The **sample space** of an experiment, sometimes denoted $S$ or in probability theory $\Omega$, is the set that consists of all possible **elementary outcomes** $\omega \in \Omega$ of that experiment (also called simple or elementary events). An **event** $A$ is a subset $A \subseteq \Omega$ of the sample space hence a union of elementary events. We say that two events $A$ and $B$ are mutually exclusive (disjoint, non-overlapping) if their intersection $A \cap B = \emptyset$ for short (the set of outcomes or elements common to both) is the empty set denoted $\emptyset$. A collection of events is mutually exclusive if any two events in the collections are disjoint. Each of the elementary events is mutually exclusive and their union is the whole sample space by definition. (Recall that the union of a collection of sets is the set of elements belonging to at least one of the sets in the collection). A collection of events is exhaustive if its union is the whole sample space. The **complement** of an event $A$ denoted $\bar{A}$ is the set of all outcomes which are not in $A$ or in other words which are in $\Omega \setminus A$ the whole space $\Omega$ minus $A$.

**Example 1**: If our experiment consists of we flip a fair coin 3 times the sample space is the set of 8 elementary outcomes consisting of a sequence of 3 head H or tail T events.  
$$\Omega = \{HHH, THH, HTH, HHT, TTH, HTT, THT, TTT\}$$

The event $A$ that we get a head on the 2nd flip of the coin consists of 4 simple events  
$$A = \{THT, HHT, HHH, THH\}$$

To say the coin is fair implies that each of the 8 elementary outcomes above is equally likely, each with probability 1/8. Then the probability of the event $A$ above that the second flip comes up heads is just the sum of the probabilities of the elementary events composing that event i.e. 4/8 or 1/2 . To say that the coin is fair means that the probability that a coin flip comes up heads is 1/2. Intuitively the relative frequency interpretation of probability means that if we do an experiment, say flip the fair coin, a large number of times, with the outcome of each experiment (flip) being independent of the preceding flips, then the fraction of times an event occurs, such as the fair coin comes up heads, is approximately the probability of the event which in this example is 1/2.

**Example 2**: **Problem 3.4** (like 3.7) An agency must decide where to locate two new computer research facilities and wants to know how many are located in California and how many in Texas. The sample space pictured in Figure 3.1 consists of the 6 pairs $\{(2,0), (1,0), (0,0), (0,1), (1,1), (0,2)\}$

Here (2,0) means two are in Texas but none in California. Etc.

Express the following events in words :

a) $F = \{((1,0), (1,1))\}$ This is the event that exactly one of the facilities is built in Texas.

b) $G = \{(0,2), (1,1), (2,0)\}$ This is the event that both facilities are built somewhere in the two states of Texas and/or California.

c) $F \cap G$ This is the event $\{(1,1)\}$ that one of the facilities is built in Texas and the other in California.

d) $\overline{F} = \{(0,0), (0,1), (2,0), (0,2)\}$ The complement of $F$ is the event that not exactly one or in other words either 0 or 2 of the facilities are built in Texas.

e) $E = \{(0,2)\}$ This is the event that both facilities are built in California.

f) $E \cap F$ This is the empty set event $\emptyset$ or said differently the events $E$ and $F$ are mutually exclusive (or disjoint or non-overlapping).

**Example 3**: (like 3.8) Decide whether it would be appropriate to use a sample space which is finite, countably infinite or continuous :
a) 8 members of a business of 100 employees are chosen to sit on the outreach council. This sample space is finite and consists of the collection of size \( \binom{100}{8} \) ways to choose 8 persons out of 100.

b) An experiment is done to measure the conductivity of a copper alloy.

This sample space is treated as continuous. While such a physical measurement is limited by the finite accuracy of the measuring device to some number of decimal places, and in this sense is discrete (finite), we can imagine devising a more accurate method and so increasing the accuracy. It is convenient to regard such measurements as continuous (For a real valued answer; there are an uncountably infinite set of real numbers in any finite interval of the number line). Whether the physical universe really is continuous or discrete is open to physical (involving either physics or a fist fight) or philosophical debate, but it is customary to idealize the situation and treat such measurements as continuous.

c) In a reliability test a light bulb is switched on and off until it fails (The tungsten core disintegrate or the switch breaks). We want to know how many times the bulb is switched before failure occurs.

Here the sample space is regarded as countably infinite. While we might not want to wait around more than two years, we really don't know in advance how many times we'll have to wait. From an ideal mathematical viewpoint it again is convenient to say that the sample space consists of all positive integers, a countably infinite set.

d) In a different light bulb test the bulb is left burning until it dies. We want to know the bulb's lifetime.

This sample space is regarded as continuous. Tungsten bulb lifetimes are approximately described by an exponential random variable. It can be shown that exponentially distributed random variables are the only ones whose probability distributions have the remarkable property that if the bulb is still burning after a 100 years say (or any other time), the length of time we have to wait until failure given that it has survived that long, has exactly the same distribution as if we were starting from time zero waiting for failure with a new light bulb. Again ignoring limits of experimental accuracy, it is convenient to regard a time measurement as a continuous variable.

Example 4: problems 3.11 and 3.12 use the Venn diagram of Figure 3.4 below: Here the numbers correspond to subregion labels not counts. Specifically the regions refer to types of motor defects with A = shaft size of the motor is too large, B = windings improper, C = electrical connections are unsatisfactory. Problem 3.11 a) Express region 5 in symbols and words: This is the intersection region
\[ 5 = B \cap \bar{A} \cap \bar{C} \] or just \( B \bar{A} \bar{C} \) for short. In words: the windings are improper but none of the other defects occur. Problem 3.12 b) a shaft is too large and the windings are improper: regions 1 or 2
or in symbols \( A \cap B = AB = ABC \cup AB \bar{C} \)
**The fundamental theorem of counting** (multiplication of choices principle):

Theorem 3.1 If sets $A_1, A_2, ..., A_k$ contain respectively $n_1, n_2, ..., n_k$ elements then the number of ways of choosing first an element of $A_1$, then an element of $A_2$, ..., and finally an element of $A_k$ is just the product of the number of elements in each set = $n_1 \cdot n_2 \cdot ... \cdot n_k$

When the size of each of the set of choices are the same, as they are in the above counting theorem, we can picture the choices by a **tree diagram** in which the branching factor is the same at each stage and given by $n_j$ the size of the set of choices at the j-th stage. Such a tree diagram is still useful even when the number of choices depends on the particular history. Then the above multiplication counting theorem fails but the tree diagram applies such as in the following problem:

**Example 5 : Tree diagram** problem 3.16.

A biomedical device can operate 0, 1, or 2 times a night. Use a tree diagram to show the 10 ways it can operate exactly 6 times in 4 nights:

```
___2__    __2__    ___2___
/                          /
0/                          /
_/1_  __1__   __2__    ___2__
\   \_1_   __2__    ___2__
  \   \_2_   __1__   ___1__
    \   ___2__   ___2__
      \   ___2__   ___2__
        \   ___2__   ___1__
          \   ___2__   ___1__
            \   __2__   __1__
              \   ___1__   ___1__
                \   __2__   __1__
                  \   ___1__   ___1__
                    \   ___2__   ___0__
```
\( nP_k = \text{“} n \text{ permute } k \text{”} = \frac{n!}{(n-k)!} = n(n-1)\cdots(n-k+1) \)

\( (n \text{ ways to chose the first, } (n-1) \text{ the } 2^{\text{nd}}, \text{ etc.}) \)

The number of unordered arrangements or combinations of \( k \) things chosen from \( n \) is the so called binomial coefficient

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

\( \binom{n}{k} \) is “\( n \) choose \( k \)” = \( \frac{nP_k}{k!} \)

To see this, note that having selected \( k \) objects, each such unordered arrangement of \( k \) objects can be arranged in \( k! \) different ordered ways (the number of permutations of \( k \) objects). Thus the number of ordered arrangements \( nP_k \) of \( k \) things chosen from \( n \) is a factor of \( k! \) times the number of unordered arrangements. It is not hard to see that the binomial coefficients satisfy the symmetry property

\[ \binom{n}{k} = \binom{n}{n-k} \quad \text{and the identity } \binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k} \]

\(<--\text{used to get “Pascal’s triangle”}\)

\[ \begin{array}{cccccc}
1 & 1 & & & & \\
1 & 2 & 1 & & & \\
1 & 3 & 3 & 1 & & \\
1 & 4 & 6 & 4 & 1 & etc
\end{array} \]

**Example 6:** problem 3.20 (like 3.17):

We have 9 cars in a race. How many different ways can they place 1st, 2nd, and 3rd? The first place winner can be chosen in any one of 9 ways, the 2nd place winner in 8 ways, the 3rd place winner in 7 ways for a total of \( 9 \cdot 8 \cdot 7 = 9P_3 = \frac{9!}{6!} = 504 \) ways.

**Example 7:** (like 3.19): 5 students can each choose from 7 different meal plans. How many meal plan assignments to these 5 students are there if the sample of size 5 from the 7 plans is done so that

a) there are no restrictions on which plan (“sampling with replacement”). In this case each student can choose in 7 ways so there are \( 7^5 = 16807 \) ways by the multiplication principle.

b) No two of the students can choose from the same meal plan. (“sampling without replacement”) In this case the first can choose in 7 ways the second in 6 ways etc. so \( 7P_5 = \frac{7!}{2!} = 2520 \) ways.

c) If now 7 students choose from the 7 meal plans, how many ways can they do this if no two students can choose from the same plan. Then there is exactly one plan per student so the answer is the number of permutations of 7 plans i.e. \( 7! \) ways total where \( 7! = 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 5040 \) ways.

**Example 8:** problem 3.24 (like 3.23)

The number of ways that 4 of 18 robotic arms can be chosen for a welding job when order doesn’t matter, only which 4 arms is

\[ \binom{18}{4} = \frac{18!}{4!15!} = 18 \cdot 17 \cdot 16 \cdot 15 \]

\[ = 18 \cdot 17 \cdot 2 \cdot 5 = (20-2)(20-3)10 = 3060 \quad \text{ways}. \]

**Example 9:** a) How many 5 card poker hands can be chosen from a 52 card deck when order doesn’t matter?

\[ \binom{52}{5} = \frac{52!}{5!47!} = \frac{52 \cdot 51 \cdot 50 \cdot 49 \cdot 48}{5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} \]

\[ = 52 \cdot 51 \cdot 10 \cdot 49 \cdot 2 = (50+2)(50+1)(50-1)20 = 2,599,000 \quad \text{ways} \]

b) Same question but for ordered poker hands: there are then 5! = 120 times as many ways as the
Example 10: (similar to 3.26 b)
a) Given we have 12 batteries of which 4 are defective, how many ways are there to choose 5 batteries from these 12 such that exactly two are defective? Note that this means we are choosing the other 3 of the 5 batteries picked from the 8 batteries which are non-defective. By our basic counting principle there are \( \binom{8}{3} \binom{4}{2} = \frac{8 \cdot 7 \cdot 6}{3 \cdot 2 \cdot 1} \cdot \frac{4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1} = 8 \cdot 7 \cdot 6 = 336 \) ways to do this.

b) What is the probability that if we randomly select 5 batteries from the 12 as above that we will obtain exactly two defective as above? Assuming that random here means that all selections are equally likely each with probability \( \frac{1}{12 \cdot 5} \) the answer is then \( \frac{8}{3} \cdot \frac{4}{2} \cdot \frac{12}{5} = 14/33 \).

The relative frequency interpretation of probability says that the probability of an event is the fraction of times the event occurs relative to the total number of times the experiment is performed in the limit when the experiment is repeated many times (\( n \to \infty \)). This definition of probability has its drawbacks. How do we know that the fraction stabilizes to a fixed limiting value for large \( n \)? The modern axiomatic treatment of probability avoids this difficulty by insuring that this limit is a consequence of the axioms, i.e. the law of large numbers follows from the axioms. If we flip a fair coin repeatedly and consider the random variable \( X_i = \begin{cases} 1 & \text{if the } i^{th} \text{ flip is a head} \\ 0 & \text{else} \end{cases} \) then the proportion of heads is just the average \( \bar{X} \) of these \( X_i \)'s and the classical law of large numbers says that as \( n \to \infty \) the sample mean approaches the population mean i.e.

\[
\bar{X} = \frac{S_n}{n} = \frac{x_1 + x_2 + \ldots + x_n}{n} \to \mu = EX_1 = \text{expected value of } X_1
\]

\[= P( \text{a given flip is a head} ) = 1/2 \] (The expected value of a 0 or 1 valued random variable is just the probability that 1 occurs as we will see when we study expected values.) Thus the relative frequency interpretation is justified by the axioms. The properties of relative frequencies also motivate the choice of axioms if we assume that the frequencies do in fact stabilize to a limit for large \( n \).

Properties of relative frequencies: Consider a finite collection of \( k \) disjoint events \( A_i \) with union \( A_1 \cup A_2 \cup \ldots \cup A_k \) then if we perform an experiment a large number \( n \) of times and count the relative frequencies \( rf_n(A_j) \) of each event i.e. the fraction which is the number of times the event occurs over the total number \( n \) of times, then we observe from basic properties of counting that

1) \( 0 \leq rf_n(A_j) \leq 1 \)
2) \( rf_n(\Omega) = 1 \) i.e. if any individual outcome of an experiment occurs which must certainly happen (\( n \) times out of \( n \) is the fraction 1) then by definition of union, the event which is the union of all such outcomes i.e. which is the sample space by definition, has occurred.
3) \( rf_n(A_1 \cup A_2 \cup \ldots \cup A_k) = rf_n(A_1) + rf_n(A_2) + \ldots + rf_n(A_k) \) follows from the disjointness of the sets since the counts of each event are just the number of elementary outcomes in each and these are distinct outcomes (which do not get counted twice).

Property 2) can be viewed as a consequence of 3) if we acknowledge that when the events \( A_j \) are exhaustive so that \( A_1 \cup A_2 \cup \ldots \cup A_n = \Omega \) then their frequencies must add to 1. This last property says that the relative frequency is a (finitely) additive set function. We must generalize this somewhat to the limiting case where \( n \to \infty \) as in the relative frequency interpretation of probability. That is we must allow a countable collection of disjoint sets so that \( k = \infty \). When countable unions are allowed
then we say that the limiting relative frequency i.e. the probability is a **countably additive set function**.

If limits of the above relative frequencies are to exist, by the properties of limits, we are drawn to the following **axioms of probability**:

1. $0 \leq P(A) \leq 1$ for any event $A$
2. $P(\Omega) = 1$ The event that the observed outcome is part of the sample space is sure to happen
3. For any countable sequence of disjoint events $A_k, k = 1, 2, 3, \ldots$ the probability of their union satisfies **countable additivity**:
   \[
   P(A_1 \cup A_2 \cup \ldots A_k \cup \ldots) = \sum_{k=0}^{\infty} P(A_k)
   \]

**Consequences of the axioms of probability**:

a) **Finite additivity**: Note that with the choice of $A_k = \emptyset$ if $k \geq n$ the countable additivity property reduces to the finite additivity case (i.e. to property 3 stated for relative frequencies with rf replace by $P$)

\[
P(A_1 \cup A_2 \cup \ldots A_k) = \sum_{j=1}^{k} P(A_j) = P(A_1) + P(A_2) + \ldots + P(A_k)
\]

(Theorem 3.4 of text) The probability $P(A)$ of any event $A$ is the sum of the probabilities of the elementary events (individual outcomes) comprising $A$.

b) **Probability of complements**: $P(\overline{A}) = 1 - P(A)$ (Theorem 3.7 of text)

To see this holds we note since for any event $A$ one has $A \cup \overline{A} = \Omega$ is a disjoint exhaustive union, properties 2) and 3) above imply that $P(A) + P(\overline{A}) = 1$ which equivalently yields the above.

c) **Inclusion-exclusion formula** or **general addition rule** for probability: for two sets this is the formula

\[
P(A \cup B) = P(A) + P(B) - P(A \cap B)
\]

(Theorem 3.6 of text)

To see this note that for any set $B = (B \cap A) \cup (B \cap \overline{A})$ is a disjoint union, and hence so is the right hand side of the equality $A \cup B = A \cup (B \cap \overline{A})$ By property 2) above we thus find $P(B) = P(B \cap A) + P(B \cap \overline{A})$ and $P(A \cup B) = P(A) + P(B \cap \overline{A})$ or equivalently $P(B \cap \overline{A}) = P(B) - P(B \cap A)$ and hence combining these we arrive at the stated formula.

With the help of a Venn diagram one can write down an analogous formula for any finite union of sets. Your homework problem 3.49 does this in the case of the union of three sets.

A completely similar relation holds for the number of elements in the union of two (or more) finite sets in terms of the number in each set separately and in their intersection(s). This is not surprising since probability reduces to counting in the finite set case. Namely denoting the size of a set $a$ by $|A|$ we have
for the case of two sets : 
|A ∪ B| = |A| + |B| − |A ∩ B| .

**Example 11 : Problem 3.28 a)**: A refrigerator manufacturer sold 2756 units of a new model and 287 of these required repairs under the warranty. Estimate the probability that a new unit which has just been sold will require repairs under warranty :

Since n= 2756 is a fairly large sample size, the relative frequency interpretation of probability says that the sample proportion \( \frac{287}{2756} \approx 0.1041 \) fraction of the random sample that required repairs under warranty is a good estimate of the actual probability of needing repairs under warranty. So about 10.4 percent chance.

**Example 12 : Problem 3.33 : a)**: Of 160 graduating engineering students, 92 are enrolled in an advanced course (A), 63 in an operations research course (B) and 40 are enrolled in both (A ∩ B).

How many students are not enrolled in either course ? 
|A ∪ B| = |A| + |B| − |A ∩ B| 
= 92 + 63 − 40 = 115 . We want |\bar{A} ∩ \bar{B}| = complement of A ∪ B | = 160 − 115 = 45 students.

b) What is the probability that a randomly selected student from this group of 160 students is not enrolled in either course? Random in the context of this finite sample space means that each of the 160 students is equally likely to be selected. The probability of any particular student being selected (an elementary event) is 1/160. So the probability that the randomly selected student lies inside the event which is the union of the 45 elementary events i.e that a student is selected who is enrolled in neither course is 45/160 = 9/32.

c) What is the probability that two randomly selected students from this group are not enrolled in either course? Random here means that each of the \( \binom{160}{2} = 80 \cdot 159 \) choices of 2 students from 160 is equally likely to occur (each with probability \( \frac{1}{\binom{160}{2}} = \frac{1}{80 \cdot 159} \)). So the probability is the number of ways to select two students from the 45 not enrolled in either course over the total number of ways to select two students from 160 or \( \frac{\binom{45}{2}}{\binom{160}{2}} = \frac{45 \cdot 22}{80 \cdot 159} = \frac{99}{1272} \).

**Example 13 : a)**: What is the probability of event A that a 5 card poker hand contains one or more aces. This is the probability of the complement of the event that the hand contains no aces.

\[ P( \text{no aces} ) = \binom{48}{5}/\binom{52}{5}. \] Thus \[ P( A ) = P( \text{1 or more ace} ) = 1 - P( \text{no aces} ) = \binom{52}{5}/\binom{48}{5} - \binom{48}{5}/\binom{52}{5} \]

where the numerator \( \binom{52}{5} − \binom{48}{5} \) is the number of hands having 1 or more aces (i.e. not having zero aces). We could also have computed this by noting that the event 1 or more aces is the union of the 4 events : exactly 1 ace (hence 4 non-aces chosen from 48 non-aces), exactly two aces (hence 3 non-aces), exactly 3 aces or exactly 4 aces. By our basic counting principles we then must have

\[ \binom{52}{5}/\binom{48}{5} = \binom{4}{1} \cdot \binom{48}{4} + \binom{4}{2} \cdot \binom{48}{3} + \binom{4}{3} \cdot \binom{48}{2} + \binom{4}{4} \cdot \binom{48}{1} \]
b) What is the probability of the event B that we are dealt a full house consisting of 3 aces and 2 kings? The probability is the number of ways to choose 3 aces from the 4 aces in the deck times the ways to choose 2 kings from 4 over the total number of 5 card poker hands or
\[
\binom{4}{3} \cdot \binom{4}{2} / \binom{52}{5}.
\]

c) What is the probability of any full house (3 of one kind two of another)? This is the same as in b) above except there are 13 ways to choose the kind that we have 3 of and 12 ways to choose the kind we have 2 of (since it can’t be the kind we already picked). So probability
\[
13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2} / \binom{52}{5}.
\]

d) What is the conditional probability \(P(B | A)\) of the full house in b) if we are told by the dealer before he gives us our 5 cards that our hand has at least 1 ace in it already. This is a conditional probability problem conditioned on the event \(A\) that our hand has at least one ace in it. In this case we saw in part a) that the number of such hands with one or more ace is the total number of hands minus the number having no aces or \(\binom{52}{5} - \binom{48}{5}\) hands. These hands constitutes the reduced sample space for the conditional event problem. We can assume that our full house is then randomly selected from these hands with each such hand being equally likely. This yields the number of full houses aces over kings found in the numerator of b) above over the size of the reduced sample space or
\[
P(B | A) = \frac{\binom{4}{3} \cdot \binom{4}{2}}{\binom{52}{5} - \binom{48}{5}} = \frac{P(B \cap A)}{P(A)}.
\]

To see the last equality, note that by dividing both the numerator and denominator above by the total number of poker hands \(\binom{52}{5}\) hands this is the same as the answer in b) divided by the answer in a) i.e. it is equal to \(P(B) / P(A)\) and since the full house containing 3 aces that is the event \(B\) is contained in the event \(A\) that at least one ace occurs we also have \(B = B \cap A\). This is how we will define conditional probability in general. I.e. we have motivated the following definition.

**Definition of conditional probability**: for any two events \(A\) and \(B\) with \(P(A)\) non-zero we define \(P(B | A)\) as \(\frac{P(B \cap A)}{P(A)}\). If we let \(p(B) = P(B | A)\) one checks that the set function measure \(p(B)\) satisfies the 3 axioms needed for it to be a probability measure. Namely it satisfies
1) \(0 \leq p(A) \leq 1\)
2) \(p(\Omega) = p(A) = \frac{P(A)}{P(A)} = 1\) (i.e. since \(\Omega \cap A = A\) the sample space \(\Omega\) has measure 1)
3) countable additivity for \(p(\cdot)\) follows directly from the same property of \(P(\cdot)\).

Thus a conditional probability measure really is a probability measure.