PROBABILITY - THE PROBABILITY FUNCTION, KOLMOGOROV’S AXIOMS

The probability function is a function defined on the set of events (subsets of S). To each event A, it assigns a real number which is its probability \( P(A) \). To define (or characterize) a probability function \( P \), it is necessary and sufficient that it satisfies the following axioms.

1. Probability of any event \( A \) over the sample space \( S \) is nonnegative: \( P(A) \geq 0 \).
2. Probability of the sample space is 1: \( P(S) = 1 \).
3. The probability of a union of two mutually exclusive events \( A \) and \( B \) is the sum of their probabilities: \( P(A \cup B) = P(A) + P(B) \) for any mutually exclusive events \( A \) and \( B \).
4. The probability of a union of infinitely many pairwise disjoint events, is the sum of their probabilities. That is, if \( A_1, A_2, \ldots \) are events over \( S \) such that \( A_i \cap A_j = \emptyset \) for \( i \neq j \), then \( P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i) \).

NOTE: Axioms 1 -3 are enough for finite sample spaces. Axiom 4 is necessary when the sample space is infinite.

PROPERTIES OF PROBABILITY FUNCTION \( P \)

Suppose \( P \) is probability function on the subsets of the sample space \( S \), and \( A \) and \( B \) are events defined over \( S \). Then, the following are true.

1. \( P(A^c) = 1 - P(A) \).
2. \( P(\emptyset) = 0 \).
3. If \( A \subseteq B \), then \( P(A) \leq P(B) \).
4. For any event \( A \), \( P(A) \leq 1 \).
5. If \( A_1, A_2, \ldots, A_k \) are events over \( S \) such that \( A_i \cap A_j = \emptyset \) for \( i \neq j \), then \( P(\bigcup_{i=1}^{k} A_i) = \sum_{i=1}^{k} P(A_i) \).
6. **Addition Rule:** For any two events \( A \) and \( B \): \( P(A \cup B) = P(A) + P(B) - P(A \cap B) \).
CONDITIONAL PROBABILITY, TOTAL PROBABILITY, AND BAYES RULE

Assume we have sample space \( S \) and events \( A \) and \( B \) defined on \( S \).

**Definition:** The conditional probability of event \( A \) given that event \( B \) occurred is

\[
P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad \text{if } P(B) \neq 0.
\]

(1)

**Theorem: Multiplication Rule:** The probability of \( A \) and \( B \), \( P(A \cap B) \) can be found using conditional probability:

\[
P(A \cap B) = P(A|B)P(B) = P(B|A)P(A) \quad \text{for } P(A \neq 0) \text{ and } P(B \neq 0).
\]

**Theorem: Multiplication Rule for more than 2 events:** Let \( A_1, A_2, \ldots, A_n \) be events over \( S \). then

\[
P(A_1 \cap A_2 \cap \ldots, A_n) = P(A_1)P(A_2|A_1) \cdots P(A_{n-1}|A_1 \cap \ldots, A_{n-2})P(A_n|A_1 \cap \ldots, A_{n-1}).
\]

**Definition:** Sets \( B_1, B_2, \ldots, B_n \) form a partition of the sample space \( S \) if:

1. They "cover" \( S \), that is \( B_1 \cup B_2 \cup \ldots \cup B_n = S \), and
2. They are pairwise disjoint.

**Theorem: Total Probability formula:** Let the sets \( B_1, B_2, \ldots, B_n \) form a partition of the sample space \( S \). Let \( A \) be an event over \( S \). Then

\[
P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).
\]

(2)

**Theorem: Bayes Formula:** (1) For any events \( A \) and \( B \) defined on sample space \( S \) and such that \( P(B) \neq 0 \) we have:

\[
P(A|B) = \frac{P(B|A)P(A)}{P(B)},
\]

(3)

(2) More generally, if the sets \( B_1, B_2, \ldots, B_n \) form a partition of the sample space \( S \), we have

\[
P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^{n} P(A|B_i)P(B_i)},
\]

(4)

for every \( j = 1, \ldots, n \).
INDEPENDENCE

Definition: Two events $A$ and $B$ are called independent if $P(A \cap B) = P(A)P(B)$.

NOTE: (1) If $A$ and $B$ are independent, then $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

(2) If $A$ and $B$ are independent, then so are their complements $A^C$ and $B^C$.

Definition: Events $A_1, A_2, \ldots, A_n$ are independent if for every subset of them we have

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

(5)

NOTE: How is independence related to sets being disjoint (i.e., mutually exclusive)?

1. Independence deals with the relationship between the probabilities of events $A$ and $B$, and the probability of their co-occurrence, $P(A \cap B)$. Independence says something about events that can co-occur, whereas disjoint events, by definition, never co-occur.

2. The notion of sets being disjoint relates to the contents of those events, and whether or not they have shared outcomes (i.e, whether or not they have an empty intersection). Mutual exclusivity describes which events cannot co-occur. Intuition should tell us that disjoint sets are NOT independent! Why? Suppose two events are disjoint. Then knowledge of one event occurring tells you quite a bit of information about whether or not the other has occurred (by definition, it has not!). For example, if you are 22 years old, I know that you are not 21 years old. In fact, disjoint sets cannot be independent except in the trivial case where one or both events has probability zero: since $P(A \cap B) = 0$ for disjoint events, they can only satisfy the definition of independence ($P(A \cap B) = P(B)P(A)$) if $P(A) = 0$ or $P(B) = 0$ (or both are true).

Example: Consider the experiment defined by one card out of a standard 52 card deck.

Let event $A$ be that the card is red (i.e., $A$ is the set of all 26 red cards) and $B$ be the event that the card is a king (i.e., $B$ is all four kings).

Are $A$ and $B$ independent? Check that they satisfy the definition, $P(A \cap B) = P(A)P(B)$:

$$P(A \cap B) = P(\text{red king}) = \frac{1}{26}$$

$$P(A)P(B) = \frac{1}{2} \cdot \frac{4}{52} = \frac{1}{26}$$

So they are independent events!

Are they disjoint? No. Their intersection $A \cap B = \text{red king}$ is not empty.
**COMBINATORICS:** counting, ordering, arranging

**Multiplication Rule:** If operation \( A \) can be performed in \( n \) different ways and operation \( B \) can be performed in \( m \) different ways, then the sequence of these two operations (say, \( AB \)) can be performed in \( n \cdot m \) ways.

**Extension of Multiplication Rule to \( k \) operations:** If operations \( A_i, i = 1, \ldots, k \) can be performed in \( n_i \) different ways, then the ordered sequence (operation \( A_1 \), operation \( A_2 \), \ldots, operation \( A_k \)) can be performed in \( n_1n_2\cdots n_k \) ways.

**Permutations:** An arrangement of \( k \) objects in a row is called a *permutation of length \( k \).*

**Number of permutations of \( k \) elements chosen from a set on \( n \) elements:** The number of permutations of length \( k \), that can be formed from a set of \( n \) distinct objects is

\[
n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}.
\]  

(6)

**Number of permutations of \( n \) elements chosen from a set on \( n \) elements:** The number of permutations of length \( n \) (ordered sequences of length \( n \)), that can be formed from a set of \( n \) distinct objects is

\[
n(n-1)(n-2)\cdots(1) = n!.
\]  

(7)

**Approximation for \( n! \) (Stirling’s Formula):** \( n! \approx \sqrt{2\pi n} \frac{n^{n+1/2}e^{-n}}{n!} \).

**Number of permutations of elements that are not all different:** The number of permutations of length \( n \), that can be formed from a set of \( n_1 \) objects of type 1, \( n_2 \) objects of type 2, \ldots, \( n_k \) objects of type \( k \), where \( \sum_{i=1}^{k} n_i = n \), is

\[
\frac{n!}{n_1!n_2!\cdots n_k!}.
\]  

(8)

**Combinations:** A set of \( k \) unordered objects is called a *combination of size \( k \).*

**Number of combinations of size \( k \) of \( n \) distinct objects:** The number of ways to form combinations of size \( k \) from a set of \( n \) distinct objects, no repetitions, is denoted by the Newton symbol (or *binomial coefficient*) \( \binom{n}{k} \), and equal to

\[
\binom{n}{k} = \frac{n!}{k!(n-k)!}.
\]  

(9)
NOTE: The number of combinations of size \( k \) of \( n \) distinct objects is the number of different subsets of size \( k \) formed from a set of \( n \) elements.

**Combinatorial probabilities - classical definition of probability:** Suppose there are \( n \) simple outcomes in a sample space \( S \). Let event \( A \) consist of \( m \) of those outcomes. Suppose also that all outcomes are equally likely. Then, the probability of event \( A \) is defined as \( P(A) = \frac{m}{n} \).