\Omega \cap \phi = \phi$. There for event Ω , ϕ are mutually exclusive and collectively exhaustive. There for from axiom (ii) and (iii) of probability, we have:

$$P(\Omega) = P(\Omega \cup \phi) = P(\Omega) + P(\phi) = 1$$
$$P(\Omega) = P(\Omega \cup \phi) = P(\Omega) + P(\phi) = 1$$
$$1 + P(\phi) = 1, \text{ since } P(\Omega) = 1, \text{ from Axiom (ii)}$$

Therefore,

$$P(\phi) = 0$$

Result 2: Probability of non-occurrence = 1-probability of occurrence, or $P(A^c) = 1 - P(A)$

Proof:

Result 3: For any two events A and B in the sample space Ω , i.e., $A \subseteq \Omega$, $B \subseteq \Omega$,

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

Proof:

Result 4: Let A , B and C be three events in the same sample space Ω , then

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

Proof:

Result 5: Let there are n arbitrary event, A_1, A_2, \dots, A_n in the sample space Ω , then

$$P(A_1 \cup A_2 \cup A_3 \cup \dots A_n) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i \cap A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k) - \dots$$

Result 6: If $A = A_1 \cup A_2 \cup A_3 \dots A_n$ where A_1, A_2, \dots, A_n are mutually exclusive events, then

$$P(A) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) \\ = \sum_{i=1}^n P(A_i)$$

Proof:

Result 7: For any two events A and B , $P(A) = P(A \cap B) + P(A \cap B^c)$

and $P(B) = P(A \cap B) + P(A^c \cap B)$.

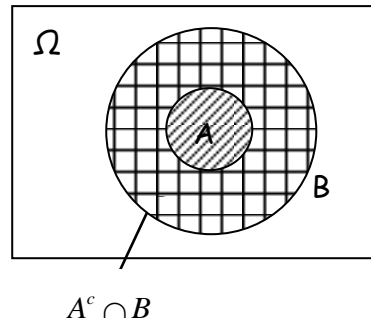
Proof:

Result 8: Let A and B be two event in the same sample space Ω , such that $A \subseteq B$, then $P(A) \leq P(B)$

Proof: From figure below we see that the event A and $A^c \cap B$ are mutually exclusive, and

$$B = A \cup (A^c \cap B)$$

Since A and $A^c \cap B$ is mutually exclusive, therefore from Axiom (iii) we have:



$P(B) = P(A) + P(A^c \cap B)$ this gives:

$$P(B) - P(A) = P(A^c \cap B)$$

Since $P(A^c \cap B) \geq 0$, therefore

$$P(B) - P(A) \geq 0$$

Thus,

$$P(A) \leq P(B)$$

Result 9: Let A be an event in the sample space Ω , then $P(A) \leq 1$.

Proof: Since Ω is sample space, there event A is subset of Ω , i.e.,

$$A \subset \Omega$$

Thus from result 9, we have:

$$P(A) \leq P(\Omega)$$

Since $P(\Omega) = 1$, therefore, $P(A) \leq 1$

Interpretation of Probability

The term probability has four interpretations:

- I. Axiomatic definition
- II. Classical Probability Theory (Equally likely)
- III. Relative Frequency Approach to The Probability (Empirical)
- IV. Subjective (Measure of belief / As Intuition)

I. Axiomatic Definition: We have already discussed the axiomatic definition. Here again review the concept in measure theoretic concept proposed by Russian mathematician **Andrei N. Kolmogorov (1903 – 1987)**.

Andrei N. Kolmogorov (1903{1987), was a brilliant Russian mathematician and physicist who made important, original contributions to probability theory, random processes, information theory, complexity theory, mechanics, uid dynamics, and

nonlinear dynamical systems theory. He proposed the axiomatic approach to probability in 1933, when he has 30 years old. He is a giant of the 20th century.

In the **axiomatic approach to probability theory**, we assume the existence of a probability function and justify its correctness after the fact via systematic testing and validation. A probability function (also known as a probability measure) is a set-function, $P(A)$, which maps events, $A \subset \Omega$, to the nonnegative real numbers. It is assumed that the probability measure satisfies the following three axioms.

Kolmogorov Probability Axioms

- i. For any event $A \subset \Omega$, $P(A) \geq 0$.
- ii. $P(\Omega) = 1$.
- iii. If the countable sequence of events A_i are mutually exclusive $A_i \cap A_j = \emptyset$, $i \neq j$ then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i).$$

Property (i) is the property of nonnegativity of $P(\cdot)$; Property (ii), normalization of $P(\cdot)$; and Property (iii), countable additivity of $P(\cdot)$. That $P(\cdot)$ obeys $P(\emptyset) = 0$ and is finitely additive is entailed by Properties (ii) and (iii).

(Kolmogorov) Probability Space: A Probability Space is a triple (Ω, \mathcal{A}, P) where

- i. Ω is a sample space of outcomes.
- ii. \mathcal{A} is a nonempty σ -algebra of Ω -events (subsets of Ω).
- iii. $P(\cdot)$ is a probability measure on \mathcal{A} which satisfies the **Kolmogorov probability axioms**.

Determination of Probability: We have already discussed some guideline for assigning the probability, $P(A)$, to the events $A \in \mathcal{A}$. Before the advent of the axiomatic approach, one would try to derive a probability model from a priori arguments (*e.g.*, “outcomes are equally likely” or “probabilities are relative frequencies”), but this could never be put on a rigorous footing which was universally applicable. **For instance, the assumption of equally likely outcomes could never handle the problem of a weighted coin or loaded die.**

In the axiomatic framework, one instead postulates that there exists a well-defined probability function, even if we don't know exactly what it is, and constructs a mathematical model based on that assumption. We then make principled choices of the probability values to be assigned to specific events. To do so, we use intuition; experience; mathematical and physical reasoning; symmetry arguments; and engineering custom. However the assignments are made, the model must then be justified after the fact via testing and experimentation. Often, a few iterations are required before an acceptable model is determined. **[For instance, the assumption of equally likely outcomes could never handle the problem of a weighted coin or loaded die]**. This is particularly the case in communications theory. Common questions about the random behavior of a wireless channel are “is it Gaussian?”; “is it Rayleigh?”, “is it Rician?”, “is it multipath?”, etc, etc, etc.

In this approach, the one which we have been following, a self-consistent mathematical model of probability and events is constructed based on the assumption of a few fundamental axioms. **(E.g., in our case we are working with the Kolmogorov probability space axiomatic model.)** Only after the axiomatic model has been constructed, are mathematical consequences and predictions of the model then compared to the measured behavior of a real-world situation or system of interest. If there is a “**reasonably good**” match between the model's predicted behavior and the corresponding measured behavior of the real-world situation, then the axiomatic model is deemed to be an acceptable mathematical model of that situation. If the match is poor, we go back to the drawing board and attempt to revise our model. Thus, **axiomatic probability models are justified after-the-fact, by how well they explain and predict measured real-world behavior**. Self-consistent axiomatic models themselves are neither true nor false; rather they are “**better or worse**” in their degree of correspondence to a measured phenomenon of interest.

II. Classical Probability Theory (*Probability as the Ration of Favorable to Total Outcomes*):

A probability space is finite if its sample space is finite. Classical probability theory assumes a finite probability space and equiprobable outcomes, $P(\omega) = \text{constant}$ for all $\omega \in \Omega$. Using the probability axioms, it is easy to show that the probability of a single outcome is $\frac{1}{N}$, where $N = N(\omega) = \#(\Omega) = \text{cardinality of } \Omega$.

An outcome, ω , is said to be favorable to A if $\omega \in A$. The number of outcomes

favorable to the event A is $N(A) = \#(A)$ = cardinality of A . Using the probability axioms it is readily shown that the probability of the event A is

$$P(A) = \frac{N(A)}{N(\Omega)} = \frac{\#(A)}{\#(\Omega)}$$

Thus, we see that the name of the game in classical probability theory is **counting**. One needs to count the number of possible outcomes to determine (e.g., how many total five-card deals are possible) and $N(A) = \#(A)$ (eg, how many ways can a *royal flush* be dealt) before one can determine the probability of A (e.g., $A = \{\text{royal flush}\}$).

Not surprisingly, then, **combinatorics** (*the theory of counting*) is an important topic in **classical probability** and much time is spent developing proficiency in computing permutations and combinations.

III. Relative Frequency Approach To Probability (Empirical) : In this probability is considered as frequency of occurrence relative to the total number of trial. Suppose that we have repeated trials so that an experiment whose sample space is Ω repeatedly performed under exactly the same conditions. After n trials (repetitions of the experiment) have been performed, for any event, $A \subset \Omega$, we define $n(A)$ to be the number of times the event A occurred. The Relative Frequency of the Event A , $f_n(A)$, is defined to be the proportion of times that A occurred in the n trials, the probability $P(A)$ is approximated as:

$$P(A) \approx f_n(A) = \frac{n_A}{n} \quad \text{for } n \text{ “large”}.$$

The above relative frequency cannot be used as the definition of $P(A)$, because it is an approximation. The approximation improves, however, as n increases. One might wonder, therefore, whether we can define $P(A)$ as limit:

$$P(A) = \lim_{n \rightarrow \infty} f_n(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n} \quad \text{for } n \text{ “large”}.$$

We cannot, of course, do so if n and n_A are experimentally determined number, because in any real experiment, the number n trials, although it might be large, it is always finite. To give meaning to the limit we must interpret above limiting definition as an assumption used to define $P(A)$ as theoretical concept. [*This approach was introduced by Von Mises in 1957 as the foundation of new theory based on above limiting definition of $P(A)$].*

Although there is little, or no, controversy about the use of an axiomatic model of probability, there is controversy about its interpretation when used to explain real-world phenomena of interest. So-called Frequentists are perhaps the *most conservative* and *accept only the Relative Frequency Interpretation of Probability*. This interpretation says that probability models can only be used to model situations where an (potentially) unlimited number of repeated trials is possible. Frequentists do not admit any probabilistic interpretations of so-called “one-off” events (events which only occur one-time). **The Frequency Interpretation of Probability assumes that repeated trials can be performed so that relative frequencies can be computed.** In this case, a model is acceptable if it can be determined that whenever relative frequencies of an event A are empirically measured we have that:

$$P(A) \approx f_n(A) = \frac{n_A}{n} \quad \text{for } n \text{ “large”}.$$

In this case, we accept the model and go on to interpret the probability, $P(B)$, of any other event as the likely relative frequency of occurrence in n trials for n \large enough. "Equivalently, for n “large enough” we expect to find that:

$$n(B) \approx n.P(B)$$

For instance, suppose a patient has been told that base on the positive outcome of a medical test he has a 10% probability of **contracting a certain genetic blood disorder** after age 60. When asked what this means, he is told by his **Frequentist doctor** that, based on data amassed in clinical trials, it means that of 1000 men who test positive on this test, one can expect about 100 of them to contract the disorder after age 60. **Note that this interpretation assumes that sufficient data exists to back up such an assertion.** The **Frequency Interpretation is sometimes called an Objectivist Interpretation** as one tries to justify it using objective, measured data collected from repeated trials.

“This approach can be partially justified by appealing to theoretical results known as Strong Laws of Large Numbers”.

IV. Subjective (Measure of belief / As Intuition):

So-called subjectivists go beyond the objective Frequency Interpretation of Probability and are willing to interpret probabilities in one-off situations where there isn't, perhaps never has been and never will be, data sufficient to construct relative frequencies. In this case, a subjectivist interprets the probability of an event A as a measure of his or her personal belief that the event A will occur. Not surprisingly, the subjectivist interpretation is controversial, even though in practice it is used extensively.

For example, suppose in the medical situation described above there is little or no data available to make a frequency interpretation (e.g., perhaps only 5 people in the world have ever even had the disease throughout history!). And suppose the doctor still tells the patient that, in his opinion, he has a 10% probability of contracting the disease. When the patient asks what that means, the doctor says that based on his professional judgment built up from years of looking at related ailments, it is his personal belief that the patient will likely get the ailment is 10%. (Whatever that means!|which is why this interpretation is considered subjective rather than *objective*.)

PART - III

Random Variable and Distribution

[Mapping sample space to numeric number. How set operations are converted to point operation. Illustrate event using point operation]

In previous section we defined and illustrated the probability function P , whose domain is a set of sets, i.e. event space \mathcal{A} , over a set of outcomes, Ω , of an experiment and range is $[0, 1]$. Since domain of P is a set, so it is not convenient to handle, moreover we can not perform very much familiar arithmetic and algebraic operation, and very well established real analysis and calculus to investigate the nature of P . Thus, it is desirable to innovate a point function equivalently to P (i.e. gives same information as P of the event space / sample space) so that we can use the knowledge of real analysis and calculus to study the properties of P .

In order to define point function equivalently to P , we must represent the domain of P

i.e, event space \mathcal{A} , with the help of real numbers. Basic building blocks of an event space are an outcome of an experiment, thus we first map the set of outcomes to real number R .

For illustration, consider an experiment of tossing of a coin twice:

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

where H, T denote head and tail respectively. Now we try to map Ω to set of real numbers R . Consider following finite, single valued function $X, \Omega \xrightarrow{X} R$, as follows:

$$X(\omega) = \text{number of } H^{\text{'s}} \text{ in } \omega.$$

Then

$$X(HH) = 2, X(HT) = X(TH) = 1, \text{ and } X(TT) = 0.$$

How to induce event space into real numbers and corresponding set operation to point operation with the help of X

$$\mathcal{A} = \left\{ \begin{array}{l} \phi, \{HH\}, \{HT\}, \{TH\}, \{TT\}, \{HH, HT\}, \{HH, TH\}, \{HH, TT\}, \\ \{HT, TH\}, \{HT, TT\}, \{TH, TT\}, \{HH, HT, TH\}, \{HH, HT, TT\}, \\ \{HT, TH, TT\}, \{TH, TT, HH\}, \{HH, HT, TH, TT\} \end{array} \right\},$$

Consider events in which number of head less than equal to 1. These events in event space are as follows: $\phi, \{HT\}, \{TH\}, \{TT\}, \{HT, TH\}, \{HT, TT\}, \{TH, TT\}, \{HT, TH, TT\}$

In many cases, one is not interested in the occurrence of the particular outcome in the sample space. Rather, one would like to know the number of heads in two tosses or more general in n -tosses.