Introduction to Probability

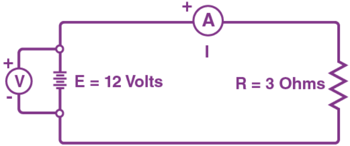
P. Sam Johnson

NITK, Surathkal, India

Mathematical Models - Deterministic Model

By this we shall mean a model which stipulates that the conditions under which an experiment is performed **determine** the outcome of the experiment.

For example, if we insert a battery into a simple circuit, the mathematical model which would presumably describe the observable flow of current would be $I = E/R$, that is, Ohm's law.



A simple circuit diagram

Mathematical Models - Deterministic Model

The model predicts the value of I as soon as E and R are given. Saying it differently, if the above experiment were repeated a number of times, each time using the same circuit (that is, keeping E and R fixed), we would presumably expect to observe the same value for I.

Any deviations that might occur would be so small that for most purposes the above description (that is, model) would suffice. The point is that the particular battery, wire, and ammeter used to generate and to observe the current, and our ability to use the measuring instrument, determine the outcome on each repetition.

Probabilistic Models

There are certain factors which may well be different from repetition to repetition that will, however, not affect the outcome in a noticeable way. For instance, the temperature and humidity in the laboratory, or the height of the person reading the ammeter can reasonably be assumed to have no influence on the outcome.

There are many examples of "experiments" in nature for which deterministic models are appropriate.

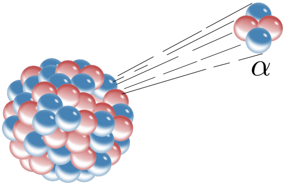
However, there are also many phenomena which require a different mathematical model for their investigation.

These are what we shall call **nondeterministic** or **probabilistic** models. Another quite commonly used term is **stochastic** model.

A probabilistic model is a mathematical description of an uncertain situation.

Mathematical Models - Radioactive Material Emitting Particles

Suppose that we have a piece of radioactive material which is emitting α-particles.



With the aid of a counting device we may be able to record the number of such particles emitted during a specified time interval. **It is clear that we cannot predict precisely the number of particles emitted**, even if we knew the exact shape, dimension, chemical composition, and mass of the object under consideration. Thus there seems to be no reasonable deterministic model yielding the number of particles emitted, say n, as a function of various pertinent characteristics of the source material. We must consider, instead, a probabilistic model.

Mathematical Models - How much precipitation will fall?

For another illustration consider the following meteorological situation. **We wish to determine how much precipitation will fall as a result of a particular storm system passing through a specified locality.** Instruments are available with which to record the precipitation that occurs.

Meteorological observations may give us considerable information concerning the approaching storm system: barometric pressure at various points, changes in pressure, wind velocity, origin and direction of the storm, and various pertinent high-altitude readings. But this information, valuable as it may be for predicting the general nature of the precipitation (light, medium, or heavy, say), simply does not make it possible to state very accurately **how much** precipitation will fall.

Again we are dealing with a phenomenon which does not lend itself to a deterministic approach. A probabilistic model describes the situation more accurately.

Probability

Probability is a very useful concept, but can be interpreted in a number of ways. As an illustration, consider the following.

A patient is admitted to the hospital and a potentially life-saving drug is administered. The following conversation takes place between the nurse and a concerned relative.

RELATIVE	:	Nurse, what is the probability that the drug will work?
NURSE	:	I hope it works, we'll know tomorrow.
RELATIVE	:	Yes, but what is the probability that it will?
NURSE	:	Each case is different, we have to wait.
RELATIVE	:	But let's see, out of a hundred patients that are treated under similar conditions, how many times would you expect it to work?
NURSE	:	I told you, every person is different, for some it works, for some it doesn't.
RELATIVE	:	Then tell me, if you had to bet whether it will work or not, which side of the bet would you take?
NURSE	:	I'd bet it will work.
RELATIVE	:	OK, now, would you be willing to lose two dollars if it doesn't work, and gain one dollar if it does?
NURSE	:	What a sick thought! You are wasting my time!

Probability

In this conversation, the relative attempts to use the concept of probability to discuss an **uncertain** situation. The nurse's initial response indicates that the meaning of "probability" is not uniformly shared or understood, and the relative tries to make it more concrete.

The first approach is to define probability in terms of frequency of occurrence, as a percentage of successes in a moderately large number of similar situations. Such an interpretation is often natural. For example, when we say that a perfectly manufactured coin lands on heads "with probability 50%," we typically mean "roughly half of the time."

Probability

The last part of the earlier conversation was an attempt to infer the nurse's beliefs in an indirect manner. Since the nurse was willing to accept a one-for-one bet that the drug would work, we may infer that the probability of success was judged to be at least 50%.

Had the nurse accepted the last proposed bet (two-for-one), this would have indicated a success probability of at least $2/3$.

Summary

Let us simply state that in a deterministic model it is supposed that the actual outcome (whether numerical or otherwise) is determined from the conditions under which the experiment or procedure is carried out. **In a nondeterministic model, however, the conditions of experimentation determine only the probabilistic behavior (more specifically, the probabilistic law) of the observable outcome.**

Saying it differently, in a deterministic model we use "physical considerations" to predict the outcome, while in a probabilistic model we use the same kind of considerations to specify a probability distribution.

Introduction to Sets

In order to discuss the basic concepts of the probabilistic model which we wish to develop, it will be very convenient to have available some ideas and concepts of the mathematical theory of sets.

A **set** is a collection of objects.

The individual objects making up the collection of the set A are called **members** or **elements** of A.

We define the **universal set** as the set of all objects under consideration. This set is usually designated by U.

We define the **empty** or **null** set to be the set containing no members. We usually designate this set by $\emptyset$.

We say that two sets are the same, $A = B$, if and only if $A \subset B$ and $B \subset A$. That is, two sets are **equal** if and only if they contain the same members.

1. For every set A, we have $\emptyset \subset A$.
2. Once the universal set has been agreed upon, then for every set A considered in the context of U, we have $A \subset U$.

Next we consider the important idea of **combining** given sets in order to form a new set. Two basic operations are considered.
We define C as the **union** of A and B (sometimes called the sum of A and B) as follows:

$$C = \{x : x \in A \quad \text{or} \quad x \in B \text{ (or both)}\}.$$

We write this as $C = A \cup B$. Thus C consists of all elements which are in A, or in B, or in both.

We define D as the **intersection** of A and B (sometimes called the product of A and B) as follows:

$$D = \{x : x \in A \text{ and } x \in B\}.$$

We write this as $D = A \cap B$. Thus D consists of all elements which are in A and in B.

Finally we introduce the idea of the **complement** of a set A with respect to the universal set U as follows: The set, denoted by $\overline{A}$, or A^c, consisting of all elements **not** in A (but in the universal set U) is called the complement of A. That is, $\overline{A} = \{x : x \notin A\}$.

A graphic device known as a **Venn diagram** can be used to considerable advantage when we are combining sets as indicated above.

Note that in describing a set (such as $A \cup B$) we list an element exactly once.

The above operations of union and intersection defined for just two sets may be extended in an obvious way to any finite number of sets. Thus we define $A \cup B \cup C$ as $A \cup (B \cup C)$ or $(A \cup B) \cup C$, which are the same, as can easily be checked. Similarly, we define $A \cap B \cap C$ as $A \cap (B \cap C)$ or $(A \cap B) \cap C$, which again can be checked to be the same. And it is clear that we may continue these constructions of new sets for **any finite** number of given sets.

We asserted that certain sets were the same, for example $A \cap (B \cap C)$ and $(A \cap B) \cap C$. It turns out that there are a number of such **equivalent** sets, some of which are listed below. If we recall that two sets are the same whenever they contain the same members, it is easy to show that the assertions stated are true.

(a) $A \cup B = B \cup A$,

(b) $A \cap B = B \cap A$,

(c) $A \cup (B \cup C) = (A \cup B) \cup C$,

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

We refer to (a) and (b) as the **commutative** laws, and (c) and (d) as the **associative** laws.

There are a number of other such **set identities** involving union, intersection, and complementation. The most important of these are listed below. In each case, their validity may be checked with the aid of a Venn diagram.

(e) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

(f) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,

(g) $A \cap \emptyset = \emptyset$,

(h) $A \cup \emptyset = A$,

(i) $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$,

(j) $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$,

(k) $\overline{\overline{A}} = A$.

We note that (g) and (h) indicate that $\emptyset$ behaves among sets (with respect to the operations $\cup$ and $\cap$) very much as does the number zero among numbers (with respect to the operation of addition and multiplication).

De Morgan's Laws

Two particularly useful properties are given by **De Morgan's laws** which state that

$$\left(\bigcup_{n=1}^{\infty} A_n\right)^c = \bigcap_{n=1}^{\infty} A_n^c, \qquad \left(\bigcap_{n=1}^{\infty} A_n\right)^c = \bigcup_{n=1}^{\infty} A_n^c.$$

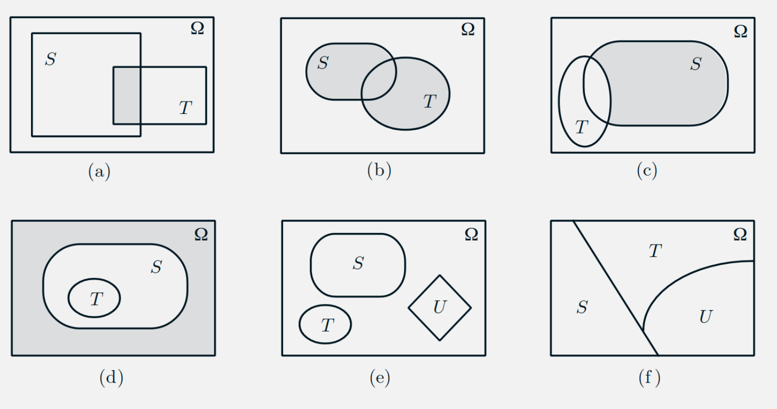


Figure 1.1: Examples of Venn diagrams. (a) The shaded region is $S \cap T$. (b) The shaded region is $S \cup T$. (c) The shaded region is $S \cap T^c$. (d) Here, $T \subset S$. The shaded region is the complement of S. (e) The sets S, T, and U are disjoint. (f) The sets S, T, and U form a partition of the set Ω.

Definition 1.

Let A and B be two sets. By the Cartesian product *of A and B, denoted by $A \times B$, we shall mean the set $\{(a,b), a \in A, b \in B\}$, that is, the set of all ordered pairs where the first element is taken from A and the second from B.*

Note: In general $A \times B \neq B \times A$.

The above notion can be extended as follows: If $A_1, \ldots, A_n$ are sets, then $A_1 \times A_2 \times \cdots \times A_n = \{(a_1, a_2, \ldots, a_n) : a_i \in A_i\}$, that is, the set of all ordered n-tuples.

The Euclidean plane, $\mathbb{R} \times \mathbb{R}$, where $\mathbb{R}$ is the set of all real numbers and Euclidean 3-space represented as $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$.

$\mathbb{R}^n$ is the n-dimensional Euclidean space.

The **number of elements** in a set will be of considerable importance to us. If there is a finite number of elements in A, say $a_1, a_2, \ldots, a_n$, we say that A is **finite**. If there is an infinite number of elements in A which may be put into a **one-to-one correspondence** with the positive integers, we say that A is **countably** or **denumerably infinite**. It can be shown for example, that the set of all rational numbers is countably infinite.

Finally we must consider the case of a **nondenumerable infinite set** (called **an uncountable set**). Such sets contain an infinite number of elements which cannot be enumerated. It can be shown, for instance, that for any two real numbers $b > a$, set $A = \{x : a \leq x \leq b\}$ has a nondenumerable number of elements. Since we may associate with each real number a point on the real number line, the above says that any (nondegenerate) interval contains more than a countable number of points.

Examples of Nondeterministic Experiments

We are now ready to discuss what we mean by a "random" or "nondeterministic" experiment. We shall give examples of phenomena for which nondeterministic models are appropriate.

Thus we shall repeatedly refer to nondeterministic or random experiments when in fact we are talking about a **nondeterministic model** for an experiment. We shall not attempt to give a precise dictionary definition of this concept. Instead, we shall cite a large number of examples.

Examples of Nondeterministic Experiments

E_1: Toss a die and observe the number that shows on top.

E_2: Toss a coin four times and observe the total number of heads obtained.

E_3: Toss a coin four times and observe the sequence of heads and tails obtained.

Examples of Nondeterministic Experiments

E_4: Manufacture items on a production line and count the number of defective items produced during a 24-hour period.

E_5: An airplane wing is assembled with a large number of **rivets**. The number of defective rivets is counted.

Examples of Nondeterministic Experiments

E_6: A light bulb is manufactured. It is then tested for its life length by inserting it into a socket and the time elapsed (in hours) until it burns out is recorded.

E_7: A lot of 10 items contains 3 defectives. One item is chosen after another (without replacing the chosen item) until the last defective item is obtained. The total number of items removed from the lot is counted.

E_8: Items are manufactured until 10 nondefective items are produced. The total number of manufactured items is counted.

Examples of Nondeterministic Experiments

E_9: A missile is launched. At a specified time t, its three velocity components, v_x, v_y, and v_z are observed.

E_{10}: A newly launched missile is observed at times, $t_1, t_2, \ldots, t_n$. At each of these times the missile's height above the ground is recorded.

Examples of Nondeterministic Experiments

E_{11}: The tensile strength of a steel beam is measured.

E_{12}: From an urn containing only black balls, a ball is chosen and its color noted.

Examples of Nondeterministic Experiments

E_{13}: A thermograph records temperature, continuously, over a 24-hour period. At a specified locality and on a specified date, such a thermograph is "read."

E_{14}: In the situation described in E_{13}, x and y, the minimum and maximum temperatures of the 24-hour period in question are recorded.

Examples of Nondeterministic Experiments

What do the above experiments have in common? The following features are pertinent for our characterization of a **random experiment**.

(a) Each experiment is capable of being repeated indefinitely under **essentially unchanged conditions**.

(b) Although we are in general not able to state what a **particular** outcome will be, we are able to describe the set of **all possible** outcomes of the experiment.

(c) As the experiment is performed repeatedly, the individual outcomes seem to occur in a haphazard manner. However, as the experiment is repeated a **large** number of times, a definite pattern or regularity appears. It is this regularity which makes it possible to construct a precise mathematical model with which to analyze the experiment. We need only think of the repeated tossings of a fair coin. Although heads and tails will appear, successively, in an almost arbitrary fashion, it is a well-known empirical fact that after a large number of tosses the proportion of heads and tails will be approximately equal.

Note that experiment E_{12} (from an urn containing only black balls, a ball is chosen and its color noted) has the peculiar feature that only one outcome is possible.

Examples of Nondeterministic Experiments

In describing the various experiments we have specified not only the procedure which is being performed but also what we are interested in observing.

For example, there is a difference between E_2 (toss a coin four times and observe the total number of heads obtained) and E_3 (toss a coin four times and observe the sequence of heads and tails obtained). This is a very important point to which we shall refer again later when discussing random variables.

Examples of Nondeterministic Experiments

Let us simply note that as a consequence of a single experimental procedure or the occurrence of a single phenomenon, **several** different numerical values could be computed.

For instance, if one person is chosen from a large group of persons (and the actual choosing would be the experimental procedure previously referred to), we might be interested in that person's height, weight, annual income, number of children, etc.

Of course in most situations we know before beginning our experimentation just what numerical characteristic we are going to be concerned about.

The Sample Space

Definition 2.

With each experiment E of the type we are considering we define the sample space *as the set of all possible outcomes of E. We usually designate this set by S which represents the universal set.*

The sample space *S is the set of all possible outcomes of an experiment.*

Consider an experiment whose outcome is not predictable with certainty. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by S.

Some examples follow.

Example 3.

If the outcome of an experiment consists in the determination of the sex of a newborn child, then $S = \{g, b\}$, where the outcome g means that the child is a girl and b that it is a boy.

Example 4.

If the outcome of an experiment is the order of finish in a race among the 7 horses having post positions $1, 2, 3, 4, 5, 6, 7$, then

$$S = \{\textit{all } 7! \textit{ permutations of } (1, 2, 3, 4, 5, 6, 7)\}.$$

The outcome $(2, 3, 1, 6, 5, 4, 7)$ means, for instance, that the number 2 horse comes in first, then the number 3 horse, then the number 1 horse, and so on.

Example 5.

If the experiment consists of flipping two coins, then the sample space consists of the following four points:

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

The outcome will be (H, H) *if both coins are heads,* (H, T) *if the first coin shows a head and the second shows a tail,* (T, H) *if the first is a tail and the second is a head, and* (T, T) *if both coins show tails.*

Example 6.

If the experiment consists of tossing two dice, then the sample space consists of the 36 points

$$S = \{(i, j) : i, j = 1, 2, 3, 4, 5, 6\}$$

where the outcome (i, j) is said to occur if i appears on the leftmost die and j on the other die.

Example 7.

If the experiment consists of measuring (in hours) the lifetime of a transistor, then the sample space consists of all nonnegative real numbers; that is,

$$S = \{x : 0 \leq x < \infty\}.$$

Any subset E of the sample space is known as an event. In other words, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred. Following are some examples of events.

In the preceding Example 3, if $A = \{g\}$, then A is the event that the child is a girl. Similarly, if $B = \{b\}$, then B is the event that the child is a boy.

Sample Space and Events

In Example 4, if

$$A = \{\text{all outcomes in } S \text{ starting with a } 3\}$$

then A is the event that horse 3 wins the race.

In Example 5, if $A = \{(H, H), (H, T)\}$, then A is the event that a head appears on the first coin.

In Example 6, if $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then A is the event that the sum of the dice equals 7.

In Example 7, if $A = \{x : 0 \leq x \leq 5\}$, then A is the event that the transistor does not last longer than 5 hours.

In Example 5, if $A = \{(H, H), (H, T)\}$ and $B = \{(T, H)\}$, then

$$A \cup B = \{(H, H), (H, T), (T, H)\}.$$

Thus, $A \cup B$ would occur if a head appeared on either coin.

For instance, in Example 5, if $A = \{(H, H), (H, T), (T, H)\}$ is the event that at least 1 head occurs and $B = \{(H, T), (T, H), (T, T)\}$ is the event that at least 1 tail occurs, then

$$A \cap B = \{(H, T), (T, H)\}$$

is the event that exactly 1 head and 1 tail occur.

Sample Space and Events

In Example 6, if $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ is the event that the sum of the dice is 7 and $B = \{(1,5),(2,4),(3,3),(4,2),(5,1)\}$ is the event that the sum is 6, then the event $A \cap B$ does not contain any outcomes and hence could not occur.

To give such an event a name, we shall refer to it as the null event and denote it by $\emptyset$. (That is, $\emptyset$ refers to the event consisting of no outcomes.) If $A \cap B = \emptyset$, then A and B are said to be *mutually exclusive*.

A^c will occur if and only if A does not occur. In Example 6, if event $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$, then A^c will occur when the sum of the dice does not equal 7.

Note that because the experiment must result in some outcome, it follows that $S^c = \emptyset$.

Sample Space and Events

For any two events A and B, if all of the outcomes in A are also in B, then we say that A is *contained* in B, or A is a *subset* of B, and write $A \subset B$ (or equivalently, $B \supset A$, which we sometimes say as B is a *superset* of A).

Thus, if $A \subset B$, then the occurrence of A implies the occurrence of B. If $A \subset B$ and $B \subset A$, we say that A and B are equal and write $A = B$.

Sample Space and Events

A graphical representation that is useful for illustrating logical relations among events is the **Venn diagram**.

The sample space S is represented as consisting of all the outcomes in a large rectangle, and the events $E, F, G, \dots$ are represented as consisting of all the outcomes in given circles within the rectangle.

Events of interest can then be indicated by shading appropriate regions of the diagram.

The operations of forming unions, intersections, and complements of events obey certain rules similar to the rules of algebra. We list a few of these rules:

Commutative laws	$A \cup B = B \cup A$	$A \cap B = B \cap A$
Associative laws	$(A \cup B) \cup C = A \cup (B \cup C)$	$(A \cap B) \cap C = A \cap (B \cap C)$
Distributive laws	$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$	$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$

These relations are verified by showing that any outcome that is contained in the event on the left side of the equality sign is also contained in the event on the right side, and vice versa. One way of showing this is by means of Venn diagrams.

Exercise 8.

Verify the distributive law by a sequence of Venn diagrams.

The following useful relationships between the three basic operations of forming unions, intersections, and complements are known as *De Morgan's laws*:

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c$$
$$\left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

Exercise 9.

Prove De Morgan's Laws.

Sample Space and Events

Let us consider some experiments and describe a sample space for each. The sample space S_i will refer to the experiment E_i.

E_1: Toss a die and observe the number that shows on top.

S_1: $\{1, 2, 3, 4, 5, 6\}$.

E_2: Toss a coin four times and observe the total number of heads obtained.

S_2: $\{0, 1, 2, 3, 4\}$.

E_3: Toss a coin four times and observe the sequence of heads and tails obtained.

S_3: $\{$all possible sequences of the form $a_1, a_2, a_3, a_4,$ where each $a_i =$ H or T depending on whether heads or tails appeared on the ith toss.$\}$

E_4: Manufacture items on a production line and count the number of defective items produced during a 24-hour period.

S_4: $\{0, 1, 2, \ldots, N\}$, where N is the maximum number that could be produced in 24 hours.

E_5: An airplane wing is assembled with a large number of **rivets**. The number of defective rivets is counted.

S_5: $\{0, 1, 2, \ldots, M\}$, where M is the number of rivets installed.

E_6: A light bulb is manufactured. It is then tested for its life length by inserting it into a socket and the time elapsed (in hours) until it burns out is recorded.

S_6: $\{t : t \geq 0\}$.

E_7: A lot of 10 items contains 3 defectives. One item is chosen after another (without replacing the chosen item) until the last defective item is obtained. The total number of items removed from the lot is counted.

S_7: $\{3, 4, 5, 6, 7, 8, 9, 10\}$.

E_8: Items are manufactured until 10 nondefective items are produced. The total number of manufactured items is counted.

S_8: $\{10, 11, 12, \ldots\}$.

The Sample Space

E_9: A missile is launched. At a specified time t, its three velocity components, v_x, v_y, and v_z are observed.
S_9: $\{(v_x, v_y, v_z) : v_x, v_y, v_z \text{ real numbers}\}$.

S_{10}: $\{(h_1, \ldots, h_n) : h_i \geq 0, i = 1, 2, \ldots, n\}$.
E_{10}: A newly launched missile is observed at times, $t_1, t_2, \ldots, t_n$. At each of these times the missile's height above the ground is recorded.

E_{11}: The tensile strength of a steel beam is measured.
S_{11}: $\{T : T \geq 0\}$.

E_{12}: From an urn containing only black balls, a ball is chosen and its color noted.
S_{12}: $\{\text{black ball}\}$.

The Sample Space

E_{13}: A thermograph records temperature, continuously, over a 24-hour period. At a specified locality and on a specified date, such a thermograph is "read."

S_{13}: This sample space is the most involved of those considered here. We may realistically suppose that the temperature at a specified locality can never get above or below certain values, say M and m. Beyond this restriction, we must allow the possibility of any graph to appear with certain qualifications. Presumably the graph will have no jumps (that is, it will represent a continuous function). In addition, the graph will have certain characteristics of smoothness which can be summarized mathematically by saying that the graph represents a differentiable function. Thus we can finally state that the sample space is

$$\{f : f \text{ a differentiable function, satisfying } m \leq f(t) \leq M, \text{ all } t\}.$$

The Sample Space

E_{14}: In the situation described in E_{13}, x and y, the minimum and maximum temperatures of the 24-hour period in question are recorded.

S_{14}: $\{(x, y) : m \leq x \leq y \leq M\}$. That is, S_{14} consists of all points in and on a triangle in the two-dimensional x, y-plane.

The Sample Space

We will not concern ourselves with sample spaces of the complexity encountered in S_{13}. However, such sample spaces do arise, but require more advanced mathematics for their study than we are presupposing.

In order to describe a sample space associated with an experiment, we must have a very clear idea of what we are measuring or observing. Hence we should speak of "a" sample space associated with an experiment rather than "the" sample space. In this connection note the difference between S_2 and S_3.

The Sample Space

Note also that the outcome of an experiment need not be a number. For example, in E_3 each outcome is a sequence of H's and T's. In E_9 and E_{10} each outcome consists of a vector, while in E_{13} each outcome consists of a function.

It will again be important to discuss the **number** of outcomes in a sample space. Three possibilities arise: the sample space may be finite, countably infinite, or noncountably infinite.

Referring to the above examples, we note that $S_1, S_2, S_3, S_4, S_5, S_7$, and S_{12} are finite, S_8 is countably infinite, and $S_6, S_9, S_{10}, S_{11}, S_{13}$, and S_{14} are noncountably infinite.

The Sample Space

At this point it might be worth while to comment on the difference between a mathematically "idealized" sample space and an experimentally realizable one.

For this purpose, let us consider experiment E_6 (A light bulb is manufactured. It is then tested for its life length by inserting it into a socket and the time elapsed (in hours) until it burns out is recorded) and its associated sample space S_6. **It is clear that when we are actually recording the total time t during which a bulb is functioning, we are "victims" of the accuracy of our measuring instrument.** Suppose that we have an instrument which is capable of recording time to two decimal places, for example, 16.43 hours. With this restriction imposed, our sample space becomes **countably infinite:** $\{0.0, 0.01, 0.02, \ldots\}$.

The Sample Space

Furthermore, it is quite realistic to suppose that no bulb could possibly last more than H hours, where H might be a very large number. Thus it appears that if we are completely realistic about the description of this sample space, we are actually dealing with a **finite** sample space: $\{0.0, 0.01, 0.02, \ldots, H\}$. The total number of outcomes would be $(H/0.01) + 1$, which would be a very large number if H is even moderately large, for example, $H = 100$. It turns out to be far simpler and convenient, mathematically, to assume that **all** values of $t \geq 0$ are possible outcomes and hence to deal with the sample space S_6 as originally defined.

In view of the above comments, a number of the sample spaces described are idealized. In all subsequent situations, the sample space considered will be that one which is mathematically most convenient. In most problems, little question arises as to the proper choice of sample space.

Sample Spaces and Events

Every probabilistic model involves an underlying process, called the experiment, that will produce exactly one out of several possible outcomes. The set of all possible outcomes is called the **sample space** of the experiment, and is denoted by S.

Another basic notion is the concept of an **event**. An event A (with respect to a particular sample space S associated with an experiment E) is simply a set of possible outcomes. In set terminology, an event is a **subset** of the sample space S.

S itself is an event and so is the empty set $\emptyset$. Any individual outcome may also be viewed as an event.

Any collection of possible outcomes, including the entire sample space S and its complement, the empty set $\emptyset$, may qualify as an event. **Strictly speaking, however, some sets have to be excluded.**

In particular, when dealing with probabilistic models involving an uncountably infinite sample space, there are certain unusual subsets for which one cannot associate meaningful probabilities. This is an intricate technical issue, involving the mathematics of measure theory. Fortunately, such pathological subsets do not arise in the problems considered in this course, and the issue can be safely ignored.

Sample Spaces and Events

There is no restriction on what constitutes an experiment. For example, it could be a single toss of a coin, or three tosses, or an infinite sequence of tosses. However, it is important to note that in our formulation of a probabilistic model, there is only one experiment. So, **three tosses of a coin constitute a single experiment, rather than three experiments.**

The sample space of an experiment may consist of a finite or an infinite number of possible outcomes. Finite sample spaces are conceptually and mathematically simpler. Still, sample spaces with an infinite number of elements are quite common. As an example, consider throwing a dart on a square target and viewing the point of impact as the outcome.

Choosing an Appropriate Sample Space

Regardless of their number, different elements of the sample space should be distinct and **mutually exclusive**, so that when the experiment is carried out there is a unique outcome. For example, the sample space associated with the roll of a die cannot contain "1 or 3" as a possible outcome and also "1 or 4" as another possible outcome. If it did, we would not be able to assign a unique outcome when the roll is a 1.

A given physical situation may be modeled in several different ways, depending on the kind of questions that we are interested in. Generally, the sample space chosen for a probabilistic model must be **collectively exhaustive**, in the sense that no matter what happens in the experiment, we always obtain an out come that has been included in the sample space. In addition, the sample space should have enough detail to distinguish between all outcomes of interest to the modeler, while avoiding irrelevant details.

Choosing an Appropriate Sample Space

Example 10.

Consider two alternative games, both involving ten successive coin tosses:
Game 1: We receive \$1 *each time a head comes up.*
Game 2: We receive \$2 *for every coin toss, up to and including the first time a head comes up. Then, we receive* \$2 *for every coin toss, up to the second time a head comes up. More generally, the dollar amount per toss is doubled each time a head comes up.*

In game 1, it is only the total number of heads in the ten-toss sequence that maters. while in game 2, the order of heads and tails is also important. Thus, in a probabilistic model for game 1, we can work with a sample space consisting of eleven possible outcomes, namely, $\{1, 2, \ldots, 10\}$. In game 2, a finer grain description of the experiment is called for, and it is more appropriate to let the sample space consist of every possible ten-long sequence of heads and tails.

Events

The following are some examples of events. A_i will refer to an event associated with the experiment E_i.

E_1: Toss a die and observe the number that shows on top.
A_1: An even number occurs; that is, $A_1 = \{2, 4, 6\}$.

E_2: Toss a coin four times and observe the total number of heads obtained.
A_2: Two heads occur ; that is $\{2\}$.

Events

E_3: Toss a coin four times and observe the sequence of heads and tails obtained.

A_3: More heads than tails showed ; that is, $\{HHHH, HHHT, HHTH, HTHH, THHH\}$.

E_4: Manufacture items on a production line and count the number of defective items produced during a 24-hour period.

A_4: All items were nondefective ; that is, $\{0\}$.

E_5: An airplane wing is assembled with a large number of **rivets**. The number of defective rivets is counted.

A_5: More than two rivets were defective ; that is, $\{3, 4, \ldots, M\}$.

E_6: A light bulb is manufactured. It is then tested for its life length by inserting it into a socket and the time elapsed (in hours) until it burns out is recorded.

A_6: The bulb burns less than three hours ; that is, $\{t : t < 3\}$.

E_{14}: A thermograph records temperature, continuously, over a 24-hour period. At a specified locality and on a specified date, such a thermograph is "read." In the situation described above, x and y, the minimum and maximum temperatures of the 24-hour period in question are recorded.

A_{14} The maximum is 20° greater than the minimum ; that is, $\{(x, y) : y = x + 20\}$.

Sample Space and Events

When the sample space S is finite or countably infinite, every subset may be considered as an event. If S has n members, there are exactly 2^n subsets (events).

However, if S is noncountably infinite, a theoretical difficulty arises. It turns out that not every conceivable subset may be considered as an event. Certain "nonadmissible" subsets must be excluded for reasons which are beyond the level of this presentation.

Fortunately such nonadmissible sets do not really arise in applications and hence will not concern us here. In all that follows it will be tacitly assumed that whenever we speak of an event it will be of the kind we are allowed to consider.

We can now use the various methods of combining sets (that is, events) and obtain the new sets (that is, events).

(a) If A and B are events, $A \cup B$ is the event which occurs if and only if A or B (or both) occur.

(b) If A and B are events, $A \cap B$ is the event which occurs if and only if A and B occur.

(c) If A is an event, $\bar{A}$ is the event which occurs if and only if A does not occur.

(d) If $A_1, \ldots, A_n$ is any finite collection of events, then $\cup_{i=1}^{n} A_i$ is the event which occurs if and only if at least one of the events A_i occurs.

(e) If $A_1, \ldots, A_n$ is any finite collection of events, then $\cap_{i=1}^{n} A_i$ is the event which occurs if and only if all the events A_i occur.

(f) If $A_1, \ldots, A_n, \ldots$ is any (countably) infinite collection of events, then $\cup_{i=1}^{\infty} A_i$ is the event which occurs if and only if at least one of the events A_i occur.

(h) Suppose that S represents the sample space associated with some experiment E and we perform E twice. Then $S \times S$ may be used to represent all outcomes of these two repetitions. That is, $(s_1, s_2) \in S \times S$ means that s_1 resulted when E was performed the first time and s_2 when E was performed the second time.

(i) The example in (h) may obviously be generalized. Consider n repetitions of an experiment E whose sample space is S. Then $S \times S \times \cdots \times S = \{(s_1, s_2, \ldots, s_n) : s_i \in S, i = 1, \ldots, n\}$ represents the set of all possible outcomes when E is performed n times. In a sense, $S \times S \times \cdots \times S$ is a sample space itself, namely the sample space associated with n repetitions of E.

Definition 11.

Two events, A and B, are said to be mutually exclusive if they cannot occur together. We express this by writing $A \cap B = \emptyset$; that is, the intersection of A and B is the empty set.

One of the basic characteristics of the concept of "experiment" is that we do not know which particular outcome will occur when the experiment is performed. Saying it differently, if A is an event associated with the experiment, then we cannot state with certainty that A will or will not occur. **Hence it becomes very important to try to associate a number with the event A which will measure, in some sense, how likely it is that the event A occurs. This task leads us to the theory of probability.**

Relative Frequency

One way of defining the probability of an event is in terms of its relative frequency. Such a definition usually goes as follows: We suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions. For each event A of the sample space S, we define $n(A)$ to be the number of times in the first n repetitions of the experiment that the event A occurs. Then $P(A)$, the probability of the event A, is defined as

$$P(A) = \lim_{n \to \infty} \frac{n(A)}{n}.$$

The null event has probability 0 of occurring.

Relative Frequency

Suppose that we repeat the experiment E n times and let A and B be two events associated with E. We let $n(A)$ and $n(B)$ be the number of times that the event A and the event B occurred among the n repetitions, respectively.

Definition 12.

$f_A = n(A)/n$ is called the relative frequency of the event A in the n repetitions of E. The relative frequency f_A has the following important properties, which are easily verified.

(1) $0 \leq f_A \leq 1$.

(2) $f_A = 1$ if and only if A occurs every time among the n repetitions.

(3) $f_A = 0$ if and only if A never occurs among the n repetitions.

(4) If A and B are two mutually exclusive events and if $f_{A \cup B}$ is the relative frequency associated with the event $A \cup B$, then $f_{A \cup B} = f_A + f_B$.

(5) f_A, based on n repetitions of the experiment and considered as a function of n, "converges" in a **certain probabilistic sense** to $P(A)$ as $n \to \infty$.

Let us simply state that Property (5) involves the fairly intuitive notion that the relative frequency based on an increasing number of observations tends to "stabilize" near some definite value. This is not the same as the usual concept of convergence encountered elsewhere in mathematics. In fact, as stated here, this is not a mathematical conclusion at all but simply an empirical fact.

Relative Frequency

Most of us are intuitively aware of this phenomenon of stabilization although we may never have checked on it. To do so requires a considerable amount of time and patience, since it involves a large number of repetitions of an experiment.

However, sometimes we may be innocent observers of this phenomenon as the following example illustrates.

Example 13.

Suppose that we are standing on a sidewalk and fix our attention on two adjacent slabs of concrete. Assume that it begins to rain in such a manner that we are actually able to distinguish individual raindrops and keep track of whether these drops land on one slab or the other. We continue to observe individual drops and note their point of impact.

Denoting the ith drop by X_i, where $X_i = I$ if the drop lands on one slab and 0 *if it lands on the other slab, we might observe a sequence such as* $1, 1, 0, 1, 0, 0, 0, 1, 0, 0, 1$.

Now it is clear that we are not able to predict where a particular drop will fall. *(Our experiment consists of some sort of meteorological situation causing the release of raindrops.)*

Example (contd...)

If we compute the relative frequency of the event
$A =$ {the drop lands on slab I}, then the above sequence of outcomes gives rise to the following relative frequencies (based on the observance of $1, 2, 3, \ldots$ drops) :

$$1, \frac{2}{2}, \frac{2}{3}, \frac{3}{4}, \frac{3}{5}, \frac{3}{6}, \frac{3}{7}, \frac{4}{8}, \frac{4}{9}, \frac{4}{10}, \frac{5}{11}, \ldots$$

These numbers show a considerable degree of variation, particularly at the beginning. It is intuitively clear that if the above experiment were continued indefinitely, these relative frequencies would stabilize near the value $\frac{1}{2}$. For we have every reason to believe that after some time had elapsed the two slabs would be equally wet.

Example (contd...)

The essence of this property is that if an experiment is performed a large number of times, the relative frequency of occurrence of some event A tends to vary less and less as the number of repetitions is increased. This characteristic is also referred to as **statistical regularity**.

We have also been somewhat vague in our definition of experiment. Just when is a procedure or mechanism an experiment in our sense, capable of being studied mathematically by means of a nondeterministic model?

We have stated previously that an experiment must be capable of being performed repeatedly under essentially unchanged conditions. We can now add another requirement. When the experiment is performed repeatedly it must exhibit the statistical regularity referred to above. Later we shall discuss a theorem (called the **Law of Large Numbers**) which shows that statistical regularity is in fact a consequence of the first requirement: **repeatability**.

To assign a number to each event A which will measure **how likely it is that A occurs** when the experiment is performed. One possible approach might be the following one: Repeat the experiment a large number of times, compute the relative frequency f_A, and use this number. When we recall the properties of f_A, it is clear that this number does give a very definite indication of how likely it is that A occurs. Furthermore, we know that as the experiment is repeated more and more times, the relative frequency f_A stabilizes near some number, say p. However, there are two serious objections to this approach. (a) It is not clear how large n should be before we know the number. 1000? 2000? $10,000$? (b) Once the experiment has been completely described and the event A specified, the number we are seeking should not depend on the experimenter or the particular streak of luck which he experiences.

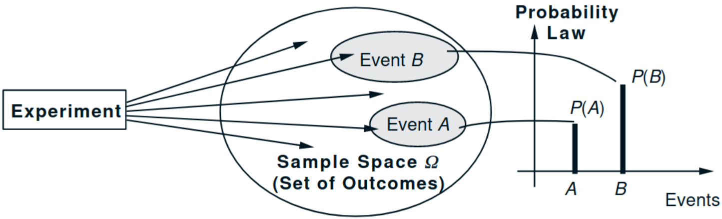
For example, it is possible for a perfectly balanced coin, when tossed 10 times, to come up with 9 heads and 1 tail. The relative frequency of the event $A\{\text{heads occur}\}$ thus equals $\frac{9}{10}$. Yet it is clear that on the next 10 tosses the pattern of heads and tails might be reversed.

What we want is a means of obtaining such a number without resorting to experimentation. Of course, for the number we stipulate to be meaningful, any subsequent experimentation should yield a relative frequency which is "close" to the stipulated value, particularly if the number of repetitions on which the computed relative frequency is based is quite large. We proceed formally as follows.

Elements of a Probabilistic Model

The main ingredients of a probabilistic model are given as follows.

- The **sample space** Ω, which is the set of all possible **outcomes** of an experiment.
- The **probability law**, which assigns to a set A of possible outcomes (also called an **event**) a nonnegative number $P(A)$ (called the **probability** of A) that encodes our knowledge or belief about the collective "likelihood" of the elements of A. The probability law must satisfy certain properties to be introduced shortly.



Definition 14.

Let E be an experiment. Let S be a sample space associated with E. With each event A we associate a real number, designated by $P(A)$ and called the probability of A satisfying the following axioms (called Probability Axioms*):*

1. **(Nonnegativity)** $P(A) \geq 0$, for every event A.
2. **(Additivity)** If A and B are mutually exclusive events ($A_i \cap A_j = \emptyset$ when $i \neq j$), $P(A \cup B) = P(A) + P(B)$.
   More generally, if the sample space has an infinite number of elements and $A_1, A_2, \ldots, A_n, \ldots$ is a sequence of **pairwise** mutually exclusive events ($A_i \cap A_j = \emptyset$ when $i \neq j$), then
   $$P(\cup_{i=1}^{\infty} A_i) = P(A_1) + P(A_2) + \cdots + P(A_i) + \cdots$$
3. **(Normalization)** The probability of the entire sample space S is equal to 1, that is, $P(S) = 1$.

The assumption of the existence of a set function P, defined on the events of a sample space S and satisfying Axioms 1, 2, and 3, constitutes the modern mathematical approach to probability theory.

Hopefully, we will agree that the axioms are natural and in accordance with our intuitive concept of probability as related to chance and randomness.

Furthermore, using these axioms we shall be able to prove that if an experiment is repeated over and over again, then the proportion of time during which any specific event A occurs will equal $P(A)$. This result, known as the strong law of large numbers. In addition, we discuss another possible interpretation of probability – as being a measure of belief.

Technical Remark

We have supposed that $P(A)$ is defined for all the events A of the sample space.

Actually, when the sample space is an uncountably infinite set, $P(A)$ **is defined only for a class of events called measurable. However, this restriction need not concern us, as all events of any practical interest are measurable.**

For instance, if the probability of obtaining a head on the toss of a coin is $\frac{3}{8}$, then the probability of obtaining a tail must be $\frac{5}{8}$. If the event A is contained in the event B, then the probability of A is no greater than the probability of B.

Example 15.

If our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we would have

$$P(\{H\}) = P(\{T\}) = \frac{1}{2}.$$

On the other hand, if the coin were biased and we felt that a head were twice as likely to appear as a tail, then we would have

$$P(\{H\}) = \frac{2}{3} \quad P(\{T\}) = \frac{1}{3}.$$

Example 16.

If a die is rolled and we suppose that all six sides are equally likely to appear, then we would have

$$P(\{1\}) = P(\{2\}) = P(\{3\}) = P(\{4\}) = P(\{5\}) = P(\{6\}) = \frac{1}{6}.$$

From Axiom 2, it would thus follow that the probability of rolling an even number would equal

$$P(\{2,4,6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}.$$

We note that from Axiom 3 it immediately follows that for any finite n,

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

We shall show that the numbers $P(A)$ and f_A are "close" to each other (in a certain sense), if f_A is based on a large number of repetitions. It is this fact which gives us the justification to use $P(A)$ for measuring how probable it is that A occurs.

Before discussing how to evaluate $P(A)$, we prove several consequences concerning $P(A)$ which follow from the above conditions, and which do not depend on how we actually compute $P(A)$.

Theorem 17.

If $\emptyset$ is the empty set, then $P(\emptyset) = 0$.

Proof: We may write, for any event A, $A = A \cup \emptyset$. Since A and $\emptyset$ are mutually exclusive, it follows from Axiom 2 that $P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$. From this the conclusion of the theorem is immediate.

Note : The converse of the above theorem is not true. That is, if $P(A) = 0$, we cannot in general conclude that $A = \emptyset$, for there are situations in which we assign probability zero to an event that can occur.

Zero Probability

A continuous sample space : Consider a probabilistic experiment whose set of possible outcomes, called sample space and denoted by S, is the unit interval $[0, 1]$.

It is possible to assign probabilities in such a way that each sub-interval has probability equal to its length. The proof that such an assignment of probabilities can be consistently performed is beyond the scope of this example, but you can find it in any elementary measure theory book.

All the possible outcomes have zero probability

As a direct consequence of this assignment, all the possible outcomes $s \in S$ have zero probability. Stated differently, every possible outcome is a zero-probability event.

This might seem counterintuitive. In everyday language, a zero-probability event is an event that never happens. However, this example illustrates that a zero-probability event can indeed happen.

Since the sample space provides an exhaustive description of the possible outcomes, one and only one of the sample points $s \in S$ will be the realized outcome.

But we have just demonstrated that all the sample points are zero-probability events: as a consequence, the realized outcome can only be a zero-probability event.

Another counter-intuitive property

Another apparently paradoxical aspect of this probability model is that the sample space S can be obtained as the union of disjoint zero-probability events:

$$S = \bigcup_{s \in S} s$$

where each $s \in S$ is a zero-probability event and all events in the union are disjoint.

If we forgot that the additivity property of probability applies only to countable collections of subsets, we would mistakenly deduce that

$$P(S) = 0$$

and we would come to a contradiction: $P(S) = 0$, when, by the properties of probability, it should be $P(S) = 1$.

Of course, the fallacy in such an argument is that S is not a countable set and, hence, the additivity property cannot be used.

Theorem 18.

If $\bar{A}$ is the complementary event of A, then

$$P(A) = 1 - P(\bar{A}).$$

Proof: We may write $S = A \cup \bar{A}$ and, using Axioms 2 and 3, we obtain $1 = P(A) + P(\bar{A})$.

- In words, Theorem 18 states that the probability that an event does not occur is 1 minus the probability that it does occur. For instance, if the probability of obtaining a head on the toss of a coin is $3/8$, then the probability of obtaining a tail must be $5/8$.
- This is a particularly useful result, for it means that whenever we wish to evaluate $P(A)$ we may instead compute $P(\bar{A})$ and then obtain the desired result by subtraction. We shall see later that in many problems it is much easier to compute $P(\bar{A})$ than $P(A)$.

Basic Notions of Probability

This result may be established easily by using mathematical induction.

Theorem 19.

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

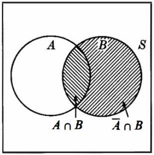
Proof: We may decompose B into two mutually exclusive events as follows: $B = A \cup (B \cap \bar{A})$. Hence $P(B) = P(A) + P(B \cap \bar{A}) \geq P(A)$, since $P(B \cap \bar{A}) \geq 0$ from Property 1.

- This result certainly is intuitively appealing. For it says that if B must occur whenever A occurs, then B is at least as probable as A. That is, it states that if the event A is contained in the event B, then the probability of A is no greater than the probability of B.
- Theorem 19 tells us, for instance, that the probability of rolling a 1 with a die is less than or equal to the probability of rolling an odd value with the die.

Theorem 20.

If A and B are any two events, then

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



Proof: The idea of this proof is to decompose $A \cup B$ and B into mutually exclusive events and then to apply Axiom 2. (See the Venn diagram in the above figure.) Thus we write

$$A \cup B = A \cup (B \cap \bar{A}),$$
$$B = (A \cap B) \cup (B \cap \bar{A}).$$

Proof (contd...)

Hence

$$\begin{aligned} P(A \cup B) &= P(A) + P(B \cap \bar{A}), \\ P(B) &= P(A \cap B) + P(B \cap \bar{A}). \end{aligned}$$

Subtracting the second equation from the first yields

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

and hence the result follows.

Note : This theorem represents an obvious extension of Axiom 2, for if $A \cap B = \emptyset$, we obtain from the above the statement of Axiom 2.

Example

Example 21.

J is taking two books along on her holiday vacation. With probability 0.5*, she will like the first book; with probability* 0.4*, she will like the second book; and with probability* 0.3*, she will like both books. What is the probability that she likes neither book?*

Solution

Let B_i denote the event that J likes book i, $i = 1, 2$. Then the probability that she likes at least one of the books is

$$P(B_1 \cup B_2) = P(B_1) + P(B_2) - P(B_1 \cap B_2) = 0.5 + 0.4 - 0.3 = 0.6.$$

Because the event that J likes neither book is the complement of the event that she likes at least one of them, we obtain the result

$$P(B_1^c \cap B_2^c) = P((B_1 \cup B_2)^c) = 1 - P(B_1 \cup B_2) = 0.4.$$

Theorem 22.

If A, B, and C are any three events, 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 : The proof consists of writing $A \cup B \cup C$ as $(A \cup B) \cup C$ and applying the result of the above theorem.

Theorem 23.

Let $A_1, \ldots, A_k$ be any k events. Then

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i) - \sum_{i<j=2}^{k} P(A_i \cap A_j) + \sum_{i<j<r=3}^{k} P(A_i \cap A_j \cap A_r) + \cdots \\ + (-1)^{k-1} P(A_1 \cap A_2 \cap \cdots \cap A_k).$$

Inclusion-Exclusion Identity

The above theorem, known as the **inclusion-exclusion identity**, can be proved by mathematical induction.

The summation

$$\sum_{i_1 < i_2 < \cdots < i_r} P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_r})$$

is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2, \ldots, n\}$.

In words, Theorem 23 states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time, minus the sum of the probabilities of these events taken two at a time, plus the sum of the probabilities of these events taken three at a time, and so on.

Note: Noninductive argument for Theorem 23 is available in the book by Sheldon Ross.

Inclusion-Exclusion Identity

The following is a succinct way of writing the inclusion-exclusion identity:

$$P(\cup_{i=1}^{n} A_i) = \sum_{r=1}^{n} (-1)^{r+1} \sum_{i_1 < \cdots < i_r} P(A_{i_1} \cap \cdots \cap A_{i_r}).$$

In the inclusion-exclusion identity, going out one term results in an upper bound on the probability of the union, going out two terms results in a lower bound on the probability, going out three terms results in an upper bound on the probability, going out four terms results in a lower bound, and so on.

Inclusion-Exclusion Identity

That is, for events $A_1, \ldots, A_n$, we have

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) \tag{1}$$

$$P(\cup_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} P(A_i) - \sum_{j<i} P(A_i \cap A_j) \tag{2}$$

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i) - \sum_{j<i} P(A_i \cap A_j) + \sum_{k<j<i} P(A_i \cap A_j \cap A_k) \tag{3}$$

and so on.

Exercise 24.

Prove the above inequalities.

Inclusion-Exclusion Identity

Notation : We say AB for $A \cap B$ when A and B are two events.

To prove the validity of these bounds, note the identity

$$\cup_{i=1}^{n} A_i = A_1 \cup A_1^c A_2 \cup A_1^c A_2^c A_3 \cup \cdots \cup A_1^c \cdots A_{n-1}^c A_n.$$

That is, at least one of the events A_i occurs if A_1 occurs, or if A_1 does not occur but A_2 does, or if A_1 and A_2 do not occur but A_3 does, and so on. Because the right-hand side is the union of disjoint events, we obtain

$$P(\cup_{i=1}^{n} A_i) = P(A_1) + P(A_1^c A_2) + P(A_1^c A_2^c A_3) + \cdots + P(A_1^c \cdots A_{n-1}^c A_n)$$

$$= P(A_1) + \sum_{i=2}^{n} P(A_1^c \cdots A_{i-1}^c A_i). \tag{4}$$

Now, let $B_i = A_1^c \cdots A_{i-1}^c = (\cup_{j<i} A_j)^c$ be the event that none of the first $i-1$ events occur.

Inclusion-Exclusion Identity

Applying the identity

$$P(A_i) = P(B_i A_i) + P(B_i^c A_i)$$

shows that

$$P(A_i) = P(A_1^c \cdots A_{i-1}^c A_i) + P(A_i \cup_{j<i} A_j)$$

or, equivalently,

$$P(A_1^c \cdots A_{i-1}^c A_i) = P(A_i) - P(\cup_{j<i} A_i A_j). \tag{5}$$

Substituting this equation into (4.4) yields

$$P(\cup_{i=1}^n A_i) = \sum_i P(A_i) - \sum_i P(\cup_{j<i} A_i A_j).$$

Inclusion-Exclusion Identity

Because probabilities are always nonnegative, Inequality (1) follows directly from Equation (5). Now, fixing i and applying Inequality (1) to $P(\cup_{j<i} A_i A_j)$ yields

$$P(\cup_{j<i} A_i A_j) \leq \sum_{j<i} P(A_i A_j)$$

which, by Equation (5), gives Inequality (2). Similarly, fixing i and applying Inequality (2) to $P(\cup_{j<i} A_i A_j)$ yields

$$\begin{aligned} P(\cup_{j<i} A_i A_j) &\geq \sum_{j<i} P(A_i A_j) - \sum_{k<j<i} P(A_i A_j A_i A_k) \\ &= \sum_{j<i} P(A_i A_j) - \sum_{k<j<i} P(A_i A_j A_k) \end{aligned}$$

which, by Equation (5), gives Inequality (3). The next inclusion-exclusion inequality is now obtained by fixing i and applying Inequality (3) to $P(\cup_{j<i} A_i A_j)$, and so on.

Continuity Property of Probabilities

(a) Let $A_1, A_2, \ldots$ be an infinite sequence of events, which is "monotonically increasing", meaning that $A_n \subset A_{n+1}$ for every n. Let $A = U_{n=1}^{\infty} A_n$. Show that $P(A) = \lim_{n\to\infty} P(A_n)$. *Hint:* Express the event A as a union of countably many disjoint sets.

(b) Suppose now that the events are "monotonically decreasing", i.e., $A_{n+1} \subset A_n$ for every n. Let $A = \cap_{n=1}^{\infty} A_n$. Show that $P(A) = \lim_{n\to\infty} P(A_n)$. *Hint:* Apply the result of part (a) to the complements of the events.

(c) Consider a probabilistic model whose sample space is the real line. Show that

$$P([0,\infty)) = \lim_{n\to\infty} P([0,n]), \quad \text{and} \quad \lim_{n\to\infty} P([n,\infty)) = 0.$$

Proof of (c) : For the first equality, use the result from part (a) with $A_n = [0, n]$ and $A = [0, \infty)$.

For the second, use the result from part (b) with $A_n = [n, \infty)$ and $A = \cap_{n=1}^{\infty} A_n = \emptyset$.

Exercises 25.

(a) *Prove that for any two events A and B, we have*

$$P(A \cap B) \geq P(A) + P(B) - 1.$$

(b) *Generalize to the case of n events $A_1, A_2, \ldots, A_n$, by showing that*

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) \geq P(A_1) + P(A_2) + \cdots + P(A_n) - (n-1).$$

Solution. We have $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and $P(A \cup B) \leq 1$. which implies part (a). For part (b), we use De Morgan's law to obtain

$$\begin{aligned} 1 - P(A_1 \cap \cdots \cap A_n) &= P((A_1 \cap \ldots \cap A_n)^c) \\ &= P(A_1^c \cup \ldots \cup A_n^c) \\ &\leq P(A_1^c) + \cdots + P(A_n^c) \\ &= (1 - P(A_1)) + \cdots + (1 - P(A_n)) \\ &= n - P(A_1) - \cdots - P(A_n). \end{aligned}$$

Exercises 26.

1. *A partition of the sample space Ω is a collection of disjoint events $S_1, \ldots, S_n$ such that $\Omega = U_{i=1}^{n} S_i$.*
   (a) *Show that for any event A, we have*

$$P(A) = \sum_{i=1}^{n} P(A \cap S_i).$$

   (b) *Use part (a) to show that for any events A, B, and C, we have*

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

2. *Show the formula*

$$P((A \cap B^c) \cup (A^c \cap B)) = P(A) + P(B) - 2P(A \cap B),$$

which gives the probability that exactly one of the events A and B will occur. [Compare with the formula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$, which gives the probability that at least one of the events A and B will occur.]

Summary of Counting Results

Here is a summary of all the counting results we have developed.

- **Permutations** of n objects: $n!$.
- **k-permutations** of n objects: $n!/(n-k)!$.
- **Combinations of k** out of n objects: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
- **Partitions** of n objects into r groups, with the ith group having n_i objects:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}.$$

References

1. Meyer, *Introductory Probability and Statistical Applications*, 2nd Edition, Oxford & IBH Publishing Company, India.
2. Sheldon Ross, *First Course in Probability*, Sixth Edition, Pearson Publisher, India.
3. Dimitri P. Bertsekas and John N. Tsitsiklis, *Introduction to Probability*, Athena Scientific, Belmont, Massachusetts.