Lecture 3: Intro to Probability Theory
Math Boot Camp

Will Terry

Department of Political Science
University of Oregon

September 19, 2012
Overview of Today’s Lecture

1. Probability
   - Some Background

2. Probability in Simple Sample Spaces
   - Sample Space and Events
   - Probability of an Event
   - Counting Techniques

3. Conditional Probability
   - Conditional and Compound Probabilities
   - Bayes’ Theorem
   - Independent Events
Overview of Today’s Lecture

1. Probability
   - Some Background

2. Probability in Simple Sample Spaces
   - Sample Space and Events
   - Probability of an Event
   - Counting Techniques

3. Conditional Probability
   - Conditional and Compound Probabilities
   - Bayes’ Theorem
   - Independent Events
Lady Luck

- When the Greeks wanted a prediction about what tomorrow might bring, they turned to the oracles instead of consulting their wisest philosophers.
- Up to the time of the Renaissance, people perceived the future as little more than a matter of luck or the result of random variations, and most of their decisions were driven by instinct.
- The inappropriateness of the numbering system was important reason of why Europeans were not induced to explore the mastery of risk.
- Without numbers, there are no odds and no probabilities; without odds and probabilities, the only way to deal with risk is to appeal to the gods and the fates.
Lady Luck

The radical concept of replacing randomness with systematic probability and its implicit suggestion that the future might be predictable and even controllable to some degree, is what separates the modern times and the past.

Yet, probability theory was not developed until the XVII century.

And, it was a result of the efforts to analyze games of chance!
Gambling in the Dark

- Gambling has been a popular pastime and often an addiction since the beginning of recorded history. However, until the XVII century, people had wagered and played games of chance without using any system of odds to determine winnings and losings.

- The Greeks understood that more things might happen in the future than actually will happen.

- In fact, the word, ευκοτς, meaning plausible or probable, had the same sense as the modern concept of probability: “to be expected with some degree of certainty.”

- However, they never made any advances in measuring this degree of certainty.
Today’s Class

- In today’s class we will apply the calculus of probabilities to a wide variety of situations requiring the use of sophisticated counting techniques.
- We shall often refer to games of chance. Yet these games have applications that extend far beyond the spin of the roulette wheel.
Overview of Today’s Class

1. Probability
   - Some Background

2. Probability in Simple Sample Spaces
   - Sample Space and Events
   - Probability of an Event
   - Counting Techniques

3. Conditional Probability
   - Conditional and Compound Probabilities
   - Bayes’ Theorem
   - Independent Events
Introduction

In everyday life, we often talk loosely about chance. What are the chances of getting a job? of meeting someone? of rain tomorrow? But for the purposes of this class, we need to give the word *chance* a definite, clear interpretation. This turns out to be hard, and, as we learned today, mathematicians have struggled with the job for centuries. We will now focus on the *frequency theory*, which works best for processes which can be repeated over and over again, independently and under the same conditions. Games of chance fail into category. And, as we already know, much of the frequency theory was developed to solve gambling problems.
Tossing Coins

One of the simplest games of chance involves betting on the toss of a coin.

- If a coin is tossed in the air, then it is certain that the coin will come down, but it is not certain that, say, it will come up “head”.

- When trying to figure chances, it is usually very helpful to list all the possible ways that a chance process can turn out.

- In the case of a coin toss, we ordinarily agree to regard “head” and “tail” as the only possible outcomes.

- If we denote these outcomes by $H$ and $T$ respectively, then each outcome would correspond to exactly one of the elements of the set $\{H, T\}$. 
Sample Spaces

Suppose now that we repeat this experiment of tossing a coin.

- The process of tossing the coin can certainly be repeated over and over again, independently and under the same conditions: the outcomes of this experiment would still correspond to exactly one of the elements of the set \( \{H, T\} \).
- This set, \( \Omega = \{H, T\} \) is called a sample space for the experiment.

Suppose now that we toss a die and observe the number that appears on top. Then the sample space consist of the six possible numbers: \( \Omega = \{1, 2, 3, 4, 5, 6\} \).
<table>
<thead>
<tr>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>The set $\Omega$ of all possible outcomes associated with a real or conceptual experiment is called the <em>sample space</em>. A particular outcome $\omega$, i.e. an element in $\Omega$, is called a <em>sample point</em> or <em>sample</em>.</td>
</tr>
</tbody>
</table>
Tossing Coins (again)

Suppose now that we perform the following experiment: we toss a coin three successive times.
Let $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$ be the associated sample space.

- We may be interested in the event, "the number of heads exceeds the number of tails".
- For any outcome of the experiment we can determine whether this event does or does not occur.
- We find that $HHH, HHT, HTH$ and $THH$ are the only elements of $\Omega$ corresponding to outcomes for which this event does occur.
Events

To say that the event “the number of heads exceeds the number of tails” occurs is the same as saying the experiment results in an outcome corresponding to an element of the set $A = \{HHH, HHT, HTH, THH\}$.

- Notice that $A$ is a subset of the sample space $\Omega$.

**Definition**

An event $A$ is a set of outcomes or, in other words, a subset of some underlying sample space $\Omega$. 
Events

- The event \( \{\omega\} \) consisting of a single sample \( \omega \in \Omega \) is called an \textit{elementary event}.
- The \textit{certain} (or \textit{sure}) event, which always occurs regardless of the outcome of the experiment, is formally identical with the whole space \( \Omega \).
- The \textit{impossible event} is the empty set \( \emptyset \), containing none of the elementary events \( \omega \).
Events

We can combine events to form new events using the various set operations:

(i) \( A_1 \cup A_2 \) is the event that occurs iff \( A_1 \) occurs or \( A_2 \) occurs (or both);

(ii) \( A_1 \cap A_2 \) is the event that occurs iff \( A_1 \) occurs and \( A_2 \) occurs;

(iii) \( A' = A - \Omega \), the complement of \( A \), is the event that occurs iff \( A \) does not occur.
Events

Given two events $A_1$ and $A_2$, suppose $A_1$ occurs if and only if $A_2$ occurs.

- Then, $A_1$ and $A_2$ are said to be \textit{identical} (or \textit{equivalent}), and we write $A_1 = A_2$.

Two events $A_1$ and $A_2$ are called \textit{mutually exclusive} if they are disjoint, i.e. if $A_1 \cap A_2 = \emptyset$.

- In other words, $A_1$ and $A_2$ are mutually exclusive if they cannot occur simultaneously.
Events: Example

Example

Suppose that we toss a die and observe the number that appears on top. The sample space consists of the six possible numbers: \( \Omega = \{1, 2, 3, 4, 5, 6\} \). Let \( A_1 \) be the event that an even number occurs, \( A_2 \) that an odd number occurs, and \( A_3 \) that a number higher than 3 occurs.

Then, \( A_1 = \{2, 4, 6\} \), \( A_2 = \{1, 3, 5\} \), and \( A_3 = \{4, 5, 6\} \). And,

\[
A_1 \cup A_3 = \{2, 4, 5, 6\} \text{ is the event that an even number or a number higher than 3 occurs;}
\]

\[
A_1 \cap A_3 = \{4, 6\} \text{ is the event that an even number higher than 3 occurs;}
\]

\[
A_3' = \{1, 2, 3\} \text{ is the event that a number higher than 3 does not occur.}
\]
Events: Example

Note that $A_1$ and $A_2$ are mutually exclusive: $A_1 \cap A_2 = \emptyset$.

In other words, an even number and an odd number cannot occur simultaneously.
Overview of today’s lecture

1. Probability
   - Some Background

2. Probability in Simple Sample Spaces
   - Sample Space and Events
   - Probability of an Event
   - Counting Techniques

3. Conditional Probability
   - Conditional and Compound Probabilities
   - Bayes’ Theorem
   - Independent Events
Probability of an Event: Definition

Let $\Omega$ be a finite sample space; say, $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$.

A finite probability space is obtained by assigning to each sample point $\omega_i \in \Omega$ a real number $p_i$, called the probability of $\omega_i$, satisfying the following properties:

(i) each $p_i$ is non-negative, $p_i \geq 0$

(ii) the sum of the $p_i$ is one, $\sum_{i=1}^{n} p_i = 1$

The probability $P(A)$ of an event $A$, is then defined to be the sum of the probabilities of the sample points in $A$. For notational convenience we write $P(\omega_i)$ for $P(\{\omega_i\})$. 
Probability of an Event: Example

Example

Let three coins be tossed and the number of heads observed; then \( \Omega = \{0, 1, 2, 3\} \) is a sample space associated with this experiment.

We obtain a probability space by the following assignment, \( P(0) = \frac{1}{8}, P(1) = \frac{3}{8}, P(2) = \frac{3}{8}, \) and \( P(3) = \frac{1}{8} \) since each probability is nonnegative and the sum of the probabilities is 1.

Let now \( A_1 \) be the event that at least one head appears and let \( A_2 \) be the event that all heads or all tails appear: \( A_1 = \{1, 2, 3\} \) and \( A_2 = \{0, 3\} \).

Then, by definition, \( P(A_1) = P(1) + P(2) + P(3) = \frac{3}{8} + \frac{3}{8} + \frac{1}{8} = \frac{7}{8} \)
and \( P(A_2) = P(0) + P(3) = \frac{1}{8} + \frac{1}{8} = \frac{1}{4} \).
Note that the probability of an event depends on the previous assignment of probabilities to the sample points.

- The following question then arises: which assignment of probabilities to sample points should be made?
- The answer to this question is not a mathematical one. Rather, it depends upon our assessment of the real-world situation to which the theory is to be applied.
Axioms of Probability

Let $\Omega$ be a sample space, let $C$ be the class of all events, and let $P$ be a real-valued function defined on $C$. Then $P$ is called a **probability function**, and $P(A)$ is called the **probability** of the event $A$ if the following axioms hold:

1. The probability of a certain event is 1, $P(\Omega) = 1$.
2. If $A$ is any event, then $0 \leq P(A) \leq 1$.
3. If $A_1$ and $A_2$ are mutually exclusive events, then

$$P(A_1 \cup A_2) = P(A_1) + P(A_2).$$
Axioms of Probability

More generally, given $n$ mutually exclusive events $A_1, A_2, \ldots, A_n$, we have the formula

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

This equation is called the \textit{addition rule for probabilities}.

We can now prove a few theorems which follow directly from our axioms.
Axioms of Probability

**Theorem**

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

**Proof.** Let $A$ be any set; then $A$ and $\emptyset$ are disjoint and $A \cup \emptyset = A$. By axiom 3,

$$P(A) = P(A \cup \emptyset) = P(A) + P(\emptyset)$$

Subtracting $P(A)$ from both sides gives our result.
Axioms of Probability

Theorem
If $A'$ is the complement of an event $A$, then $P(A') = 1 - P(A)$.

In words, the probability that $A$ does not occur is obtained by subtracting from 1 the probability that $A$ does occur.

Proof. The sample space $\Omega$ can be decomposed into mutually exclusive events $A$ and $A'$; that is, $\Omega = A \cup A'$. By axioms 1 and 3 we obtain

$$1 = P(\Omega) = P(A \cup A') = P(A) + P(A')$$

from which our result follows.
Axioms of Probability

Theorem

If \( A_1 \subseteq A_2 \), then \( P(A_1) \leq P(A_2) \).

- In words, if \( A_1 \) implies \( A_2 \), then the probability of \( A_1 \) cannot exceed the probability of \( A_2 \).

Proof. If \( A_1 \subseteq A_2 \), then \( A_2 \) can be decomposed into the mutually exclusive events \( A_1 \) and \( A_2 \setminus A_1 \). Therefore,

\[
P(A_2) = P(A_1) + P(A_2 \setminus A_1)
\]

The result now follows from the fact that \( P(A_2 \setminus A_1) \geq 1 \).
Theorem

If $A_1$ and $A_2$ are any two events, then

$$P(A_1 \setminus A_2) = P(A_1) - P(A_1 \cap A_2).$$

Proof. Now $A_1$ can be decomposed into the mutually exclusive events $A_1 \setminus A_2$ and $A_1 \cap A_2$; that is $A_1 = (A_1 \setminus A_2) \cup (A_1 \cap A_2)$. Thus by axiom 3,

$$P(A_1) = P(A_1 \setminus A_2) + P(A_1 \cap A_2)$$

from which our result follows.
The addition rule (axiom 3) becomes a little more complicated if we drop the requirement that the events be mutually exclusive. The following theorem, called the *general addition law* is similar to the inclusion-exclusion principle for sets.

**Theorem**

*If* $A_1$ *and* $A_2$ *are any two events, then*

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

In words, the probability that at least one of the events $A_1$ and $A_2$ occurs is obtained by adding the probability that $A_1$ occurs and the probability that $A_2$ occurs, and then subtracting the probability that both $A_1$ and $A_2$ occur.
Proof. Note that \( A_1 \cup A_2 \) can be decomposed into the mutually exclusive events \( A_1 \setminus A_2 \) and \( A_2 \); that is \( A_1 \cup A_2 = (A_1 \setminus A_2) \cup A_2 \). Thus by axiom 3 and our previous theorem,

\[
P(A_1 \cup A_2) = P(A_1 \setminus A_2) + P(A_2) \\
= P(A_1) - P(A_1 \cap A_2) + P(A_2) \\
= P(A_1) + P(A_2) - P(A_1 \cap A_2)
\]

which is the desired result.
For any events $A_1$, $A_2$, $A_3$,

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) + P(A_1 \cap A_2 \cap A_3).$$

More generally, given any $n$ events $A_1$, $A_2$, ..., $A_n$,

$$P\left( \bigcup_{k=1}^{n} A_k \right) = P_1 - P_2 + P_3 - P_4 + \ldots \pm P_n.$$
The following examples illustrate how our formulas can be used to compute probabilities:

**Example**

A card is dealt off the top of a well-shuffled ordinary deck of 52 cards. There is 1 chance in 4 for it to be a heart. There is 1 chance in 4 for it to be a spade.

- What is the chance for it to be in a major suit (hearts or spades)?

The question asks for the chance that one of the following two things will happen:

- the card is a heart;
- the card is a spade.
Well, if the card is a heart then it cannot be a spade: these are mutually exclusive events.

- So it is legitimate to add the chances.
- The chance of getting a card in a major suit is \( \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \).

(A check on the reasoning: there are 13 hearts and 13 spades, so \( \frac{26}{52} = \frac{1}{2} \) of the cards in the deck are in a major suit.)
Example

Three coins are tossed. We assign equal probabilities to the eight sample points in the sample space, \(\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}\).

- What is the chance of getting at least one head?
- If \(A\) is the event “at least one head,” then the complementary event \(A'\) is “no heads.” By theorem 2,

\[
P(A) = 1 - P(A') \\
= 1 - P(\{TTT\}) \\
= 1 - \frac{1}{8} = \frac{7}{8}.
\]
Example

Someone throws a pair of dice. True or false: The chance of getting at least one ace is \( \frac{1}{6} + \frac{1}{6} = \frac{1}{3} \).

This is false. Imagine one of the dice is white, the other red.

- The question asks for the chance that one of the two following things will happen:
  - the white die lands ace;
  - the red die lands ace.

A white ace does not prevent a red ace. These two events are not mutually exclusive, so the simple addition rule does not apply. Adding the chances gives the wrong answer.
Look at the following figure: there are six ways for the white die to show an ace; there are six ways for the red die to show an ace; but the number of ways to get at least one ace is not $6 + 6$.

- Addition double counts the outcome “snake eyes” at the top left corner.
The chance of getting at least one ace is: \( \frac{6+6-1}{36} = \frac{11}{36} \), or 30.5% (and not 33.3%).

If you want to find the probability that at least one event occurs, and the events are not mutually exclusive, do not add the probabilities; the sum will be too big.
As some of our examples show, many times the physical characteristics of an experiment suggest that the various outcomes of the sample space be assigned equal probabilities.

- Such a finite probability space $\Omega$, where each sample point has the same probability, is called an *equiprobable* or *uniform space*.
- In particular, if $\Omega$ contains $n$ points then the probability of each point is $\frac{1}{n}$.
- Furthermore, if an event $A$ contains $r$ points then its probability is $r \cdot \frac{1}{n} = \frac{r}{n}$. 
In other words, the probability \( P(A) \) of the event \( A \) is defined as the fraction of the outcomes in which \( A \) occurs:

\[
P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } \Omega} = \frac{n(A)}{n(\Omega)}.
\]

For example, in tossing a die an even number can occur in 3 ways out of 6 “equally likely” ways; hence \( p = \frac{3}{6} = \frac{1}{2} \).
Theorem

Given a finite sample space $\Omega$, for any subset $A$ of $\Omega$, let $P(A) = \frac{n(A)}{n(\Omega)}$. Then, $P$ satisfies the probability axioms presented above.

The expression “at random” will be used only with respect to an equiprobable space; formally, the statement “choose a point at random from a set $\Omega$” shall mean that $\Omega$ is an equiprobable space, i.e. that each sample point in $\Omega$ has the same probability.
Example

A card is dealt off the top of a well-shuffled deck. Let $A_1 = \{\text{the card is a spade}\}$, and $A_2 = \{\text{the card is a face card (i.e. a jack, queen or king)}\}$. What is the chance for it to be a face card of spades?

We compute $P(A_1)$, $P(A_2)$, and $P(A_1 \cap A_2)$. Since we have an equiprobable space,

$$P(A_1) = \frac{\text{number of spades}}{\text{number of cards}} = \frac{13}{52} = \frac{1}{4},$$

$$P(A_2) = \frac{\text{number of face cards}}{\text{number of cards}} = \frac{12}{52} = \frac{3}{13},$$
\[ P(A_1 \cap A_2) = \frac{\text{number of spade face cards}}{\text{number of cards}} = \frac{3}{52}. \]
1. Probability
   - Some Background

2. Probability in Simple Sample Spaces
   - Sample Space and Events
   - Probability of an Event
   - Counting Techniques

3. Conditional Probability
   - Conditional and Compound Probabilities
   - Bayes’ Theorem
   - Independent Events
The rule for computing probabilities delineated above is applicable only when all sample points have been assigned the same probability.

- Thus, the formula does not apply to a wide variety of important problems where it is not reasonable to make this special assignment of probabilities to sample points.
Suppose a license plate contains two distinct letters followed by three digits with the first digit not zero.

- How many different license plates can be printed?
- The first letter can be printed in 26 different ways, the second letter in 25 different ways (since the letter printed first cannot be chosen for the second letter), the first digit in 9 ways and each of the other two digits in 10 ways.
- Hence

\[ 26 \times 25 \times 9 \times 10 \times 10 = 585,000 \]

different plates can be printed.
If some procedure can be performed in $n_1$ different ways, and if, following this procedure, a second procedure can be performed in $n_2$ different ways, and if, following this second procedure, a third procedure can be performed in $n_3$ different ways, and so forth; then the number of ways the procedures can be performed in the order indicated is the product $n_1 \times n_2 \times n_3$....
The product of the positive integers from 1 to \( n \) inclusive occurs very often in mathematics and hence is denoted by the special symbol \( n! \) (read “\( n \) factorial”):

\[
n! = 1 \times 2 \times 3 \times \ldots \times (n - 2)(n - 1)n
\]

In words, we use the exclamation mark (!) to indicate the result of multiplying together a number and all the numbers which come before it. So, for example: \( 4! = 4 \times 3 \times 2 \times 1 = 24 \) (and is read “four-factorial equal twenty-four.”)

It is also convenient to define \( 1! = 1 \), and \( 0! = 1 \) (it is a convention of mathematics that zero-factorial equals 1).
Permutations

An arrangement of a set of \( n \) objects in a given order is called a *permutation* of the objects (taken all at a time).

- An arrangement of any \( r \leq n \) of these objects in a given order is called an \( r \)-permutation or a permutation of the \( n \) objects taken \( r \) at a time.
Permutations

Example

Consider the set of letters $a, b, c,$ and $d$. Then:

(i) $bdca, dcba, \text{ and } acdb$ are permutations of the 4 letters (taken all at a time);

(ii) $bad, adb, cbd, \text{ and } bca$ are permutations of the 4 letters taken 3 at a time;

(iii) $ad, cb, da, \text{ and } bd$ are permutations of the 4 letters taken 2 at a time.
Permutations

The number of permutations of \( n \) objects taken \( r \) at a time will be denoted by

\[
P(n, r)
\]

- Before we derive the general formula for \( P(n, r) \) we consider a special case:
- Suppose we want to find the number of permutations of 6 objects, say \( a, b, c, d, e, f \), taken three at a time. In other words, we want to find the number of “three letter words” with distinct letters that can be formed from the above six letters.
Permutations

Let the general three letter word be represented by three boxes:

□ □ □

Now the first letter can be chosen in 6 different ways; following this, the second letter can be chosen in 5 different ways; and, following this, the last letter can be chosen in 4 different ways.

- If we write each number in its appropriate box, they should look like this:

   6 5 4
Permutations

By the fundamental principle of counting there are $6 \times 5 \times 4 = 120$ possible three letter words without repetitions from the six letters, or there are 120 permutations of 6 objects taken 3 at a time. That is,

$$P(6, 3) = 120$$

The derivation of the formula for $P(n, r)$ follows the procedure in the preceding example.

- The first element in an $r$-permutation of $n$-objects can be chosen in $n$ different ways; following this, the second element in the permutation can be chosen in $n - 1$ ways; and following this, the third element in the permutation can be chosen in $n - 2$ ways.
Permutations

Continuing in this manner, we have that the $r^{th}$ (last) element in the permutation can be chosen in $n - (r - 1) = n - r + 1$ ways.

**Theorem**

$$P(n, r) = n(n - 1)(n - 2) \cdots (n - r + 1) = \frac{n!}{(n-r)!}$$

The second part of the formula follows from the fact that

$$P(n, r) =$$

$$= \frac{n(n - 1)(n - 2) \cdots (n - r + 1) \times (n - r)!}{(n - r)!}$$

$$= \frac{n!}{(n-r)!}$$

So, $P(6, 3) = \frac{6!}{3!} = \frac{720}{6} = 120$. 
Permutations

In the special case that $r = n$, we have

$$P(n, n) = n(n - 1)(n - 2) \cdots 3 \times 2 \times 1 = n!$$

- Namely, there are $n!$ permutations of $n$ objects (taken all at a time).
- For example, if we have 3 objects, say $a, b, c$, there are 3! permutations: $abc, acb, bac, bca, cab, cba$. 
Permutations with repetitions

Suppose we want to know the number of permutations of objects some of which are alike.

- For example, say an urn contains one red ball and nine green ones. Five draws are made at random. Each time a ball is drawn, we put it back into the box.

We interested in the situation where two red balls (no more and no less) are drawn. One way this can happen is that the first two draws are red an the final three are green.

- With $R$ for red and $G$ for green, this possibility can be written

$$\text{R R G G G}$$
Permutations with repetitions

Of course, there are many other ways to get two reds. For example, the second and the fifth draws might be red, while all the rest are green:

\[ G \ R \ G \ G \ R \]

To figure out all the possible ways in which two red balls are drawn, we must first calculate the chance of each possible way, and then use the addition rule to add up the chances.

- The chance of the pattern R R G G G is:

\[
\frac{1}{10} \times \frac{1}{10} \times \frac{9}{10} \times \frac{9}{10} \times \frac{9}{10} = \left( \frac{1}{10} \right)^2 \left( \frac{9}{10} \right)^3
\]
Permutations with repetitions

Similarly, the chance of the pattern G R G G R equals

\[
\frac{9}{10} \times \frac{1}{10} \times \frac{9}{10} \times \frac{9}{10} \times \frac{1}{10} = \left( \frac{1}{10} \right)^2 \left( \frac{9}{10} \right)^3
\]

The pattern G R G G R has the same chance as the pattern R R G G G. In fact, each pattern with 2 reds and 3 greens has the same chance, \( \left( \frac{1}{10} \right)^2 \left( \frac{9}{10} \right)^3 \), since the 2 reds will contribute \( \left( \frac{1}{10} \right)^2 \) to the product and the 3 greens will contribute \( \left( \frac{9}{10} \right)^3 \).

- The sum of the chances of all the patterns, therefore, equals the number of patterns times the common chance.
Permutations with repetitions

How many patterns are there? Each pattern is specified by writing down in a row 2 R’s and 3 G’s, in some order. The number of patterns is given by:

\[
\frac{5!}{2!3!} = 10
\]

In words, this formula gives us the number of different ways of arranging 2 R’s and 3 G’s in a row. The 5 in the numerator of the formula is the sum of 2 and 3 in the denominator.
Permutations with repetitions

For example, the number of ways to arrange four R’s and one G in a row is:

$$\frac{5!}{4!1!} = 5$$

The patterns are:
R R R R G, R R R G R, R R G R R, R G R R R, G R R R R

R R R R R
Permutations with repetitions

The general formula follows:

**Theorem**

The number of permutations of \( n \) objects of which \( n_1 \) are alike, \( n_2 \) are alike, ..., \( n_r \) are alike is

\[
\frac{n!}{n_1!n_2! \cdots n_r!}
\]
Ordered Samples

Our last example is representative of many other problems in combinatorial analysis.

- In particular, a lot of problems in probability are concerned with choosing a ball from an urn containing \( n \) balls (or a card from a deck, or a person from a population).
- When we choose one ball after another from the urn, say \( r \) times, we call the choice an ordered sample of size \( r \).
Ordered samples

We consider two cases:

(i) *Sampling with replacement.* Here the ball is replaced in the urn before the next ball is chosen. Now since there are \( n \) different ways to choose each ball, there are by the fundamental principle of counting

\[
\begin{align*}
\underbrace{n \times n \times n \cdots n}^{	ext{r times}} &= n^r
\end{align*}
\]

different ordered samples with replacement of size \( r \).
Ordered samples

(ii) *Sampling without replacement.* Here the ball is not replaced in the urn before the next ball is chosen. In other words, an ordered sample of size \( r \) without replacement is simply an \( r \)-permutation of the objects in the urn. Thus there are

\[
P(n, r) = n(n - 1)(n - 2) \cdots (n - r + 1) = \frac{n!}{(n - r)!}
\]

different ordered samples of size \( r \) without replacement from a population of \( n \) objects.
The Binomial Coefficient and Formula

Let’s go back to the example of the urn containing one red ball and nine green ones. Five draws were made at random with replacement.

We learned that there are 10 different patterns in which 2 draws are red:

\[
\frac{5!}{2!3!} = 10
\]

An expression of this form is called a binomial coefficient.
The Binomial Coefficient and Formula

We will usually write the binomial coefficient as follows:

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

The symbol \( \binom{n}{r} \) reads as “\( n \) choose \( r \),” the idea being that the formula gives the number of ways to choose \( r \) things out of \( n \).

Some books write the binomial coefficient as “\( nC_r \)” or \( C^n_r \), the “number of combinations of \( n \) things taken \( r \) at a time.”
The Binomial Coefficient and Formula

We are finally ready to answer the following question.

- Given our example of the red and green balls: What is the probability that exactly two draws will be red?
- We know that the chance of red is $\frac{1}{10}$, and the chance of green is $\frac{9}{10}$.
- We also know that there are 10 different patterns with 2 R's and 3 G's.
- So the probability of drawing exactly 2 reds is:

$$10 \times \left( \frac{1}{10} \right)^2 \left( \frac{9}{10} \right)^3 \approx 7\%$$
The Binomial Coefficient and Formula

The reasoning of this example is summarized in the *binomial formula*.

- Suppose a chance process is carried out as a sequence of trials; an example would be rolling a die 10 times, where each roll counts as a trial.
- There is an event of interest which may or may not occur at each trial; the die may or may not land ace.
- The problem is to calculate the probability that the event will occur a specified number of times.
The Binomial Coefficient and Formula

**The Binomial Formula.** The chance that an event will occur exactly \( r \) times out of \( n \) is given by the binomial formula

\[
\frac{n!}{r!(n-r)!} p^r (1-p)^{n-r}.
\]

In this formula, \( n \) is the number of trials, \( r \) is the number of times the event is to occur, and \( p \) is the probability that the event will occur on any particular trial. The assumptions:

- The value of \( n \) must be fixed in advance.
- \( p \) must be the same from trial to trial.
- The trials must be independent.
The Binomial Coefficient and Formula

Example
A die is rolled 10 times. What is the chance of getting exactly 2 aces?

The number of trials is fixed in advance. It is 10. So \( n = 10 \). The event of interest is rolling an ace. The probability of rolling an ace is the same from trial to trial. It is \( \frac{1}{6} \). So, \( p = \frac{1}{6} \). The trials are independent.

The binomial formula can be used, and the answer is:

\[
\frac{10!}{2!8!} \left( \frac{1}{6} \right)^2 \left( \frac{5}{6} \right)^8 \approx 29\%.
\]
Overview of Today’s Class

1. Probability
   • Some Background

2. Probability in Simple Sample Spaces
   • Sample Space and Events
   • Probability of an Event
   • Counting Techniques

3. Conditional Probability
   • Conditional and Compound Probabilities
   • Bayes’ Theorem
   • Independent Events
Sometimes when we observe two outcomes we may be interested in how the outcome of one event $E$ is influenced by that of another event $F$.

For example, in one extreme case the relation between $E$ and $F$ may be such that $E$ always occur if $F$ does, while in the other case, $E$ never occurs if $F$ does.

To characterize the relation between $E$ and $F$, I will introduce the conditional probability of $E$ on the hypothesis $F$, i.e., the “probability of $E$ occurring under the condition that $F$ is known to have occurred.”
Conditional Probability: Example

It will be helpful first to take a look at an example in which we can see how this works on intuitive grounds.

**Example**

A deck of cards is shuffled and the top two cards are put on a table, face down. You win $1 if the second card is the queen of hearts.

(a) What is your chance of winning the dollar?

(b) You turn over the first card. It is the seven of clubs. Now what is your chance of winning?
Conditional Probability: Example

Solution. Part (a). The bet is about the second card, not the first. For example,

- If the first card is the two of spades and the second is the queen of hearts, you win.
- If the first card is the seven of clubs and the second is the king of hearts, you lose.

The bet can be settled without even looking at the first card: the second card is all you need to know. The chance of winning is $\frac{1}{52}$. A well-shuffled deck brings the cards into random order. The queen of hearts has to wind up somewhere. There are 52 possible positions, and they are all equally likely.
Conditional Probability: Example

Part (b). There are 51 cards left. They are in random order, and the queen of hearts is one of them. She has 1 chance in 51 to be on the table. Your chance goes up a little, to \( \frac{1}{51} \). That is the answer.

- The \( \frac{1}{51} \) in part (b) is called a *conditional* chance. The problem puts a condition on the first card: it has to be the seven of clubs.
- We will refer to this as the conditional probability that the second card is the queen of hearts *given* the first card is the seven of clubs.
- To emphasize the contrast, the \( \frac{1}{52} \) in part (a) is called an *unconditional* chance: the problem puts no conditions on the first card.
Conditional Probability: Definition

This informal and intuitive reasoning can be described in another way. In general, given a sample space $\Omega$ and an acceptable assignment of probabilities to the sample points in $\Omega$, we compute the probability of an event $E$ by adding the probabilities of the sample points whose union is $E$.

- Since $P(\Omega) = 1$ and $E \cap \Omega = E$, we can write the identity

$$P(E) = \frac{P(E \cap \Omega)}{P(\Omega)},$$

which shows that $P(E)$ is the ratio of the probability of that part of $E$ included in $\Omega$ (which happens to be all of $E$) to the probability of $\Omega$ itself (which happens to be 1).
Conditional Probability: Definition

But if we are told that event $F$ has occurred, then the outcomes corresponding to elements of $F'$, the complement of $F$, are no longer possible.

Therefore, in light of our added information, the event $F$ replaces the sample space $\Omega$ as the set whose elements correspond to all possible outcomes.
Conditional Probability: Definition

With this in mind, observe how reasonable it appears to write

\[ P(E|F) = \frac{P(E \cap F)}{P(F)}, \]

which says that \( P(E|F) \), the conditional probability of \( E \) given \( F \) (the vertical bar is read “given”), is the ratio of the probability of that part of \( E \) included in \( F \) (which is \( E \cap F \)) to the probability of \( F \) itself.
Conditional Probability: Definition

Let $F$ be an arbitrary event in a sample space $\Omega$ with $P(F) > 0$. The probability that an event $E$ occurs once $F$ has occurred or, in other words, the conditional probability of $E$ given $F$, written $P(E|F)$, is defined as follows:

$$P(E|F) = \frac{P(E \cap F)}{P(F)}.$$
Conditional Probability: Definition

If $\Omega$ is a finite equiprobable space and $|E|$ denotes the number of elements in an event $E$, then

$$P(E \cap F) = \frac{|E \cap F|}{|\Omega|}, \quad P(F) = \frac{|F|}{|\Omega|} \text{ and so}$$

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{|E \cap F|}{|F|}.$$
Conditional Probability: Example

Example

Three fair coins are tossed, one after the other. Let $E$ be the event “at least two heads” and $F$ be the event “first coin falls heads.”

- The sample space for this experiment is $\Omega = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
- Thus, the event $E$ is the union of four sample points. The event $F$ is also the union of four sample points. And, $E \cap F$ is the union of three sample points.
Conditional Probability: Example

Therefore, the conditional probability of $E$ given $F$ is

$$P(E|F) = \frac{P(E \cap F)}{P(F)} = \frac{\frac{3}{8}}{\frac{4}{8}} = \frac{3}{4}.$$ 

Note that the added knowledge that the first coin falls heads increases the probability of getting at least two heads. Before this additional information is revealed, $P(E) = \frac{1}{2}$. Afterwards, $P(E|F) = \frac{3}{4}$. 
Conditional Probability

**Theorem**

Let $\Omega$ be a finite equiprobable space with events $E$ and $F$. Then

$$P(E|F) = \frac{\text{number of elements in } E \cap F}{\text{number of elements in } F}.$$ 

or

$$P(E|F) = \frac{\text{number of ways } E \text{ and } F \text{ can occur}}{\text{number of ways } F \text{ can occur}}.$$
Conditional Probability: Example

Example
Let a pair of dice be tossed. If the sum is 6, find the probability that one of the dice is a 2.

- We can restate this problem as $F = \{\text{sum is 6}\} = \{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$. And, $E = \{\text{a 2 appears on at least one die}\}$, find $P(E|F)$.
- Now $F$ consists of five elements and two of them, $(2, 4)$ and $(4, 2)$ belong to $E$: $E \cap F = \{(2, 4), (4, 2)\}$. Then $P(E|F) = \frac{2}{5} \approx 0.4$.
- As before, the added knowledge that the sum of the dice is 6 increases the probability that one of the dice is a 2. Before this additional information is revealed, $P(E) = \frac{11}{36} \approx 0.3$. 
Compound Probabilities

Given two events $E$ and $F$, we already learned how to calculate $P(E|F)$.

- Notice now that if $P(E) > 0$, then the roles of $E$ and $F$ can be interchanged.
- The conditional probability of $F$ given $E$ is

$$P(F|E) = \frac{P(F \cap E)}{P(E)} = \frac{P(E \cap F)}{P(E)}$$

the last equality following from the commutative law for the intersection of two sets.
Compound Probabilities

If we cross-multiply the above equation defining conditional probability, we obtain the following useful formula:

\[ P(E \cap F) = P(E)P(F|E) = P(F)P(E|F). \]

In words, the chance that two things will both happen equals the chance that the first will happen, multiplied by the chance that the second will happen given that the first has happened.
Compound Probabilities

The last result can be extended by induction as follows:

Theorem

If $n$ is any integer ($n \geq 2$) and $E_1, E_2, ..., E_n$ are any $n$ events for which $P(E_1 \cap E_2 \cap ... \cap E_{n-1}) \neq 0$, then $P(E_1 \cap E_2 \cap ... \cap E_n)$ can be written as

$$P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)...P(E_n|E_1 \cap E_2 ... \cap E_{n-1}).$$

This theorem is sometimes referred to as the theorem on compound probabilities.
Compound Probabilities: Examples

Example
Two cards will be dealt off the top of a well-shuffled deck. What is the chance that the first card will be the seven of clubs and the second card will be the queen of hearts?

This is very similar to the first example. The chance that the first card will be the seven of clubs is $\frac{1}{52}$. Given that the first card was the seven of clubs, the chance that the second card will be the queen of hearts is $\frac{1}{51}$. The chance of getting both cards is,

$$\frac{1}{52} \times \frac{1}{51} = \frac{1}{2652}.$$ 

- This is a very small chance: about 4 in 10,000, or 0.04 of 1%. 