AXIOMATIC DEFINITION OF PROBABILITY

The mathematical and statistical definitions of probability have their own disadvantages. So they do not contribute much to the growth of the probability theory. The axiomatic definition is due to A.N. Kolmogorov (1933), a Russian mathematician, and is mathematically the best definition of probability since it eliminates most of the difficulties that are encountered in using other definitions. This axiomatic approach is based on measure theory. Here we introduce it by means of set operations.

Sample space

A sample space is the set of all conceivable outcome of a random experiment. The sample space is usually denoted by S or W. The notion of a sample space comes from Richard Von Mises.

Every indecomposable outcome of a random experiment is known as a *sample point or elementary outcome.* The number of sample points in the sample space may be finite, countably infinite or noncountably infinite. Sample space with finite or countably infinite number of elements is called discrete sample space. Sample space with continuum of points is called continuous sample space.

Example

- The sample space obtained in the throw of a single die is a finite sample space, ie. S = {1, 2, 3, 4. 5, 6}
- The sample space obtained in connection with the random experiment of tossing a coin again and again until a head appears is a countably infinite sample space.

  ie. S {H, TH, TTH, TTTH .... }
- Consider the life time of a machine. The outcomes of this experiment form a continuous sample space.

  ie., $S = \{ t : 0 < t < \infty\}$

Event

An event is a subset of the sample space. In other words, "of all the possible outcomes in the sample space of an experiment, some outcomes satisfy a specified description, which we call an event."

Field of events (F)

Let S be the sample space of a random experiment. Then the collection or class of sets F is called a field or algebra if it satisfies the following conditions.

- F is nonempty
- the elements of F are subsets of S.
- if $A \in F$, then $A^C \in F$
- if $A \in F$ and $B \in F$ then $A \cup B \in F$

For example, let S = { 1, 2, 3, 4, 5, 6 }

Choose F as the set with elements ϕ, S, { 5, 6} and {1, 2, 3,4) Then F satisfies all the four conditions. So F is a field.

More generally, when $A \subset S$, $F = \{\phi, A, A^C, S\}$ forms a field. Trivially, F with just two elements ϕ and S forms a field.

σ-'field or σ-algebra of events

Let S be a nonempty set and F be a collection of subsets of S. Then F is called a σ-field or σ-algebra if

- F is nonempty
- Tile elements of F are subsets of S
- If $A \in F$,then $A^C \in F$ and
- The union of any countable collection of elements of F is an element of F.

$$\text{i.e., if } A_i \in F,\ i = 1, 2, 3, \dots n, \text{ then } \bigcup_{i=1}^{\infty} A_i \in F$$

The σ algebra F is also called *Borel field* and is often denoted by B.

Examples

- $B = \{\phi, S\}$
- $B = \{0, A, A^C, S\}$
- $B = \{\phi, A, B, S\}$ provided $A \cup B = S$ and $A \cap B = \phi$
- The powerset of S always form a Borel field.

Function and Measure

We know that a function or mapping is a correspondence between the elements of the set X (called domain) and the set Y (called range) by a rule or principle. When the elements of the domain are sets and the elements of the range are real numbers, the function is said to be a 'set function'. A set function is usually denoted by P(A) or $\mu(A)$ where A represents an arbitrary set in the domain.

In a set function if $A_1, A_2, \ldots, A_n$ are disjoint sets in the domain and if

$$\mu(A_1 \cup A_2 \cup A_3 \cup \ldots \cup A_n) = \mu(A_1) \cup \mu(A_2) \cup \mu(A_3) \cup \ldots \mu(A_n)$$

then the set function is said to be *additive.*

If a set S is partitioned into a countable number of disjoint sets $A_1, A_2, \ldots$ and if a set function defined on satisfies the property.

$$\mu(A_1 \cup A_2 \cup \ldots) = \mu(A_1) + \mu(A_2) + \ldots$$

i.e., $$\mu\left(\bigcup_{1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$$

then the set function is said to be *countably additive.*

Measure

A set function which is non negative and totally additive is called a measure. A *measure* will be called a probability measure if

$$\mu(A_1 \cup A_2 \cup A_3 \cup \ldots A_n) = \mu(A_1) + \mu(A_2) + \mu(A_3) + \ldots + \mu(A_n) = 1$$

where $\bigcup_i Ai = S$, $Ai \cap Aj = \phi$, $i \neq j$

In probability theory, the probability measure is denoted by P instead of μ.

Axiomatic definition

Let S be the sample space. Let B be the class of events constituting the Borel field. Then for each $A \in B$, we can find a real valued set function P (A), known as the probability for the occurrence of A if P(A) satisfies the following three axioms,

Axiom 1. **(Non negativity)**

$0 \leq P(A) \leq 1$ for each $A \in B$

Axiom 2. **(Norming)**

$P(S) = 1$

Axiom 3**. (Countable additivity)**

If $A_1, A_2, \ldots, A_n$ is a finite or infinite sequence of elements in B such that $Ai \cap Aj = \phi$, $i \neq j$.

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

Probability Space

From the axiomatic definition of probability we can conceive of a probability space constituting the triplet (S, B, P) where S represents the sample space, B is the class of all subsets of S constituting a Borel field, and P is the probability function with domain B and satisfying the axioms 1, 2 and 3 of probability given above.

Probability space is a single term that gives us an expedient way to assume the existence of all three components in its notation. The three components are related; B is a collection of subsets of S and P is a function that has B as its domain. The probability space's main rise is in providing a convenient method of stating background assumptions for future definitions and theorems etc.

Note: The axiomatic definition of probability proposed by Kolmogorov reveals that the numbers in the interval [0, 1] can be assigned as probabilities of events in some initial class of elementary events. Using these probabilities we can determine the probability of any event which may be of interest. The calculus of probability begin after the assignment of probabilities represented by the symbols p_1, p_2, p_3 ...... which are usually determined on the basis of some past experience or on the basis of some empirical study.

Theorems in Probability

The following are some consequences of the axioms of probability, which have got general applications and so they are called theorems. We can make use of Venn diagrams for the better understanding of these theorems.

Theorem I

The probability of an impossible event is ZERO.

ie., $P(\phi) = 0$,

Proof

Let ϕ be the impossible event.

Then $S \in B$ and $\phi \in B$

We have $S \cup \phi = S$

♣ $P(S \cup \phi) = P(S)$

i.e., $P(S) + P(\phi) = P(S)$, since S and ϕ are disjoint.

i.e., $1 + P(\phi) = 1$ – by axiom 2

$\therefore$ $P(\phi) = 0$

Note:

The condition $P(A) = 0$ does not imply that $A = \phi$

Example

Consider an experiment of tossing a coin infinitely many times. The outcomes may be represented as infinite sequences of the form HHTHTTTHHT….. so that the sample space S consists of infinitely many such sequences. The event 'head only' given by the sequence {HHHH….} is not empty. However, the chance of such an outcome is, atleast intuitively, zero. Tails should come up sooner or later.

Theorem 2

Probability is infinitely additive

i.e., $P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n)$

where $A_i \cap A_j = \phi$, $i \neq j$.

Proof

Consider an infinite sequence of events $A_1, A_2, A_3, \dots A_n, \phi, \phi, \phi, \dots$ which are pairwisely disjoint since Ai's are disjoint.

Then by axiom 3

$P(A_1 \cup A_2 \cup \dots A_n \cup \phi \cup \phi \cup \dots) =$

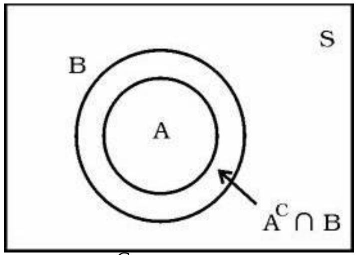
$P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n) + P(\phi) + P(\phi) \dots$ i.e.,

$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + P(A_3) + \dots + P(A_n)$

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

Theorem 3 (Monotonicity)

If $A \subset B$, then $P(A) \leq P(B)$



Proof

From the Venn diagram

We have $B = A \cup (A^C \cap B)$

$\therefore P(B) = P[A \cup (A^C \cap B)]$

$= P(A) + P(A^C \cap B)$ since A and $A^C \cap B$ are disjoint

$= P(A) + a$ +ve quantity i.e., $P(B) \geq P(A)$ or $P(A) \leq P(B)$

Note: From the above, we get $P(B - A) = P(B) - P(A)$ since $A^C \cap B = B - A$

Theorem 4

Probability is countably subadditive. i.e., for every sequence of events $A_1, A_2, \ldots$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq P(A_1) + P(A_2) + P(A_3) + \ldots$$

Proof

By considering the infinite operations on events, we can write the union of events into union of disjoint events,

i.e., $$\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_1^C \cap A_2) \cup (A_1^C \cap A_2^C \cap A_3) \cup \ldots$$

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq P(A_1) + P(A_1^C \cap A_2) + P(A_1^C \cap A_2^C \cap A_3) + \ldots$$

$$\leq P(A_1) + P(A_2) + P(A_3) + \ldots$$

Since $P(A_1^C \cap A_2) \leq P(A_2)$ etc., i.e., $P\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} P(A_i)$

Theorem 5 (Complementation)

$P(A^C) = 1 - P(A)$

Proof

We have $A \cup A^C = S$ $\therefore P(A \cup A^C) = P(S)$

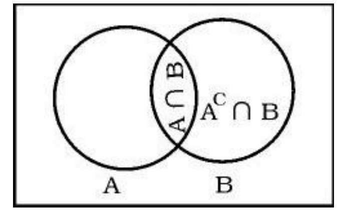
i.e., $P(A) + P(A^C) = 1$, by axiom 2 and 3 $\therefore P(A^C) = 1 - P(A)$

i.e., P [Non occurrence of an event] = 1 – P[Occurrence of that event]

Theorem 6 (Addition theorem of two events)

If A and B are any two events,

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



Proof

From the Venn diagram,

We can write

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

$\therefore P(A \cup B) = P[A \cup (A^C \cap B)]$

$= P(A) + P(A^C \cap B)$ … (1) since $A \cap (A^C \cap B) = \phi$

On the other hand,

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

$\therefore P(B) = P[(A \cap B) + P(A^C \cap B)]$ since $(A \cap B) \cap (A^C \cap B) = \phi$

$P(A^C \cap B) = P(B) - P(A \cap B)$ .. (2)

On substituting (2) in (1) we get,

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

Corollary (1) If $A \cap B = \phi$, $P(A \cup B) = P(A) + P(B)$

□ $P(A \cup B) = 1 - P(A \cup B)^C$

$= 1 - P(A^C \cap B^C)$

i.e., P[the occurrence of atleast one event] = 1 – P[None of them is occurring]

Theorem 7 (Addition theorem for 3 events)

If A, B, C are any three events,

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

$P(A \cap C) + P(A \cap B \cap C)$

Proof

Let $B \cup C = D$, Then $P(A \cup B \cup C) = P(A \cup D)$

$= P(A) + P(D) - P(A \cap D)$, by theorem 6

$\Rightarrow P(A) + P(B \cup C) - P[A \cap (B \cup C)]$

$\Rightarrow P(A) + P(B) + P(C) - P(B \cap C) - P[(A \cap B) \cup (A \cap C)]$

$\Rightarrow P(A) + P(B) + P(C) - P(B \cap C) - \{P(A \cap B) +$

$P(A \cap C) - P(A \cap B \cap C)\}$

$\Rightarrow P(A) + P(B) + P(C) - P(A \cap B) -$

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

Corollary

- If the event A, B, C are mutually exclusive

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

- $P(A \cup B \cup C) = 1 - P(A \cup B \cup C)^C = 1 - P(A^C \cap B^C \cap C^C)$

Probability in finite sample space with equally likely points

For certain random experiment there is a finite number of outcomes, say n and the probability attached to each outcome is 1/n. The classical definition of probability is generally adopted for these problems. But we can see that the axiomatic definition is applicable as well.

Definition : Let $E_1, E_2, \ldots E_n$ be n sample points or simple events in a discrete or finite sample space S. Suppose the set function P with domain the collection of all subsets of S satisfies the following conditions.

(i) $P(E_1) = P(E_2) = P(E_3) = \ldots = P(E_n) = \frac{1}{n}$

(ii) $P(S) = P(E_1 \cup E_2 \cup \ldots E_n) = P(E_1) + P(E_2) + \ldots + P(E_n)$

$\frac{1}{n} + \frac{1}{n} + \ldots \frac{1}{n}$ (n terms) $= \frac{n}{n} = 1$

- If A is any event which contains m sample points, say $E_1, E_2, \ldots E_m$

then, $P(A) = P(E_1 \cup E_2 \cup \ldots E_m)$

$= P(E_1) + P(E_2) + \ldots + P(E_m)$, since $E_i \cap E_j = \phi, i \neq j$

$= \frac{1}{n} + \frac{1}{n} + \ldots \frac{1}{n}$ (m terms) $= \frac{m}{n}$

This shows that P(A) satisfies all the axioms of probability. Thus we can see that the classical definition is a particular case of axiomatic definition. In other words, the axiomatic definition can be deduced to classical definition of probability if it is defined on a discrete or finite sample space with equally likely points.

SOLVED PROBLEMS

Example 10

A die is rolled. If x is the number shown on the die. 7x coins are tossed, If y is the number of heads (x , y) is recorded. Write down the sample space of this experiment.

Solution

If x is 1, 7 coins are tossed. If x = 2, 14 coins are tossed and so on. If x = 6, 42 coins are tossed. When y denotes the number of heads obtained, with x = 1, the pair (x, y) takes the values (1, 1) (1 , 2) (1 , 3) … (1, 7}. Thus the required sample space is

S = {(1,1}, (1.2), (1,3)........ (1, 7)}
(2.1), (2,2), (2.3) .... (2, 14)
(3.1), (3.2), (3.3) .... (3, 21)
(4.1), (4.2), (4,3) .... (4,28)
(5.1), (5.2), (5,3) .... (5,35)
(6,1), (6,2), (6.3) .... (6,42)}

Example 11

If A_1, A_2, A_3 are three events which are exhaustive, show that $B_1 = A_1$, $B_2 = A_1^C \cap A_2$, $B_3 = A_1^C \cap A_2^C \cap A_3$ are exhaustive and mutually exclusive.

Solution

Since A_1, A_2, and A_3 are exhaustive, we have

$A_1 \cup A_2 \cup A_3 = S$

We have to show that $B_1 \cup B_2 \cup B_3 = S$, Now $B_1 \cup B_2 \cup B_3 = A_1 \cup (A_1^C \cap A_2) \cup (A_1^C \cap A_2^C \cap A_2)$

$\{(A_1 \cup A_1^C) \cap (A_1 \cup A_2)\} \cup (A_1^C \cap A_2^C \cap A_3)$

$\{S \cap (A_1 \cup A_2)\} \cup (A_1^C \cap A_2^C \cap A_3)$

$(A_1 \cup A_2) \cup \{(A_1 \cup A_2)^C\} \cap A_3$

$(A_1 \cup A_2) \cup \{(A_1 \cup A_2)^C\} \cap (A_1 \cup A_2 \cup A_3)$

$S \cap S = S$

i.e., the events B_1, B_2 and B_3 are exhaustive.

To show that B_1, B_2 and B_3 are mutually exclusive,

$B_1 \cap B_2 \cap B_3 = A_1 \cap (A_1^C \cap A_2) \cap (A_1^C \cap A_2^C \cap A_3)$

$(A_1 \cap A_1^C) \cap A_2 \cap (A_1^C \cap A_2^C \cap A_3)$

$(\phi_1 \cap A_2) \cap (A_1^C \cap A_2^C \cap A_3)$

$\phi \cap (A_1^C \cap A_2^C \cap A_3)$

ϕ $\therefore$ The events B_1, B_2 and B_3 are mutually exclusive.

Example 12

In a swimming race the odds that A will win are 2 to 3 and the odds that B will win are 1 to 4. Find the probability and the odds that A or B wins the race?

Solution

We have $P(A) = \frac{2}{2+3} = \frac{2}{5}$

$P(B) = \frac{1}{1+4} = \frac{1}{5}$

P(A or B) = P(A) + P(B) since A and B are m.e

$= \frac{2}{5} + \frac{1}{5} = \frac{3}{5}$

$\therefore$ Odds that A or B wins are 3 to 2.

Example 13

Given P(A) = 0.30, P(B) = 0.78 and $P(A \cap B) = 0.16$. Find

i. $P(A^C \cap B^C)$ ii, $P(A^C \cup B^C)$ iii. $P(A \cap B^C)$

Solution

Given P(A) = 0.30, P(B) = 0.78 and $P(A \cap B) = 0.16$.

(i) $P(A^C \cap B^C) = P\{(A \cup B)^C\} = 1 - P(A \cup B)$

$1 - \{P(A) + P(B) - P(A \cap B)\}$

$1 - \{0.30 + 0.78 - 0.16\} = \mathbf{0.08}$

$P(A^C \cup B^C) = P\{(A \cap B)^C\} = 1 - P(A \cap B)$

$1 - 0.16 = \mathbf{0.84}$

(iii) $P(A \cap B^C) = P[A - (A \cap B)]$

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

$0.30 - 0.16 = \mathbf{0.14}$

Example 14

The probability that a student passes statistics test is 2/3 and the probability that he passes both statistics and Mathematics test is 14/45. The probability that he passes at least one test is 4/5. What is the probability that he passes Mathematics test?

Solution

Define, A - the student passes statistics test.

B -he passes the Mathematics test.

Given P(A) = 2/3. $P(A \cap B) = 14/45$. $P(A \cup B) = 4/5$

We have to find P(B). By addition theorem,

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

ie., 4/5 = 2/3 + P(B) – 14/45.

$\therefore P(B) = 4/5 - 2/3 + 14/45 = \frac{70}{\ }$