# ECE 302: Probability and Applications[1]

## Week 1 Topics

- Probability Models
  - Random Experiments
  - Relative Frequency
  - Sample Spaces and Events
- Set Operations
- Axioms of Probability
- Discrete vs Continuous Sample Spaces

# 1 Probability Models

Many systems in engineering involve phenomena that exhibit unpredictable variation and randomness. Probability models these systems and phenomena in terms of random experiments

- A **random experiment** is an experiment in which the outcome varies in an unpredictable fashion when the experiment is repeated under the same conditions.
- Each time we perform a random experiment we observe a single unique **outcome**.
- We call the set $S$ of all possible outcomes the **sample space**.
- Examples:
  - Flip a coin: S={H,T}
  - Observe a binary signal S={0,1}
  - Pick a numbered ball from an urn S={1,2,...N}
  - Measure the lifetime of a lightbulb S=[0,50000]
  - Measure the distance of a dart to the center of a target S=[0,15]

Figure 1 shows the outcomes in 100 repetitions (trials) of a computer simulation of an urn experiment with N=3. It is clear that the outcome of this experiment cannot consistently be predicted correctly.

## 1.1 Statistical Regularity

- If a random experiment is repeated under *completely identical conditions* then the outcome of an experiment cannot depend on the outcome of any other repetitions of the experiment.
- We find in practice that averages obtained in long sequences of repetitions (trials) of random experiments consistently to the same value.
- This property is called **statistical regularity**.

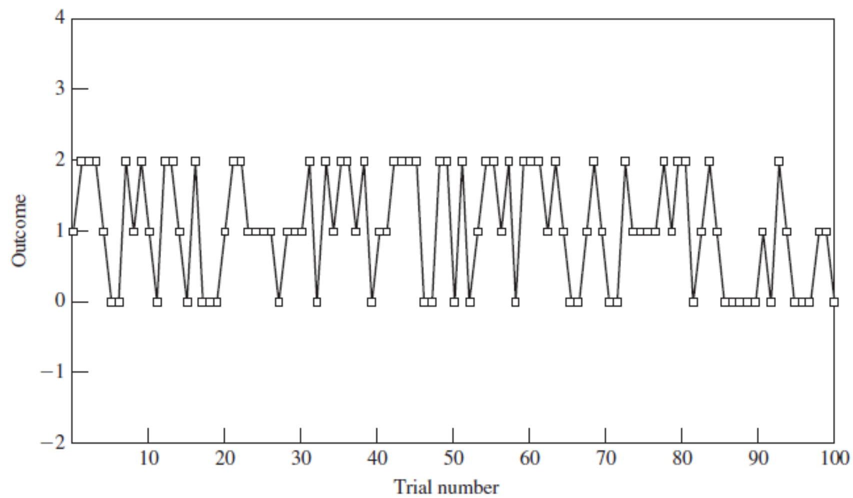



**Figure 1:** Outcomes of urn experiment

- Probability models are useful model because they allow us to make predictions about the future behavior of a system, specifically in terms of the regularity of long-term averages.

- Suppose that an experiment has sample space $\{1, 2, ..., K\}$, for example, and an urn with $K$ numbered balls.

- Suppose the experiment is repeated $n$ times under identical conditions. Let $N_k(n)$ be the number of times in which the outcome is "k", where $0 <= k <= K$.

- The **relative frequency** of outcome $k$ is defined as the fraction of times (out of n trials) that the outcome is "k":

$$f_k(n) = \frac{N_k(n)}{n} \tag{1}$$

- By statistical regularity we mean that $f_k(n)$ varies less and less about a constant value as $n$ is made large, that is,

$$\lim_{n\to\infty} f_k(n) = p_k \tag{2}$$

- We call the constant $p_k$ the **probability** of outcome $k$.

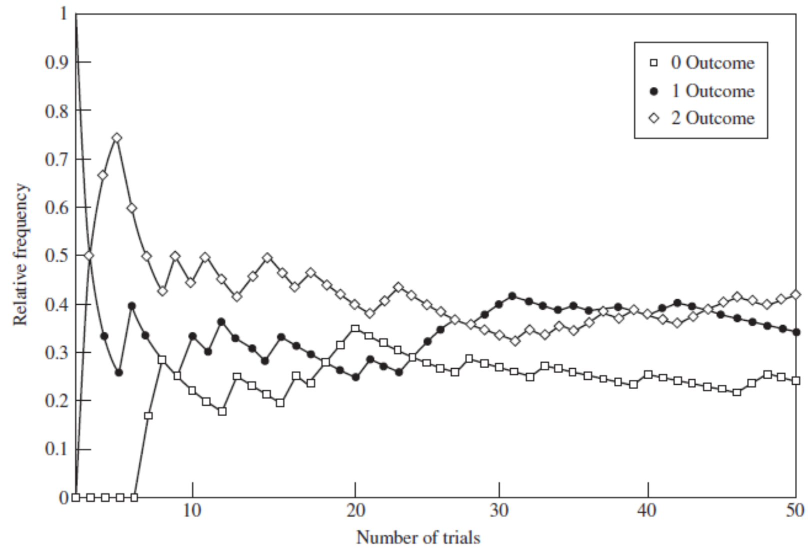



**Figure 2:** Relative frequencies in urn experiment

- Equation (2) states that the probability of an outcome is the long-term proportion of times it arises in a long sequence of trials.

- Eq. (2) provides a key connection between measurement of random quantities and probability models.

- Figures 2 and 3 show the relative frequencies for the three outcomes in the urn experiment as $n$ is increased. In this case relative frequencies of the three possible outcomes are converging to the same value 1/3. This corresponds to the case where the three outcomes are equiprobable.

- Suppose we modify the urn experiment by placing in the urn two additional identical balls with the numbers 1 and 2. The probability of the outcome 0 is now 1/5 and the probabilities of the outcomes 1 and 2 increase to 2/5 each. This illustrates a key property of probability models, namely, *the conditions under which a random experiment is performed determine the probabilities of the outcomes of an experiment.*

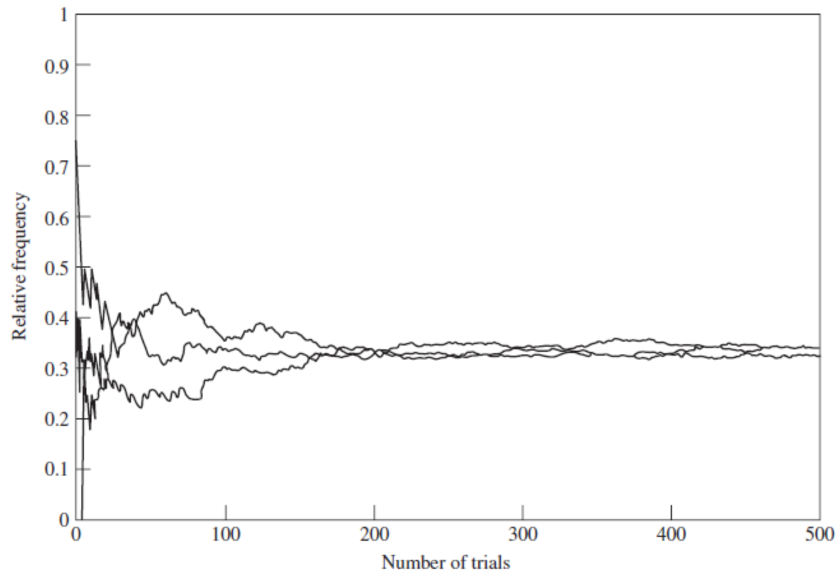



**Figure 3:** Relative frequencies in urn experiment

## 1.2 Properties of Relative Frequency

- Suppose that a random experiment has $S = \{1, 2, \ldots, K\}$. Since the number of occurrences of any outcome in $n$ trials is a number between zero and $n$,

$$0 \leq N_k(n) \leq n \text{ for } k = 1, 2, \ldots, K.$$

- Dividing the above equation by $n$, we find that the relative frequencies are a number between zero and one:

$$0 \leq f_k(n) \leq 1 \quad \text{for } k = 1, 2, \ldots, K \tag{3}$$

- The sum of the number of occurrences of all possible outcomes must be $n$ :

$$\sum_{k=1}^{K} N_k(n) = n$$

- If we divide both sides of the above equation by $n$, we find that the sum of all the relative frequencies equals one:

$$\sum_{k=1}^{K} f_k(n) = 1 \tag{4}$$

- Sometimes we are only interested in whether the occurrence satisfies come condition, that is, whether the occurrence belongs to the set of outcomes that satisfy the condition of interest. An **event** is a set of outcomes.

- This example shows that the **relative frequency of an event** is the sum of the relative frequencies of the associated outcomes.

- More generally, let $C$ be the event "$A$ or $B$ occurs," where $A$ and $B$ are two events that cannot occur simultaneously, then the number of times when $C$ occurs is $N_C(n) = N_A(n) + N_B(n)$, so

$$f_C(n) = f_A(n) + f_B(n) \tag{5}$$

- Equations (3), (4), and (5) are the three basic properties of relative frequency from which we can derive many other useful results.

## 1.3 The Axiomatic Approach to a Theory of Probability

- Equation (2) suggests that we define the probability of an event by its long-term relative frequency. But there are problems with using this definition to develop a mathematical theory of probability.
  - First of all, it is not clear when and in what mathematical sense the limit in Eq. (2) exists.
  - Second, we can never perform an experiment an infinite number of times so we can never know the probabilities $p_k$ exactly.
  - Finally, the use of relative frequency to define probability would rule out the applicability of probability theory to situations in which an experiment cannot be repeated.

- It is better to develop a mathematical theory of probability that is not tied to any particular notion of what probability "means".
- We only require that the definition of probability be consistent with the relative frequency interpretation, that is properties in Eqs. (3) through (5) are satisfied.
- The theory of probability is based on a set of axioms that specify that probabilities of events must satisfy certain properties. It supposes that: (1) A random experiment has been defined, and a set $S$ of all possible outcomes has been identified; (2) a class of subsets of $S$ called events has been specified as well; and (3) each event $A$ is assigned a number, $P[A]$, so that the following axioms are satisfied:
  1. $0 \le P[A] \le 1$.
  2. $P[S] = 1$.
  3. If $A$ and $B$ are events that cannot occur simultaneously, then $P[A \text{ or } B] = P[A] + P[B]$.
- The three axioms correspond to the properties of relative frequency stated in Eqs. (3) through (5). These three axioms lead to many useful and powerful results.
- The theory of probability does not concern itself with how the probabilities are obtained or with what they mean. Any assignment of probabilities to events that satisfies the above axioms is legitimate.
- The user of the theory, the model builder, determines what the probability assignment should be and what interpretation of probability makes sense in any given application.

# 2 Specifying Random Experiments

- *A random experiment is specified by stating an experimental procedure and a set of one or more measurements or observations.*
- When a random experiment is performed, one and only one outcome occurs. But the outcome can vary as the experiment is repeated.

**Example 1: Some Random Experiments**

*Experiment* $E_1$: Select a ball from an urn containing balls numbered 1 to 50. Note the number of the ball.

*Experiment* $E_2$: Select a ball from an urn containing balls numbered 1 to 4. Suppose that balls 1 and 2 are black and that balls 3 and 4 are white. Note the number and color of the ball.

*Experiment* $E_3$: Toss a coin three times and note the sequence of heads and tails.

*Experiment* $E_4$: Toss a coin three times and note the number of heads.

*Experiment* $E_5$: Count the number of spam emails in a group of $N$ emails.

*Experiment* $E_6$: Count the number of coin tosses until heads appears.

*Experiment* $E_7$: Pick a number at random between zero and one.

*Experiment* $E_8$: Measure the time between consecutive message arrivals at a server.

*Experiment* $E_9$: Measure the lifetime of a given computer chip.

*Experiment* $E_{10}$: Determine the value of a voltage waveform at time $t_1$.

*Experiment* $E_{11}$: Determine the values of a voltage waveform at times $t_1$ and $t_2$.

*Experiment* $E_{12}$: Pick two numbers at random between zero and one.

*Experiment* $E_{13}$: Pick a number $X$ at random between zero and one, then pick a number $Y$ at random between $X$ and one.

*Experiment* $E_{14}$: A system component is installed at time $t = 0$. For $t \geq 0$ let $X(t) = 1$ as long as the component is functioning, and let $X(t) = 0$ after the component fails.

- A random experiment must be specified with an unambiguous statement of exactly what is measured or observed. For example, random experiments may consist of the same procedure but differ in the observations made, as illustrated by $E_3$ and $E_4$.

- A random experiment may involve more than one measurement or observation, as illustrated by $E_2$, $E_3$, $E_{11}$, $E_{12}$, and $E_{13}$.

- A random experiment may even involve a continuum of measurements, as shown by $E_{14}$.

- Experiments $E_3$, $E_4$, $E_5$, $E_6$, $E_{12}$, and $E_{13}$ are examples of sequential experiments that can be viewed as consisting of a sequence of simple subexperiments.

- In $E_{13}$ the second sub-experiment depends on the outcome of the first sub-experiment.

## 2.1 The Sample Space

- *The **Sample Space** is the set of all possible outcomes of a random experiment.*

- We define an ***outcome*** or ***sample point*** of a random experiment as a result that is "finest grain" in the sense that it cannot be decomposed into other results. For example, if the outcome is the color green, then we cannot distinguish different shades of green.

- When we perform a random experiment, *one and only one outcome occurs*. It then follows that outcomes are mutually exclusive in the sense that they cannot occur simultaneously.

- We will denote an outcome of an experiment by $\zeta$, where $\zeta$ i*s an element or point in $S$*.

- Each performance of a random experiment can then be viewed as the selection at random of a single point (outcome) from $S$.

- *The sample space $S$ can be specified compactly by using set notation.*

- We can specify sets in many ways: tables, diagrams, intervals of the real line, regions of the plane, and more.

**Example 2: Some Sample Spaces**

The sample spaces corresponding to the experiments described in Example 1 are given below using set notation:

$$S_1 = \{1, 2, \ldots, 50\}$$
$$S_2 = \{(1, b), (2, b), (3, w), (4, w)\}$$
$$S_3 = \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{TTH}, \text{THT}, \text{HTT}, \text{TTT}\}$$
$$S_4 = \{0, 1, 2, 3\}$$
$$S_5 = \{0, 1, 2, \ldots, N\}$$
$$S_6 = \{1, 2, 3, \ldots\}$$
$$S_7 = \{x : 0 \leq x \leq 1\} = [0, 1] \quad \text{See Fig. 4(a).}$$
$$S_8 = \{t : t \geq 0\} = [0, \infty)$$
$$S_9 = \{t : t \geq 0\} = [0, \infty) \quad \text{See Fig. 4(b).}$$
$$S_{10} = \{v : -\infty < v < \infty\} = (-\infty, \infty)$$
$$S_{11} = \{(v_1, v_2) : -\infty < v_1 < \infty \text{ and } -\infty < v_2 < \infty\}$$
$$S_{12} = \{(x, y) : 0 \leq x \leq 1 \text{ and } 0 \leq y \leq 1\} \quad \text{See Fig. 4(c).}$$
$$S_{13} = \{(x, y) : 0 \leq y \leq x \leq 1\} \quad \text{See Fig. 4(d).}$$

$S_{14}$ = set of functions $X(t)$ for which $X(t) = 1$ for $0 \leq t < t_0$ and $X(t) = 0$ for $t \geq t_0$, where $t_0 > 0$ is the time when the component fails.

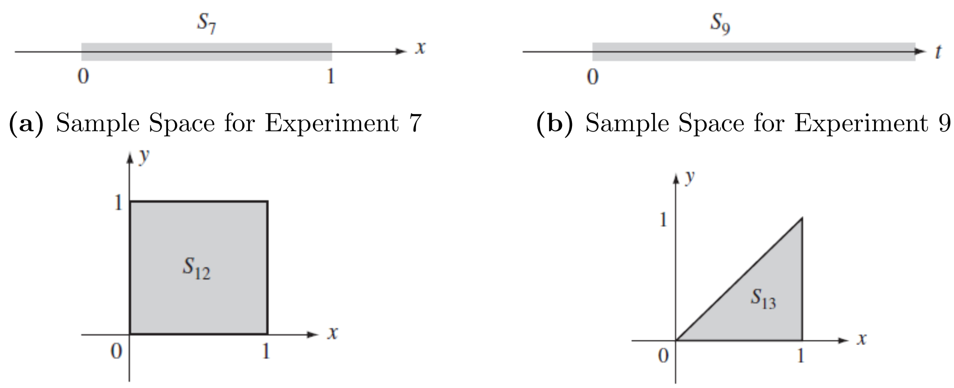
---

- *There are three possibilities for the number of outcomes in a sample space:*
  - sample space is **finite**
  - sample space is **countably infinite**,that is, its outcomes can be put into one-to-one correspondence with the positive integers.
  - sample space is **uncountably infinite**.

- We say that $S$ is a **discrete sample space** if $S$ is finite or countably infinite.
- We say that $S$ is a **continuous sample space** if $S$ is uncountably infinite.
- Experiments $E_1, E_2, E_3, E_4$, and $E_5$ have finite discrete sample spaces.
- Experiments $E_6$ has a countably infinite discrete sample space. Experiments $E_7$ through $E_{13}$ have continuous sample spaces.
- If an experiment consists of one or more observations or measurements, then the sample space $S$ can be multi-dimensional.
  - The outcomes in Experiments $E_2$, $E_{11}$, $E_{12}$, and $E_{13}$ are two-dimensional, and those in Experiment $E_3$ are three-dimensional.
- Sometimes, the sample space is the Cartesian product of other sets.1
  - For example, $S_{11} = R \times R$, where $R$ is the real line, and $S_3 = S \times S \times S$, where $S =${H, T}.
- Sometimes for convenience we let the sample space include outcomes that are not possible.
  - in Experiment $E_9$ it is convenient to define the sample space as the positive real line, even though a device cannot have an infinite lifetime.

## 2.2 Events

- *We are usually on the occurrence of some event (i.e., whether the outcome satisfies certain conditions). that is, whether the outcome belongs to some set.*
- The conditions of interest define a subset of the sample space, namely the set of points $\zeta$, from that satisfy the given conditions.
- The **event** occurs if and only if the outcome of the experiment $\zeta$ is in this subset.




**(a)** Sample Space for Experiment 7 **(b)** Sample Space for Experiment 9

**(c)** Sample Space for Experiment 12 **(d)** Sample Space for Experiment 13

**Figure 4:** Sample Spaces for Select Experiments

- Clearly every event is a subset of $S$.
- *There are two special events:*
  - The **certain event**, $S$, which consists of all outcomes and hence always occurs;
  - the **impossible** or **null event**, $\varnothing$, which contains no outcomes and hence never occurs.

**Example 3: Some Example Events**

In the following examples, $A_k$ refers to an event corresponding to an experiment $E_k$ in Example 1.

$E_1$ : "An even-numbered ball is selected," $A_1=\{2, 4, \ldots, 48, 50\}$.

$E_2$ : "The ball is white and even numbered," $A_2 = \{(4, w)\}$.

$E_3$ : "The three tosses give the same outcome," $A_3=\{\text{HHH},\text{TTT}\}$.

$E_4$ : "The number of heads equals the number of tails," $A_4 = \varnothing$.

$E_5$ : "No active packets are produced," $A_5 = \{0\}$.

$E_6$ : "Fewer than 10 transmissions are required," $A_6 = \{1, \ldots, 9\}$.

$E_7$ : "The number selected is nonnegative," $A_7=S_7$.

$E_8$ : "Less than $t_0$ seconds elapse between message arrivals," $A_8 = \{t : 0 < t < t_0\} = [0, t_0)$.

$E_9$ : "The chip lasts more than 1000 hours but less than 1500 hours," $A_9 = \{t : 1000 < t < 1500\} = (1000, 1500)$.

$E_{10}$ : "The absolute value of the voltage is less than 1 volt," $A_{10} = \{v : -1 < v < 1\} = (-1, 1)$.

$E_{11}$ : "The two voltages have opposite polarities," $A_{11} = \{(v_1, v_2) : (v_1 < 0$ and $v_2 > 0$ ) or ( $v_1 > 0$ and $v_2 < 0)\}$.

$E_{12}$ : "The two numbers differ by less than 1/10," $A_{12} = \{(x, y) : (x, y)$ in $S_{12}$ and $|x - y| < 1/10\}$.

$E_{13}$ : "The two numbers differ by less than 1/10," $A_{13} = \{(x, y) : (x, y)$ in $S_{13}$ and $|x - y| < 1/10\}$.

$E_{14}$ : "The system is functioning at time $t_1$," $A_{14}$ = subset of $S_{14}$ for which $X(t_1) = 1$.

---

- An event may consist of a single outcome, as in $A_2$ and $A_5$.
- An event from a discrete sample space that consists of a single outcome is called an **elementary event**. Events $A_2$ and $A_5$ are elementary events.
- An event may also consist of the entire sample space, as in $A_7$.
- The null event, $\varnothing$, arises when none of the outcomes satisfy the conditions that specify a given event, as in $A_4$.

## 2.3 Set Operations

- *We can apply set operations to events because they correspond to sets.*
- We can combine events using **set operations** to obtain other events.
- We can also express complicated events as combinations of simple events.

- The **union** of two events $A$ and $B$ is denoted by $A \cup B$ and is defined as the set of outcomes that are either in $A$ or in $B$, or both. The event $A \cup B$ occurs if either $A$, or $B$, or both $A$ and $B$ occur.

- The **intersection** of two events $A$ and $B$ is denoted by $A \cap B$ and is defined as the set of outcomes that are in both $A$ and $B$. The event $A \cap B$ occurs when both $A$ and $B$ occur. Two events are said to be **mutually exclusive** if their intersection is the null event, $A \cap B = \varnothing$. Mutually exclusive events cannot occur simultaneously.

- The **complement** of an event $A$ is denoted by $A^c$ and is defined as the set of all outcomes not in $A$. The event $A^c$ occurs when the event $A$ does not occur and vice versa.

- Figures 5(a), 5(b), and 5(c) show the basic set operations using Venn diagrams. In these diagrams the rectangle represents the sample space $S$, and the shaded regions represent the various events.

- Figure 5 (d) shows two mutually exclusive events.

- If an event $A$ is a subset of an event $B$, that is $A \subset B$, then event $B$ will occur whenever event $A$ occurs because all the outcomes in $A$ are also in $B$ (see Fig. 5(e)). For this reason we say that event $A$ **implies** event $B$.

- The events $A$ and $B$ are equal, $A = B$, if they contain the same outcomes.

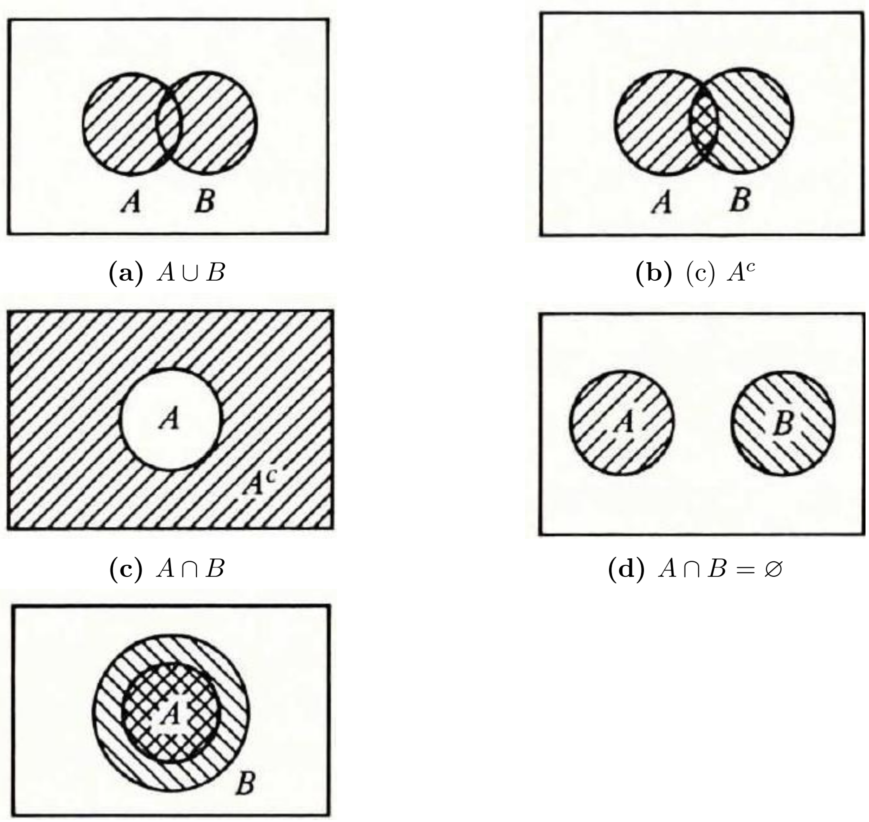
The three basic operations can be combined to form other events. The following properties of set operations and their combinations are useful:

*Commutative Properties:*

$$A \cup B = B \cup A \quad \text{and} \quad A \cap B = B \cap A. \tag{6}$$

*Associative Properties:*

$$\begin{aligned} &A \cup (B \cup C) = (A \cup B) \cup C \quad \text{and} \\ &A \cap (B \cap C) = (A \cap B) \cap C. \end{aligned} \tag{7}$$




**(a)** $A \cup B$ **(b)** (c) $A^c$

**(c)** $A \cap B$ **(d)** $A \cap B = \varnothing$

**(e)** $A \subset B$

**Figure 5:** Set operations and set relations

*Distributive Properties:*

$$
\begin{aligned}
&A \cup (B \cap C) = (A \cup B) \cap (A \cup C) \quad \text{and} \\
&A \cap (B \cup C) = (A \cap B) \cup (A \cap C). 
\end{aligned} \tag{8}
$$

*DeMorgan's Rules:*

$$(A \cap B)^c = A^c \cup B^c \quad \text{and} \quad (A \cup B)^c = A^c \cap B^c. \tag{9}$$

- *The union and intersection operations can be repeated for an arbitrary lumber of events.*

- Thus the event

$$\bigcup_{k=1}^{n} A_k = A_1 \cup A_2 \cup \cdots \cup A_n$$

occurs if one or more of the events $A_k$ occur. The event

$$\bigcap_{k=1}^{n} A_k = A_1 \cap A_2 \cap \cdots \cap A_n$$

occurs when all of the events $A_1, \ldots, A_n$ occur.

- The operations can also be applied to a countably infinite sequence of events. Thus we can also have events of the form

$$\bigcup_{k=1}^{\infty} A_k \quad \text{and} \quad \bigcap_{k=1}^{\infty} A_k.$$

- *Suppose we are given a basic set of events $A_1, A_2, ...A_n$, then we can generate a collection of events by applying the set operations "union, intersection, and complement" to all possible subsets of these events.*

- We say that a collection of events $\mathscr{F}$ is a **field** if any set operation on events in the class produce an event that is in the class.

# 3 The Axioms of Probability

- *A probability is a number assigned to an event.*

- Intuitively the probability of an event indicates how likely it is to occur when the corresponding random experiment is performed.

- A **probability law** for a random experiment is a function that assigns probabilities to the events of the experiment.

- The **axioms of probability** formally state that a probability law must satisfy a set of properties.

| | |
|---|---|
| *Axiom I* | $0 \leq P[A]$ |
| *Axiom II* | $P[S] = 1$ |
| *Axiom III* | If $A \cap B = \varnothing$, then $P[A \cup B] = P[A] + P[B]$ |
| *Axiom III$'$* | If $A_1, A_2, \ldots$ is a sequence of events such that $A_i \cap A_j = \varnothing$ for $i \neq j$, then $P\left[\bigcup_{k=1}^{\infty} A_k\right] = \sum_{k=1}^{\infty} P\left[A_k\right]$. |

Let $E$ be a random experiment with sample space $S$ and let $\mathscr{F}$ be a field for a collection of events. A **probability law** for the experiment $E$ is a rule that assigns to each event $A$ in $\mathscr{F}$ a number $P[A]$, called the **probability of $\boldsymbol{A}$**, that satisfies the following axioms:

- Axioms I, II, and III are enough to deal with experiments with finite sample spaces because the number of possible events is finite.
- In order to handle experiments with infinite sample spaces, Axiom III needs to be replaced by Axiom III$'$.
- Note the Axiom III$'$ cannot be proven from Axiom III.
- Note also that Axiom III$'$ includes Axiom III as a special case, by letting $A_k = \varnothing$ for $k \geq 3$. Thus we really only need Axioms I, II, and III$'$. However we will gain greater insight by starting with Axioms I, II, and III.
- *We view events as objects that possess a property (i.e., their probability) that has attributes similar to physical mass.*
  - Axiom I states that the probability (mass) is nonnegative
  - Axiom II states that there is a fixed total amount of probability (mass), namely 1 unit.
  - Axiom III states that the total probability (mass) in two disjoint objects is the sum of the individual probabilities (masses).
- *The axioms imply a set of consistency rules that any valid probability assignment must satisfy.*
- These rules are useful in the computation of probabilities.

- First, if we partition the sample space into two mutually exclusive events, $A$ and $A^c$, then the probabilities of these two events add up to one.

**Corollary 1.** $P[A^c] = 1 - P[A]$

*Proof.* Since an event $A$ and its complement $A^c$ are mutually exclusive, $A \cap A^c = \varnothing$, we have from Axiom III that

$$P[A \cup A^c] = P[A] + P[A^c]$$

Since $S = A \cup A^c$, by Axiom II,

$$1 = P[S] = P[A \cup A^c] = P[A] + P[A^c]$$

The corollary follows after solving for $P[A^c]$.

□

- Second, the probability of an event is always less than or equal to one.

**Corollary 2.** $P[A] \leq 1$

*Proof.* From Corollary 1,

$$P[A] = 1 - P[A^c] \leq 1,$$

since $P[A^c] \geq 0$. □

- Combined with Axiom I we have a good check in problem solving: If your probabilities are negative or are greater than one, you have made a mistake!

- Three, the impossible event has probability zero.

**Corollary 3.** $P[\varnothing] = 0$

*Proof.* Let $A = S$ and $A^c = \varnothing$ in Corollary 1:

$$P[\varnothing] = 1 - P[S] = 0$$

□

- Four, we can compute the probability of a complicated event $A$ by decomposing it into the union of disjoint events $A_1, A_2, \ldots, A_n$. The probability of $A$ is the sum of the probabilities of the $A_k$ 's.

**Corollary 4.** *If $A_1, A_2, \ldots, A_n$ are pairwise mutually exclusive, then*

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P\left[A_k\right] \qquad \textit{for } n \geq 2$$

*Proof.* We use mathematical induction. Axiom III implies that the result if true for $n = 2$. Next we need to show that if the result is true for some $n$, then it is also true for $n+1$. This combined with the fact that the result is true for $n = 2$, implies that the result is true for $n \geq 2$.

Suppose that the result is true for $n$, that is,

$$P\left[\bigcup_{k=1}^{n} A_k\right] = \sum_{k=1}^{n} P\left[A_k\right] \tag{10}$$

and consider the $n+1$ case

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = P\left[\left\{\bigcup_{k=1}^{n} A_k\right\} \cup A_{n+1}\right] = P\left[\bigcup_{k=1}^{n} A_k\right] + P\left[A_{n+1}\right] \tag{11}$$

where we have applied Axiom III to to the second expression after noting that the union of events $A_1$ to $A_n$ is mutually exclusive with $A_{n+1}$ :

$$\left\{\bigcup_{k=1}^{n} A_k\right\} \cap A_{n+1} = \bigcup_{k=1}^{n} \left\{A_k \cap A_{n+1}\right\} = \bigcup_{k=1}^{n} \varnothing = \varnothing$$

Substitution of Eq.(10) into Eq. (11) gives the $n+1$ case

$$P\left[\bigcup_{k=1}^{n+1} A_k\right] = \sum_{k=1}^{n+1} P\left[A_k\right]$$

□

- Five, the probability of the union of two events that are not necessarily mutually exclusive is found as follows:

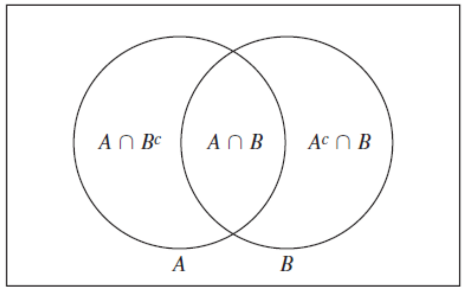



**Figure 6:** Decomposition of $A \cup B$ into three disjoint sets

**Corollary 5.** $P[A \cup B] = P[A] + P[B] - P[A \cap B]$

*Proof.* First we decompose $A \cup B$, $A$, and $B$ as unions of disjoint events. From the Venn diagram in Fig. 6,

$$
\begin{aligned}
&P[A \cup B] = P\left[A \cap B^c\right] + P\left[B \cap A^c\right] + P[A \cap B] \\
&P[A] = P\left[A \cap B^c\right] + P[A \cap B] \\
&P[B] = P\left[B \cap A^c\right] + P[A \cap B]
\end{aligned}
$$

By substituting $P\left[A \cap B^c\right]$ and $P\left[B \cap A^c\right]$ from the two lower equations into the top equation, we obtain the corollary. □

- From the Venn diagram in Fig. 6, we see that the sum $P[A] + P[B]$ counts the probability (mass) of the set $A \cap B$ twice. The expression in Corollary 5 makes the appropriate correction.

Corollary 5 is easily generalized to three events,

$$
\begin{aligned}
P[A \cup B \cup C] = &P[A] + P[B] + P[C] - P[A \cap B] \\
&- P[A \cap C] - P[B \cap C] + P[A \cap B \cap C],
\end{aligned}
\tag{12}
$$

and in general to $n$ events, as shown in Corollary 6.

**Corollary 6.**

$$
\begin{aligned}
P\left[\bigcup_{k=1}^{n} A_k\right] = &\sum_{j=1}^{n} P\left[A_j\right] - \sum_{j<k} P\left[A_j \cap A_k\right] + \cdots \\
&+ (-1)^{n+1} P\left[A_1 \cap \cdots \cap A_n\right]
\end{aligned}
$$

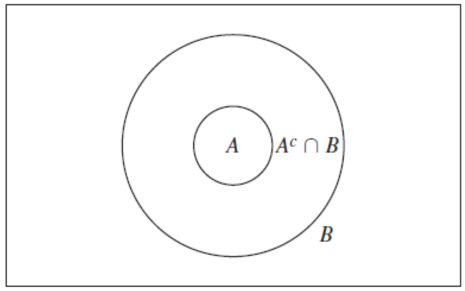



**Figure 7:** If $A \subset B$, then $P(A) \leq P(B)$

*Proof* is by induction.

- Since probabilities are nonnegative, Corollary 5 implies that the probability of the union of n events is no greater than the sum of the individual event probabilities.
- This is called the **union bound**.
- A typical use of this approach arises when we are interested in an event $A$ whose probability is difficult to find; so we find an event $B$ for which the probability can be found and that includes $A$ as a subset.

**Corollary 7.** *If $A \subset B$, then $P[A] \leq P[B]$.*

*Proof.* In Fig. 6, $B$ is the union of $A$ and $A^c \cap B$, thus

$$P[B] = P[A] + P\left[A^c \cap B\right] \geq P[A]$$

since $P\left[A^c \cap B\right] \geq 0$. □

- The axioms and corollaries give a set of rules for computing the probability of events in terms of other events.
- Next, we will consider how to select a basic set of events from which we can calculate the probability of all other events.

## 3.1 Discrete Sample Spaces

- *The probability law for an experiment with a countable sample space can be found from the probabilities of the elementary events.*

- If the sample space is *finite*, $S = \{a_1, a_2, \ldots, a_n\}$. All distinct elementary events are mutually exclusive, so by Corollary 4 the probability of any event $B = \{a_{k_1}, a_{k_2}, \ldots, a_{k_m}\}$ is given by

$$\begin{aligned} P[B] &= P\left[\{a_{k_1}, a_{k_2}, \ldots, a_{k_m}\}\right] \\ &= P\left[\{a_{k_1}\}\right] + P\left[\{a_{k_2}\}\right] + \cdots + P\left[\{a_{k_m}\}\right] \end{aligned} \tag{13}$$

- If $S$ is countably infinite, then Axiom III' $'$ implies that the probability of an event such as $D = \{b_{k_1}, b_{k_2}, \ldots\}$ is given by

$$P[D] = P\left[\{b_{k_1}\}\right] + P\left[\{b_{k_2}\}\right] + \cdots \tag{14}$$

- If the sample space has $n$ elements, $S = \{a_1, \ldots, a_n\}$, we say that probability assignment has **equally likely outcomes**, if

$$P\left[\{a_1\}\right] = P\left[\{a_2\}\right] = \cdots = P\left[\{a_n\}\right] = \frac{1}{n} \tag{15}$$

- The probability of any event that consists of $k$ outcomes, say $B = \{a_{j_1}, \ldots, a_{j_k}$, is

$$P[B] = P\left[\{a_{j_1}\}\right] + \cdots + P\left[\{a_{j_k}\}\right] = \frac{k}{n} \tag{16}$$

**Example 4**

An urn contains ten identical balls numbered $0, 1, \ldots, 9$. A random experiment involves selecting a ball from the urn and noting the number of the ball. Find the probability of the following events:

$A =$ "number of ball selected is odd,"

$B =$ "number of ball selected is a multiple of three,"

$C =$ "number of ball selected is less than" 5,

and of $A \cup B$ and $A \cup B \cup C$.

The sample space is $S = \{0, 1, \ldots, 9\}$, so the sets of outcomes corresponding to the above events are

$$A = \{1, 3, 5, 7, 9\}, \quad B = \{3, 6, 9\}, \quad \text{and} \quad C = \{0, 1, 2, 3, 4\}.$$

If we assume that the outcomes are equally likely, then

$$\begin{aligned}
P[A] &= P[\{1\}] + P[\{3\}] + P[\{5\}] + P[\{7\}] + P[\{9\}] = \frac{5}{10} \\
P[B] &= P[\{3\}] + P[\{6\}] + P[\{9\}] = \frac{3}{10} \\
P[C] &= P[\{0\}] + P[\{1\}] + P[\{2\}] + P[\{3\}] + P[\{4\}] = \frac{5}{10}
\end{aligned}$$

From Corollary 5

$$P[A \cup B] = P[A] + P[B] - P[A \cap B] = \frac{5}{10} + \frac{3}{10} - \frac{2}{10} = \frac{6}{10}$$

where we have used the fact that $A \cap B = \{3, 9\}$, so $P[A \cap B] = 2/10$. From Corollary 6

$$\begin{aligned}
P[A \cup B \cup C] =& P[A] + P[B] + P[C] - P[A \cap B] \\
& - P[A \cap C] - P[B \cap C] + P[A \cap B \cap C] \\
=& \frac{5}{10} + \frac{3}{10} + \frac{5}{10} - \frac{2}{10} - \frac{2}{10} - \frac{1}{10} + \frac{1}{10} \\
=& \frac{9}{10}
\end{aligned}$$

You should verify the answers for $P[A \cup B]$ and $P[A \cup B \cup C]$ by enumerating the outcomes in the events.

---

- We obtain different probability models for experiments with the same sample space by varying the probabilities of the elementary events: all we need is $n$ nonnegative numbers that add up to one.

- In any particular situation, we select the probability assignment to reflect experimental observations or prior knowledge to the extent possible.

- The following example shows a situation where there is more than one "reasonable" probability assignment and where experimental evidence is required to decide on the appropriate assignment.

**Example 5**

A coin is tossed three times. If we observe the sequence of heads and tails, then there are eight possible outcomes $S_3$ = HHH, HHT, HTH, THH, TTH, THT, HTT, TTT. If we assume that the outcomes of $S_3$ are equiprobable, then the probability of each of the eight elementary events is $1/8$. This probability assignment implies that the probability of obtaining two heads in three tosses is by Corollary 3

$$\begin{aligned} P[\text{“2 heads in 3 tosses”}] &= P[\{\text{HHT, HTH, THH}\}] \\ &= P[\{\text{HHT}\}] + P[\{\text{HTH}\}] + P[\{\text{THH}\}] = \frac{3}{8}. \end{aligned}$$

Suppose that we toss a coin three times but we count the number of heads in three tosses. The sample space is now $S_4 = \{0, 1, 2, 3\}$. If we assume the outcomes of $S_4$ to be equiprobable, then each of the elementary events of $S_4$ has probability $1/4$. This second probability assignment predicts that the probability of obtaining two heads in three tosses is

$$P[\text{“2 heads in tosses”}] = P[\{2\}] = \frac{1}{4}.$$

- The first probability assignment implies that the probability of two heads in three tosses is $3/8$, and the second probability assignment predicts that the probability is $1/4$. Thus the two assignments are not consistent with each other.

- As far as the theory is concerned, either one of the assignments is acceptable. It is up to us to decide which assignment is more appropriate.

**Example 6: Fair Coin Toss**

A fair coin is tossed repeatedly until the first head shows up; the outcome of the experiment is the number of tosses until the first head occurs. Find a probability law for this experiment.

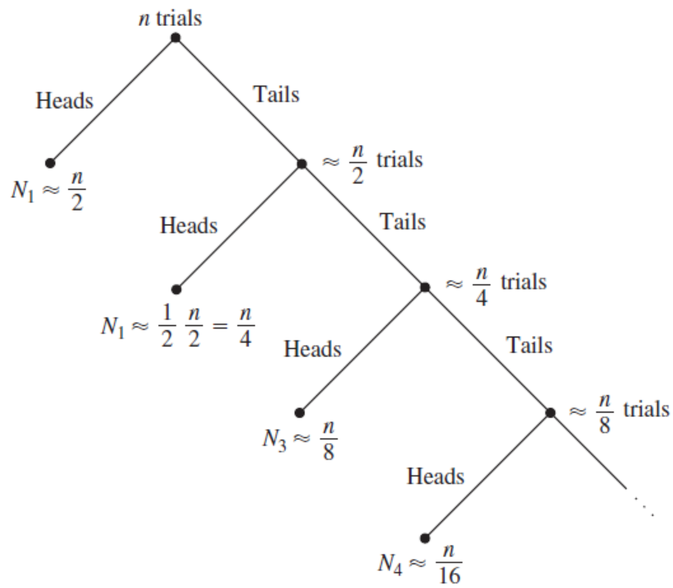



**Figure 8:** In $n$ trials "heads" comes up in the first toss approximately $n/2$ times; in the second toss approximately $n/4$ times; and so on.

An arbitrarily large number of tosses could be required if a head never occurs, so the sample space is $S = \{1, 2, 3, \ldots\}$. Suppose the experiment is repeated $n$ times. Let $N_j$ be the number of trials in which the $j$th toss results in the first head. If $n$ is very large, we expect $N_1$ to be approximately $n/2$ since the coin is fair. This implies that a second toss is necessarily about $n - N_1 \approx n/2$ times, and again we expect that about half of these, that is $n/4$, will result in heads, and so on as shown in Fig. 8. Thus for large $n$, the relative frequencies are

$$f_j \approx \frac{N_j}{n} = \left(\frac{1}{2}\right)^j \quad j = 1, 2, \ldots$$

We therefore conclude that a reasonable probability law for this experiment is

$$P[j \text{ tosses till first head }] = \left(\frac{1}{2}\right)^j \quad j = 1, 2, \ldots \tag{17}$$

We can verify that these probabilities add up to one by using the geometric series with $\alpha = 1/2$

$$\sum_{j=1}^{\infty} \alpha^j = \left.\frac{\alpha}{1-\alpha}\right|_{\alpha=1/2} = 1.$$

## 3.2 Continuous Sample Spaces

- *Continuous sample spaces arise in experiments in which the outcomes are numbers from an uncountably infinite set. The events of interest in these experiments consist of intervals of the real line, and of complements, unions, and intersections of intervals.*
- The field $\mathscr{F}$ of events of interest is then constructed from intervals of the real line, for example the semi-infinite interval, $(-\infty, x)$, where x is a real number.
- Thus, *probability laws in experiments with continuous sample spaces specify a rule for assigning a number to intervals of the real line.*

**Example 7**

Suppose we "pick a number $x$ at random between zero and one." The sample space $S$ for this experiment is the unit interval $[0, 1]$, which is uncountably infinite.

Consider the probability law: "The probability that the outcome falls in a subinterval of $S$ is equal to the length of the subinterval," that is,

$$P[[a, b]] = (b - a) \quad \text{for } 0 \le a \le b \le 1, \tag{18}$$

where by $P[[a, b]]$ we mean the probability of the event corresponding to the interval $[a, b]$. Clearly, Axiom I is satisfied since $b \ge a \ge 0$. Axiom II follows from $S = [a, b]$ with $a = 0$ and $b = 1$.

The following shows that this probability law is consistent with notion that the outcome is equally likely to occur anywhere in the interval, thus for the events $[0, 1/2], [1/2, 1]$, and $\{1/2\}$, we have :

$$P[[0, 0.5]] = 0.5 - 0 = .5$$
$$P[[0.5, 1]] = 1 - 0.5 = .5$$

In addition, if $x_0$ is any point in $S$, then $P\left[[x_0, x_0]\right] = 0$ since individual points have zero width.

Now consider the event "the outcome is at least 0.3 away from the center of the unit interval," that is $A = [0, 0.2] \cup [0.8, 1]$. Since the two intervals are disjoint, we have by Axiom III

$$P[A] = P[[0, 0.2]] + P[[0.8, 1]] = .4.$$

---

- We now consider an example where we the initial probability assignment specifies the probability of semi-infinite intervals.

**Example 8**

Suppose that the lifetime of a computer chip is measured, and we find that "the proportion of chips whose lifetime exceeds $t$ decreases exponentially at a rate $\alpha$." Find an appropriate probability law.

Let the sample space in this experiment be $S = (0, \infty)$. If we interpret the above finding as "the probability that a chip's lifetime exceeds $t$ decreases exponentially at a rate $\alpha$," we then obtain the following assignment of probabilities to events of the form $(t, \infty)$ :

$$P[(t, \infty)] = e^{-\alpha t}, \qquad \text{for } t > 0 \tag{19}$$

where $\alpha > 0$. Note that the exponential is a number between 0 and 1 for $t > 0$, so Axiom I is satisfied. Axiom II is satisfied since

$$P[S] = P[(0, \infty)] = 1$$

The probability that the lifetime is in the interval $[r, s]$ is found by noting in Fig. 9 that $(r, s] \cup (s, \infty) = (r, \infty)$, thus by Axiom III,

$$P[(r, \infty)] = P[(r, s]] + P[(s, \infty)]$$

By rearranging the above equation we obtain

$$P[(r, s]] = P[(r, \infty)] - P[(s, \infty)] = e^{-\alpha r} - e^{-\alpha s}$$

We thus obtain the probability of arbitrary intervals in $S$.

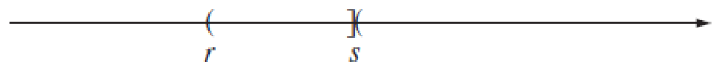



**Figure 9:** $(r, \infty) = (r, s] \cup (s, \infty)$

- In Examples 4 and 5, the probability that the outcome takes on a specific value is zero.

- How can all the outcomes in a continuuous sample space have probability zero? We can explain this paradox by using the relative frequency interpretation of probability. An event that occurs only once in an infinite number of trials will have relative frequency zero. The fact that an event has relative frequency zero, does not imply that it cannot occur, but rather that it occurs very *infrequently*. For continuous sample spaces, the set of possible outcomes is so rich that all outcomes occur infrequently enough that their relative frequencies are zero.

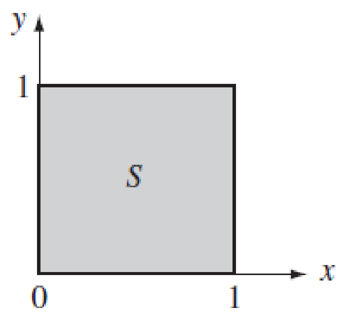
Finally, we consider an example where the events are regions in the plane.

**Example 9**

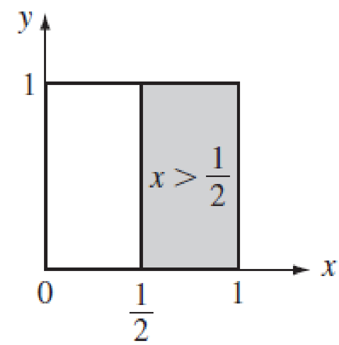
Consider experiment $E_{12}$ where we picked two numbers $x$ and $y$ at random between zero and one. The sample space is then the unit square shown in Fig. 10(a). If we suppose that all pairs of numbers in the unit square are equally likely to be selecte, then it is reasonable to use a probability assignment in which the probability of any region $R$ inside the unit square is equal to the area of $R$. Find the probability of the following events: $A = \{x > 0.5\}, B = \{y > 0.5\}$, and $C = \{x > y\}$.

Figures 10(b) and 10(c) show the regions corresponding to the events $A, B$, and $C$. Clearly each of these regions has area 1/2. Thus

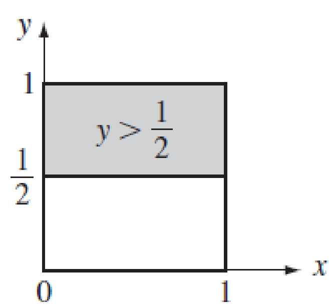
$$P[A] = \frac{1}{2}, \quad P[B] = \frac{1}{2}, \quad P[C] = \frac{1}{2}.$$



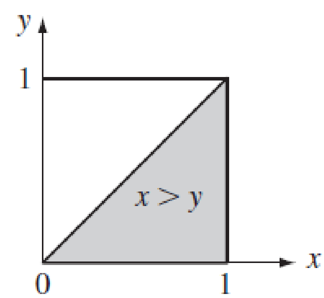

**(a)** Sample space




**(b)** Event $\{x > \frac{1}{2}\}$




**(c)** Event $\{y > \frac{1}{2}\}$




**(d)** Event $\{x > y\}$

**Figure 10:** A two-dimensional sample space and three events