Basic Probability Theory

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Basic Probability Theory

1. Introduction
2. Set Theory
   - Definition of a Set
   - Complement, Union, Intersection, and Difference
   - Set Partition and Mutually Exclusive Sets
3. Probabilistic Experiments and Events
   - Probabilistic Experiments and their Sample Spaces
   - Elementary and Compound Events
4. Axioms and Basic Theorems of Probability
5. Basic Rules for Computing Probability
   - The General Rule
6. Joint Events, Conditional Probability, and Independence
   - Joint and Marginal Events
   - Conditional Probability
7. Sequences and the Multiplicative Rules
   - Sequences as Intersections
   - The Multiplicative Rules
8. The Keep It Alive Strategy
9. Probability Distributions
   - Discrete Probability Distributions
   - Continuous Probability Distributions
   - Continuous Probability Distributions
   - Summing it up in Simple Terms
   - Calculating Probability with R

Introduction

- The Gravetter-Walnau treatment of probability is extremely basic and in my view, too incomplete.
- I'm going to develop probability theory "from the ground up," using set theory as the point of departure.
- Gravetter and Walnau use the notation $p(A)$ to stand for the probability of an event A, while I use the notation $\Pr(A)$.

Introduction

- The Gravetter-Walnau treatment of probability is extremely basic and in my view, too incomplete.
- I'm going to develop probability theory "from the ground up," using set theory as the point of departure.
- Gravetter and Walnau use the notation $p(A)$ to stand for the probability of an event A, while I use the notation $\Pr(A)$.

Introduction

- The Gravetter-Walnau treatment of probability is extremely basic and in my view, too incomplete.
- I'm going to develop probability theory "from the ground up," using set theory as the point of departure.
- Gravetter and Walnau use the notation $p(A)$ to stand for the probability of an event A, while I use the notation $\Pr(A)$.

Set Theory

- Some grounding in basic set theory is very useful to gaining a solid understanding of probability theory.
- Many books attempt to discuss probability theory while never formally defining set theory, and instead relying on an "informal" version of set theory. This saves time, but also reduces precision, depending on the sophistication of the reader.

Set Theory

- Some grounding in basic set theory is very useful to gaining a solid understanding of probability theory.
- Many books attempt to discuss probability theory while never formally defining set theory, and instead relying on an "informal" version of set theory. This saves time, but also reduces precision, depending on the sophistication of the reader.

Definition of a Set

- A set is any well-defined collection of "objects."
- The term "object" is used loosely. It could refer to numbers, ideas, people, or fruits.
- The *elements* of a set are the objects in the set.

Definition of a Set

- A set is any well-defined collection of "objects."
- The term "object" is used loosely. It could refer to numbers, ideas, people, or fruits.
- The *elements* of a set are the objects in the set.

Definition of a Set

- A set is any well-defined collection of "objects."
- The term "object" is used loosely. It could refer to numbers, ideas, people, or fruits.
- The *elements* of a set are the objects in the set.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Complement, Union, Intersection, and Difference

- The *universal set* Ω is the set of all objects currently under consideration.
- The *null (empty) set* $\emptyset$ is a set with no elements.
- The *complement* of A, i.e., $\bar{A}$, is the set of all elements not in A (but in Ω).
- Consider two sets A and B:
- The *union* of A and B, i.e., $A \cup B$, is the set of all elements in A, B, or both.
- The *intersection* of A and B, i.e., $A \cap B$, is the set of all elements in A and B. Informal discussions of probability will refer to $\Pr(A \text{ and } B)$ to signify $\Pr(A \cap B)$.
- The *set difference*, $A - B$, is defined as the set of elements in A but not in B. That is, $A - B = A \cap \overline{B}$.

Mutually Exclusive and Exhaustive Events

- Two sets A and B are said to be *mutually exclusive* if they have no elements in common, i.e., $A \cap B = \emptyset$
- A set of events is *exhaustive* of another set if their union is equal to that set.

Mutually Exclusive and Exhaustive Events

- Two sets A and B are said to be *mutually exclusive* if they have no elements in common, i.e., $A \cap B = \emptyset$
- A set of events is *exhaustive* of another set if their union is equal to that set.

Set Partition

- A group of n sets $E_i, i = 1, \ldots n$ is said to *partition* the set A if the E_i are mutually exclusive and exhaustive with respect to A.
- Note the following:
  1. Any set is the union of its elements.
  2. Any set is partitioned by its elements.

Set Partition

- A group of n sets $E_i, i = 1, ...n$ is said to *partition* the set A if the E_i are mutually exclusive and exhaustive with respect to A.
- Note the following:
  1. Any set is the union of its elements.
  2. Any set is partitioned by its elements.

Set Partition

- A group of n sets $E_i, i = 1, \ldots n$ is said to *partition* the set A if the E_i are mutually exclusive and exhaustive with respect to A.
- Note the following:
  1. Any set is the union of its elements.
  2. Any set is partitioned by its elements.

Set Partition

- A group of n sets $E_i, i = 1, ...n$ is said to *partition* the set A if the E_i are mutually exclusive and exhaustive with respect to A.
- Note the following:
  1. Any set is the union of its elements.
  2. Any set is partitioned by its elements.

Probabilistic Experiment

- A probabilistic experiment E is a situation in which
  1. Something can happen.
  2. What can happen is well-defined.
  3. The outcome is potentially uncertain.

Probabilistic Experiment

- A probabilistic experiment E is a situation in which
  1. Something can happen.
  2. What can happen is well-defined.
  3. The outcome is potentially uncertain.

Probabilistic Experiment

- A probabilistic experiment E is a situation in which
  1. Something can happen.
  2. What can happen is well-defined.
  3. The outcome is potentially uncertain.

Probabilistic Experiment

- A probabilistic experiment E is a situation in which
  1. Something can happen.
  2. What can happen is well-defined.
  3. The outcome is potentially uncertain.

Sample Space

- Associated with every probabilistic experiment E_i is a *sample space* S_i.
- The sample space is defined as the universal set of all possible outcomes of E_i.

Sample Space

- Associated with every probabilistic experiment E_i is a *sample space* S_i.
- The sample space is defined as the universal set of all possible outcomes of E_i.

Elementary Events

- The *elementary events* in a sample space are the elements of S.
- As such, they constitute the "finest" possible partition of S.

Example (Elementary Events)

You throw a fair die, and observe the number that comes up. The elementary events are 1,2,3,4,5,6.

Elementary Events

- The *elementary events* in a sample space are the elements of S.
- As such, they constitute the "finest" possible partition of S.

Example (Elementary Events)

You throw a fair die, and observe the number that comes up. The elementary events are 1,2,3,4,5,6.

Elementary Events

- The *elementary events* in a sample space are the elements of S.
- As such, they constitute the "finest" possible partition of S.

Example (Elementary Events)

You throw a fair die, and observe the number that comes up. The elementary events are 1,2,3,4,5,6.

Compound Events

- Compound events are the union of two or more elementary events.

Example (Compound Events)

In the die throw experiment, the event
Even $= 2 \cup 4 \cup 6 = \{2, 4, 6\}$ is a compound event.

Compound Events

- Compound events are the union of two or more elementary events.

Example (Compound Events)

In the die throw experiment, the event
Even $= 2 \cup 4 \cup 6 = \{2, 4, 6\}$ is a compound event.

Axioms of Probability Theory

- Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:
  1. $\Pr(A) \geq 0$
  2. $\Pr(S) = 1$
  3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- The above 3 *axioms* of discrete probability theory guarantee that probabilities act like relative frequencies.
- We can build on these axioms by immediately deriving 3 fundamental theorems.

Axioms of Probability Theory

- Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:
  1. $\Pr(A) \geq 0$
  2. $\Pr(S) = 1$
  3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- The above 3 *axioms* of discrete probability theory guarantee that probabilities act like relative frequencies.
- We can build on these axioms by immediately deriving 3 fundamental theorems.

Axioms of Probability Theory

- Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:
  1. $\Pr(A) \geq 0$
  2. $\Pr(S) = 1$
  3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- The above 3 *axioms* of discrete probability theory guarantee that probabilities act like relative frequencies.
- We can build on these axioms by immediately deriving 3 fundamental theorems.

Axioms of Probability Theory

- Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:
  1. $\Pr(A) \geq 0$
  2. $\Pr(S) = 1$
  3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- The above 3 *axioms* of discrete probability theory guarantee that probabilities act like relative frequencies.
- We can build on these axioms by immediately deriving 3 fundamental theorems.

Axioms of Probability Theory

- Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:
  1. $\Pr(A) \geq 0$
  2. $\Pr(S) = 1$
  3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- The above 3 *axioms* of discrete probability theory guarantee that probabilities act like relative frequencies.
- We can build on these axioms by immediately deriving 3 fundamental theorems.

Axioms of Probability Theory

- Given a sample space S and any event A_i within S, we assign to each event a number called the *probability* of A. The probabilities in S must satisfy the following:
    1. $\Pr(A) \geq 0$
    2. $\Pr(S) = 1$
    3. If two events A and B in S are mutually exclusive, then $\Pr(A \cup B) = \Pr(A) + \Pr(B)$.
- The above 3 *axioms* of discrete probability theory guarantee that probabilities act like relative frequencies.
- We can build on these axioms by immediately deriving 3 fundamental theorems.

Fundamental Theorems

- $\Pr(A) = 1 - \Pr(\bar{A})$

  Proof. For any set A, since S is the universal set in this context, $S = A \cup \bar{A}$, and $A \cap \bar{A} = \emptyset$. But, by the third axiom, this means that $\Pr(S) = \Pr(A) + \Pr(\bar{A})$. However, the second axiom states that $\Pr(S) = 1$, and so we get $1 = \Pr(A) + \Pr(\bar{A})$, and the result follows by subtraction.

- $\Pr(\emptyset) = 0$

  Proof. For any set A, $A = A \cup \emptyset$, and $A \cap \emptyset = \emptyset$. But, using the third axiom, this means that $\Pr(A) = \Pr(A) + \Pr(\emptyset)$. By subtraction, you get the result.

- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

  Proof. We will do this proof in class.

Fundamental Theorems

- $\Pr(A) = 1 - \Pr(\bar{A})$

  Proof. For any set A, since S is the universal set in this context, $S = A \cup \bar{A}$, and $A \cap \bar{A} = \emptyset$. But, by the third axiom, this means that $\Pr(S) = \Pr(A) + \Pr(\bar{A})$. However, the second axiom states that $\Pr(S) = 1$, and so we get $1 = \Pr(A) + \Pr(\bar{A})$, and the result follows by subtraction.

- $\Pr(\emptyset) = 0$

  Proof. For any set A, $A = A \cup \emptyset$, and $A \cap \emptyset = \emptyset$. But, using the third axiom, this means that $\Pr(A) = \Pr(A) + \Pr(\emptyset)$. By subtraction, you get the result.

- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

  Proof. We will do this proof in class.

Fundamental Theorems

- $\Pr(A) = 1 - \Pr(\bar{A})$

  Proof. For any set A, since S is the universal set in this context, $S = A \cup \bar{A}$, and $A \cap \bar{A} = \emptyset$. But, by the third axiom, this means that $\Pr(S) = \Pr(A) + \Pr(\bar{A})$. However, the second axiom states that $\Pr(S) = 1$, and so we get $1 = \Pr(A) + \Pr(\bar{A})$, and the result follows by subtraction.

- $\Pr(\emptyset) = 0$

  Proof. For any set A, $A = A \cup \emptyset$, and $A \cap \emptyset = \emptyset$. But, using the third axiom, this means that $\Pr(A) = \Pr(A) + \Pr(\emptyset)$. By subtraction, you get the result.

- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$

  Proof. We will do this proof in class.

Fundamental Theorems

Examples

Example (The Third Theorem.)

Jody is a student in Psychology 666. There are 2 exams in Psychology 666. The probability that *Jody passes Exam A, Exam B, or both* is .97. The probability that she passes Exam A is .90, the probability that she passes Exam B is .80. What is the probability that she passes *both Exam A and Exam B*?

Fundamental Theorems

Examples

Example (The Third Theorem.)

Jody is a student in Psychology 666. There are 2 exams in Psychology 666. The probability that *Jody passes exam Exam A, Exam B, or both* is .97. The probability that she passes Exam A is .90, the probability that she passes Exam B is .80. What is the probability that she passes *both Exam A and Exam B*?

Answer. We have to translate the English into probability theory. $\Pr(A \cup B) = 0.97$. $\Pr(A) = 0.90$, $\Pr(B) = 0.80$. We wish to find $\Pr(A \cap B)$. But the Third Theorem states that $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$, which leads to

$$
\begin{aligned}
\Pr(A \cup B) &= \Pr(A) + \Pr(B) - \Pr(A \cap B) \\
0.97 &= 0.90 + 0.80 - \Pr(A \cap B) \\
0.97 &= 1.70 - \Pr(A \cap B) \\
-0.73 &= -\Pr(A \cap B) \\
0.73 &= \Pr(A \cap B)
\end{aligned}
$$

The General Rule

- For any event A composed of elementary events e_i, i.e., $A = \cup_{i=1}^{n} e_i$, the probability of A is the sum of the probabilities of the elementary events in A, i.e., $\Pr(A) = \sum_{i=1}^{n} \Pr(e_i)$

  Comment. This follows immediately from the Third Axiom, because any event is partitioned by its elementary events, so the probability of the union must be the sum of the probabilities of the elemetary events.

Equally Likely Elementary Events

- Since the probabilities of the elementary events must sum to 1, if the elementary events are equally likely, then every one of the n_S elementary event has a probability of $1/n_S$.
- Consequently, the total probability of an event A can be computed as

$$\Pr(A) = \frac{n_A}{n_S} \tag{1}$$

  where n_A is the number of elementary events in A, and n_S is the number of elementary events in the sample space.
- This rule is essentially where Gravetter and Walnau begin their treatment of probability theory.

Equally Likely Elementary Events

- Since the probabilities of the elementary events must sum to 1, if the elementary events are equally likely, then every one of the n_S elementary event has a probability of $1/n_S$.
- Consequently, the total probability of an event A can be computed as

$$\Pr(A) = \frac{n_A}{n_S} \tag{1}$$

where n_A is the number of elementary events in A, and n_S is the number of elementary events in the sample space.
- This rule is essentially where Gravetter and Walnau begin their treatment of probability theory.

Equally Likely Elementary Events

- Since the probabilities of the elementary events must sum to 1, if the elementary events are equally likely, then every one of the n_S elementary event has a probability of $1/n_S$.
- Consequently, the total probability of an event A can be computed as
$$\Pr(A) = \frac{n_A}{n_S} \tag{1}$$
where n_A is the number of elementary events in A, and n_S is the number of elementary events in the sample space.
- This rule is essentially where Gravetter and Walnau begin their treatment of probability theory.

Equally Likely Elementary Events

Example (Marble Problem)

A bag of marbles has 4 red, 6 blue, 5 black, and 5 white marbles in it. A marble is *colored* if it is red or blue.

- What is the probability that a marble is colored or white?
- What is the probability that a marble is not white?
- What is the probability that a marble is colored or not white?

Equally Likely Elementary Events

Example (Marble Problem)

A bag of marbles has 4 red, 6 blue, 5 black, and 5 white marbles in it. A marble is *colored* if it is red or blue.

- What is the probability that a marble is colored or white?
- What is the probability that a marble is not white?
- What is the probability that a marble is colored or not white?

Equally Likely Elementary Events

Example (Marble Problem)

A bag of marbles has 4 red, 6 blue, 5 black, and 5 white marbles in it. A marble is *colored* if it is red or blue.

- What is the probability that a marble is colored or white?
- What is the probability that a marble is not white?
- What is the probability that a marble is colored or not white?

Equally Likely Elementary Events

Example (Marble Problem)

A bag of marbles has 4 red, 6 blue, 5 black, and 5 white marbles in it. A marble is *colored* if it is red or blue.

- What is the probability that a marble is colored or white?
- What is the probability that a marble is not white?
- What is the probability that a marble is colored or not white?

Joint Events and Marginal Events

- In many situations, two or more physical processes are occurring simultaneously, and we can define the elementary events of the sample space to be the intersection of the outcomes on the two processes.
- For example, I might throw a die and a coin simultaneously.
- In that case, we have 12 elementary events, each one of which is an intersection. They are $H \cap 1$, $H \cap 2 \ldots, T \cap 1, \ldots T \cap 6$.
- The *marginal* events are the outcomes on the individual processes, i.e., $H, T, 1, 2, 3, 4, 5, 6$.

Joint Events and Marginal Events

- In many situations, two or more physical processes are occurring simultaneously, and we can define the elementary events of the sample space to be the intersection of the outcomes on the two processes.
- For example, I might throw a die and a coin simultaneously.
- In that case, we have 12 elementary events, each one of which is an intersection. They are $H \cap 1$, $H \cap 2 \ldots, T \cap 1, \ldots T \cap 6$.
- The *marginal* events are the outcomes on the individual processes, i.e., $H, T, 1, 2, 3, 4, 5, 6$.

Joint Events and Marginal Events

- In many situations, two or more physical processes are occurring simultaneously, and we can define the elementary events of the sample space to be the intersection of the outcomes on the two processes.
- For example, I might throw a die and a coin simultaneously.
- In that case, we have 12 elementary events, each one of which is an intersection. They are $H \cap 1$, $H \cap 2 \ldots, T \cap 1, \ldots T \cap 6$.
- The *marginal* events are the outcomes on the individual processes, i.e., $H, T, 1, 2, 3, 4, 5, 6$.

Joint Events and Marginal Events

- In many situations, two or more physical processes are occurring simultaneously, and we can define the elementary events of the sample space to be the intersection of the outcomes on the two processes.
- For example, I might throw a die and a coin simultaneously.
- In that case, we have 12 elementary events, each one of which is an intersection. They are $H \cap 1$, $H \cap 2 \ldots, T \cap 1, \ldots T \cap 6$.
- The *marginal* events are the outcomes on the individual processes, i.e., $H, T, 1, 2, 3, 4, 5, 6$.

Joint Events and Marginal Events

- Here is a tabular representation.
- Note that the probability of a marginal event in a row or column is the sum of the probabilities of the joint events in that row or column. (Why?)

	Die					
Coin	1	2	3	4	5	6
H	$H \cap 1$	$H \cap 2$	$H \cap 3$	$H \cap 4$	$H \cap 5$	$H \cap 6$
T	$T \cap 1$	$T \cap 2$	$T \cap 3$	$T \cap 4$	$T \cap 5$	$T \cap 6$

Joint Events and Marginal Events

- Here is a tabular representation.
- Note that the probability of a marginal event in a row or column is the sum of the probabilities of the joint events in that row or column. (Why?)

	Die					
Coin	1	2	3	4	5	6
H	$H \cap 1$	$H \cap 2$	$H \cap 3$	$H \cap 4$	$H \cap 5$	$H \cap 6$
T	$T \cap 1$	$T \cap 2$	$T \cap 3$	$T \cap 4$	$T \cap 5$	$T \cap 6$

Conditional Probability

- The *conditional probability* of A given B, written as $\Pr(A|B)$, is the probability of A within the reduced sample space defined by B.
- To evaluate conditional probability, we simply move inside the class of events defined by B, and calculate what proportion of the events in B are examples of A.
- For example, suppose the probabilities of our joint events were as given in the table on the next slide.

Conditional Probability

- The *conditional probability* of A given B, written as $\Pr(A|B)$, is the probability of A within the reduced sample space defined by B.
- To evaluate conditional probability, we simply move inside the class of events defined by B, and calculate what proportion of the events in B are examples of A.
- For example, suppose the probabilities of our joint events were as given in the table on the next slide.

Conditional Probability

- The *conditional probability* of A given B, written as $\Pr(A|B)$, is the probability of A within the reduced sample space defined by B.
- To evaluate conditional probability, we simply move inside the class of events defined by B, and calculate what proportion of the events in B are examples of A.
- For example, suppose the probabilities of our joint events were as given in the table on the next slide.

Conditional Probability

		Die					
		1/6	1/6	1/6	1/6	1/6	1/6
	Coin	**1**	**2**	**3**	**4**	**5**	**6**
1/2	**H**	1/6	0	1/6	0	1/6	0
1/2	**T**	0	1/6	0	1/6	0	1/6

- What is $\Pr(1|H)$? $\Pr(1|T)$?
- What is $\Pr(H|1)$? $\Pr(H|2)$?
- Notice that the conditional probabilities within any conditionalizing event are simply the joint probabilities inside that event, re-standardized so that they add up to 1.
- They are re-standardized to add up to 1 by dividing by what they currently add up to, i.e., $\Pr(B)$.
- This leads to the formula for conditional probability often given as its definition:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{2}$$

Conditional Probability

		Die					
		1/6	1/6	1/6	1/6	1/6	1/6
	Coin	**1**	**2**	**3**	**4**	**5**	**6**
1/2	**H**	1/6	0	1/6	0	1/6	0
1/2	**T**	0	1/6	0	1/6	0	1/6

- What is $\Pr(1|H)$? $\Pr(1|T)$?
- What is $\Pr(H|1)$? $\Pr(H|2)$?
- Notice that the conditional probabilities within any conditionalizing event are simply the joint probabilities inside that event, re-standardized so that they add up to 1.
- They are re-standardized to add up to 1 by dividing by what they currently add up to, i.e., $\Pr(B)$.
- This leads to the formula for conditional probability often given as its definition:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \qquad (2)$$

Conditional Probability

		Die					
		1/6	1/6	1/6	1/6	1/6	1/6
	Coin	**1**	**2**	**3**	**4**	**5**	**6**
1/2	**H**	1/6	0	1/6	0	1/6	0
1/2	**T**	0	1/6	0	1/6	0	1/6

- What is $\Pr(1|H)$? $\Pr(1|T)$?
- What is $\Pr(H|1)$? $\Pr(H|2)$?
- Notice that the conditional probabilities within any conditionalizing event are simply the joint probabilities inside that event, re-standardized so that they add up to 1.
- They are re-standardized to add up to 1 by dividing by what they currently add up to, i.e., $\Pr(B)$.
- This leads to the formula for conditional probability often given as its definition:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{2}$$

Conditional Probability

		Die					
		1/6	1/6	1/6	1/6	1/6	1/6
	Coin	**1**	**2**	**3**	**4**	**5**	**6**
1/2	**H**	1/6	0	1/6	0	1/6	0
1/2	**T**	0	1/6	0	1/6	0	1/6

- What is $\Pr(1|H)$? $\Pr(1|T)$?
- What is $\Pr(H|1)$? $\Pr(H|2)$?
- Notice that the conditional probabilities within any conditionalizing event are simply the joint probabilities inside that event, re-standardized so that they add up to 1.
- They are re-standardized to add up to 1 by dividing by what they currently add up to, i.e., $\Pr(B)$.
- This leads to the formula for conditional probability often given as its definition:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{2}$$

Conditional Probability

		Die					
		1/6	1/6	1/6	1/6	1/6	1/6
	Coin	**1**	**2**	**3**	**4**	**5**	**6**
1/2	**H**	1/6	0	1/6	0	1/6	0
1/2	**T**	0	1/6	0	1/6	0	1/6

- What is $\Pr(1|H)$? $\Pr(1|T)$?
- What is $\Pr(H|1)$? $\Pr(H|2)$?
- Notice that the conditional probabilities within any conditionalizing event are simply the joint probabilities inside that event, re-standardized so that they add up to 1.
- They are re-standardized to add up to 1 by dividing by what they currently add up to, i.e., $\Pr(B)$.
- This leads to the formula for conditional probability often given as its definition:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)} \tag{2}$$

The Meaning of Conditional Probability

- Conditional probability reflects how the probabilities of the outcomes on one process change after we are informed about the status of the outcome of the other process.
- If knowledge of one process changes the probability structure for the other, we say that the processes are *dependent*.
- If knowledge of one process does not change the probability structure for the other, then we say that the processes are *independent*.
- This leads to two formally equivalent definitions of independence. Two events A and B are independent if

$$\Pr(A|B) = \Pr(A) \tag{3}$$

or

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{4}$$

The Meaning of Conditional Probability

- Conditional probability reflects how the probabilities of the outcomes on one process change after we are informed about the status of the outcome of the other process.
- If knowledge of one process changes the probability structure for the other, we say that the processes are *dependent*.
- If knowledge of one process does not change the probability structure for the other, then we say that the processes are *independent*.
- This leads to two formally equivalent definitions of independence. Two events A and B are independent if

$$\Pr(A|B) = \Pr(A) \tag{3}$$

or

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{4}$$

The Meaning of Conditional Probability

- Conditional probability reflects how the probabilities of the outcomes on one process change after we are informed about the status of the outcome of the other process.
- If knowledge of one process changes the probability structure for the other, we say that the processes are *dependent*.
- If knowledge of one process does not change the probability structure for the other, then we say that the processes are *independent*.
- This leads to two formally equivalent definitions of independence. Two events A and B are independent if

$$\Pr(A|B) = \Pr(A) \tag{3}$$

or

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{4}$$

The Meaning of Conditional Probability

- Conditional probability reflects how the probabilities of the outcomes on one process change after we are informed about the status of the outcome of the other process.
- If knowledge of one process changes the probability structure for the other, we say that the processes are *dependent*.
- If knowledge of one process does not change the probability structure for the other, then we say that the processes are *independent*.
- This leads to two formally equivalent definitions of independence. Two events A and B are independent if

$$\Pr(A|B) = \Pr(A) \tag{3}$$

or

$$\Pr(A \cap B) = \Pr(A)\Pr(B) \tag{4}$$

Sequences as Intersections

- Sequences of events are, of course, events themselves.
- Consider when we shuffle a poker deck and draw two cards off the top.
- The sequence of cards involves Card_1 followed by Card_2.
- If we now consider the event "Ace on the first card followed by Ace on the second card," we quickly realize that this is the intersection of two events, i.e. $A_1 \cap A_2$.
- How can we compute the probability of this event?

Sequences as Intersections

- Sequences of events are, of course, events themselves.
- Consider when we shuffle a poker deck and draw two cards off the top.
- The sequence of cards involves Card_1 followed by Card_2.
- If we now consider the event "Ace on the first card followed by Ace on the second card," we quickly realize that this is the intersection of two events, i.e. $A_1 \cap A_2$.
- How can we compute the probability of this event?

Sequences as Intersections

- Sequences of events are, of course, events themselves.
- Consider when we shuffle a poker deck and draw two cards off the top.
- The sequence of cards involves Card_1 followed by Card_2.
- If we now consider the event "Ace on the first card followed by Ace on the second card," we quickly realize that this is the intersection of two events, i.e. $A_1 \cap A_2$.
- How can we compute the probability of this event?

Sequences as Intersections

- Sequences of events are, of course, events themselves.
- Consider when we shuffle a poker deck and draw two cards off the top.
- The sequence of cards involves Card_1 followed by Card_2.
- If we now consider the event "Ace on the first card followed by Ace on the second card," we quickly realize that this is the intersection of two events, i.e. $A_1 \cap A_2$.
- How can we compute the probability of this event?

Sequences as Intersections

- Sequences of events are, of course, events themselves.
- Consider when we shuffle a poker deck and draw two cards off the top.
- The sequence of cards involves Card_1 followed by Card_2.
- If we now consider the event "Ace on the first card followed by Ace on the second card," we quickly realize that this is the intersection of two events, i.e. $A_1 \cap A_2$.
- How can we compute the probability of this event?

The Multiplicative Rules

- Recall that

$$\Pr(B|A) = \Pr(B \cap A)/\Pr(A) = \Pr(A \cap B)/\Pr(A)$$

- This implies that

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B|A) \tag{5}$$

- If A and B are independent, the rule simplifies to

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

.

The Multiplicative Rules

- Recall that

$$\Pr(B|A) = \Pr(B \cap A)/\Pr(A) = \Pr(A \cap B)/\Pr(A)$$

- This implies that

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B|A) \tag{5}$$

- If A and B are independent, the rule simplifies to

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

The Multiplicative Rules

- Recall that

$$\Pr(B|A) = \Pr(B \cap A)/\Pr(A) = \Pr(A \cap B)/\Pr(A)$$

- This implies that

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B|A) \tag{5}$$

- If A and B are independent, the rule simplifies to

$$\Pr(A \cap B) = \Pr(A) \times \Pr(B)$$

.

The Multiplicative Rules

- These rules are sometimes referred to as the *multiplicative rules of probability*. They generalize to sequences of any length. For example, suppose you draw 3 cards from a poker deck without replacement. Since the outcome on each card affects the probability structure of the succeeding draws, the events are not independent, and the probability of drawing 3 aces can be written using the general multiplicative rule as:

$$\Pr(A_1 \cap A_2 \cap A_3) = \Pr(A_1) \times \Pr(A_2 | A_1) \times \Pr(A_3 | A_1 \cap A_2) \quad (6)$$

The Multiplicative Rules

- The general multiplicative rule for a sequence can be stated as follows. *The probability of a sequence is the product of the probabilities of each event in the sequence conditionalized on everything that has gone before.*

The Multiplicative Rules

Example (The Multiplicative Rule)

Suppose you shuffle a poker deck and draw two cards completely randomly. What is the probability that they are both aces?

Answer. We are trying to solve for $\Pr(A_1 \cap A_2)$. By the general rule of multiplicative probability, this is $\Pr(A_1 \cap A_2) = \Pr(A_1) \times \Pr(A_2|A_1)$. When you go to draw the first card from the deck, $\Pr(A_1) = 4/52$, because there are 4 aces in the deck of 52 cards. If the first card is an ace, there are only 3 aces remaining out of 51 cards, so $Pr(A_2|A_1) = 3/51$. So $\Pr(A_1 \cap A_2) = 4/52 \times 3/51 = 1/13 \times 1/17 = 1/221$.

The Keep It Alive Strategy

A General Approach to Sequences

- The general multiplicative rule is much more powerful than it appears at first glance.
- Many very difficult problems can be solved with this rule.
- The key is one special insight.

The Keep It Alive Strategy

A General Approach to Sequences

- The general multiplicative rule is much more powerful than it appears at first glance.
- Many very difficult problems can be solved with this rule.
- The key is one special insight.

The Keep It Alive Strategy

A General Approach to Sequences

- The general multiplicative rule is much more powerful than it appears at first glance.
- Many very difficult problems can be solved with this rule.
- The key is one special insight.

The Keep It Alive Strategy

The Special Insight

- For an event to occur during a sequence, it has to *stay alive* during the sequence.
- So the probability of an event occurring during a sequence is the probability that it *stays alive* during the sequence.
- Let's try a few examples.

The Keep It Alive Strategy
The Special Insight

- For an event to occur during a sequence, it has to *stay alive* during the sequence.
- So the probability of an event occurring during a sequence is the probability that it *stays alive* during the sequence.
- Let's try a few examples.

The Keep It Alive Strategy

The Special Insight

- For an event to occur during a sequence, it has to *stay alive* during the sequence.
- So the probability of an event occurring during a sequence is the probability that it *stays alive* during the sequence.
- Let's try a few examples.

The Keep It Alive Strategy

A General Approach to Sequences

- A bag has 10 marbles, 5 of which are black, and 5 of which are white.
- If you select 2 marbles at random, what is the probability that they are both black?
- *Answer.* $\Pr(B_1 \cap B_2) = \Pr(B_1) \times \Pr(B_2|B_1) = \frac{5}{10} \times \frac{4}{9} = \frac{2}{9}$

The Keep It Alive Strategy

A General Approach to Sequences

- A bag has 10 marbles, 5 of which are black, and 5 of which are white.
- If you select 2 marbles at random, what is the probability that they are both black?
- *Answer.* $\Pr(B_1 \cap B_2) = \Pr(B_1) \times \Pr(B_2|B_1) = \frac{5}{10} \times \frac{4}{9} = \frac{2}{9}$

The Keep It Alive Strategy

A General Approach to Sequences

- A bag has 10 marbles, 5 of which are black, and 5 of which are white.
- If you select 2 marbles at random, what is the probability that they are both black?
- *Answer.* $\Pr(B_1 \cap B_2) = \Pr(B_1) \times \Pr(B_2 | B_1) = \frac{5}{10} \times \frac{4}{9} = \frac{2}{9}$

The Keep It Alive Strategy
Probability of a Flush

- Suppose you deal a poker hand of 5 cards at random from a deck of 52. A *flush* occurs if all 5 cards are the same suit (either 5 spades, 5 hearts, 5 diamonds, or 5 clubs).
- What is the probability of getting a flush (including straight flushes)?

The Keep It Alive Strategy
Probability of a Flush

- Suppose you deal a poker hand of 5 cards at random from a deck of 52. A *flush* occurs if all 5 cards are the same suit (either 5 spades, 5 hearts, 5 diamonds, or 5 clubs).
- What is the probability of getting a flush (including straight flushes)?

The Keep It Alive Strategy
Probability of a Flush

- *Answer.* This solution epitomizes the *keep it alive* strategy. The probability of a flush is the product of 5 probabilities, each of which is the conditional probability of staying alive for a flush on that draw.
- What is the probability that the first card you draw leaves you alive for a flush? (C.P.)
- You now have your first card. What is the probabiltiy that the next card leaves you alive for a flush, given everything that has gone before? (C.P.)
- Now you have two cards of the same suit. What is the probability that the third card leaves you alive, given what has gone before?
- Continuing this strategy, we arrive at the solution shown on the next slide.

The Keep It Alive Strategy
Probability of a Flush

- *Answer.* This solution epitomizes the *keep it alive* strategy. The probability of a flush is the product of 5 probabilities, each of which is the conditional probability of staying alive for a flush on that draw.
- What is the probability that the first card you draw leaves you alive for a flush? (C.P.)
- You now have your first card. What is the probabiltiy that the next card leaves you alive for a flush, given everything that has gone before? (C.P.)
- Now you have two cards of the same suit. What is the probability that the third card leaves you alive, given what has gone before?
- Continuing this strategy, we arrive at the solution shown on the next slide.

The Keep It Alive Strategy
Probability of a Flush

- *Answer.* This solution epitomizes the *keep it alive* strategy. The probability of a flush is the product of 5 probabilities, each of which is the conditional probability of staying alive for a flush on that draw.
- What is the probability that the first card you draw leaves you alive for a flush? (C.P.)
- You now have your first card. What is the probabiltiy that the next card leaves you alive for a flush, given everything that has gone before? (C.P.)
- Now you have two cards of the same suit. What is the probability that the third card leaves you alive, given what has gone before?
- Continuing this strategy, we arrive at the solution shown on the next slide.

The Keep It Alive Strategy

Probability of a Flush

- *Answer.* This solution epitomizes the *keep it alive* strategy. The probability of a flush is the product of 5 probabilities, each of which is the conditional probability of staying alive for a flush on that draw.
- What is the probability that the first card you draw leaves you alive for a flush? (C.P.)
- You now have your first card. What is the probabiltiy that the next card leaves you alive for a flush, given everything that has gone before? (C.P.)
- Now you have two cards of the same suit. What is the probability that the third card leaves you alive, given what has gone before?
- Continuing this strategy, we arrive at the solution shown on the next slide.

The Keep It Alive Strategy

Probability of a Flush

- *Answer.* This solution epitomizes the *keep it alive* strategy. The probability of a flush is the product of 5 probabilities, each of which is the conditional probability of staying alive for a flush on that draw.
- What is the probability that the first card you draw leaves you alive for a flush? (C.P.)
- You now have your first card. What is the probabiltiy that the next card leaves you alive for a flush, given everything that has gone before? (C.P.)
- Now you have two cards of the same suit. What is the probability that the third card leaves you alive, given what has gone before?
- Continuing this strategy, we arrive at the solution shown on the next slide.

The Keep It Alive Strategy

Probability of a Flush

$$
\begin{aligned}
\Pr(\mathit{Flush}) &= 1 \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48} \\
&= \frac{1}{51} \times \frac{11}{5} \times \frac{1}{49} \times \frac{9}{4} \\
&= \frac{1}{17} \times \frac{11}{5} \times \frac{1}{49} \times \frac{3}{4} \\
&= \frac{11 \times 3}{17 \times 5 \times 49 \times 4} \\
&= \frac{33}{16660} \\
&= 0.001981
\end{aligned}
$$

Introduction

- There are two fundamental types of random variables, *discrete* and *continuous*
- With discrete random variables, the number of events is countable.
- With continous random variables, the number of events is not countable and infinite, and continuous over at least some range on the number line.

Introduction

- There are two fundamental types of random variables, *discrete* and *continuous*
- With discrete random variables, the number of events is countable.
- With continous random variables, the number of events is not countable and infinite, and continuous over at least some range on the number line.

Introduction

- There are two fundamental types of random variables, *discrete* and *continuous*
- With discrete random variables, the number of events is countable.
- With continous random variables, the number of events is not countable and infinite, and continuous over at least some range on the number line.

Discrete Probability Distributions

- When the number of events is countable and finite, the probability of each elementary event can be calculated, and the sum of these probabilities is 1.
- In this case, each event has a probability, and the probability density function, or pdf, is usually denoted $p(x)$. This is the probability that the random variable X takes on the value x, i.e., $p(x) = \Pr(X = x)$.
- The *cumulative probability function*, $F(x)$, is the cumulative probability up to and including x, i.e., $F(x) = \Pr(X \leq x)$.

Discrete Probability Distributions

- When the number of events is countable and finite, the probability of each elementary event can be calculated, and the sum of these probabilities is 1.
- In this case, each event has a probability, and the probability density function, or pdf, is usually denoted $p(x)$. This is the probability that the random variable X takes on the value x, i.e., $p(x) = \Pr(X = x)$.
- The *cumulative probability function*, $F(x)$, is the cumulative probability up to and including x, i.e., $F(x) = \Pr(X \leq x)$.

Discrete Probability Distributions

- When the number of events is countable and finite, the probability of each elementary event can be calculated, and the sum of these probabilities is 1.
- In this case, each event has a probability, and the probability density function, or pdf, is usually denoted $p(x)$. This is the probability that the random variable X takes on the value x, i.e., $p(x) = \Pr(X = x)$.
- The *cumulative probability function*, $F(x)$, is the cumulative probability up to and including x, i.e., $F(x) = \Pr(X \leq x)$.

Continuous Probability Distributions

- When a random variable X is continuous, it takes on all values over some continuous range, and so the number of outcomes is uncountably infinite.
- For continuous random variables, the probability of a particular outcome x cannot be defined, even if x can occur. The probability is infinitesimally small.
- Rather than defining the probability of x, we instead define an alternative concept, *probability density*.

Continuous Probability Distributions

- When a random variable X is continuous, it takes on all values over some continuous range, and so the number of outcomes is uncountably infinite.
- For continuous random variables, the probability of a particular outcome x cannot be defined, even if x can occur. The probability is infinitesimally small.
- Rather than defining the probability of x, we instead define an alternative concept, *probability density*.

Continuous Probability Distributions

- When a random variable X is continuous, it takes on all values over some continuous range, and so the number of outcomes is uncountably infinite.
- For continuous random variables, the probability of a particular outcome x cannot be defined, even if x can occur. The probability is infinitesimally small.
- Rather than defining the probability of x, we instead define an alternative concept, *probability density*.

Continuous Probability Distributions

- Probability density $f(x)$ is not probability. Rather, it is the instantaneous rate at which the cumulative probability F is increasing at x. That is, it is the slope (derivative) of $F(x)$ at x.
- In a similar fashion, $F(x)$ is the area under the probability density curve. It is the integral of $f(x)$.
- Although the probability of a single value x cannot be defined with continuous random variables, the probability of an interval can be. To compute the probability of an interval, we simply take the difference of the two cumulative probabilities. That is,

$$\Pr(a \leq X \leq b) = F(b) - F(a)$$

Continuous Probability Distributions

- Probability density $f(x)$ is not probability. Rather, it is the instantaneous rate at which the cumulative probability F is increasing at x. That is, it is the slope (derivative) of $F(x)$ at x.
- In a similar fashion, $F(x)$ is the area under the probability density curve. It is the integral of $f(x)$.
- Although the probability of a single value x cannot be defined with continuous random variables, the probability of an interval can be. To compute the probability of an interval, we simply take the difference of the two cumulative probabilities. That is,

$$\Pr(a \leq X \leq b) = F(b) - F(a)$$

Continuous Probability Distributions

- Probability density $f(x)$ is not probability. Rather, it is the instantaneous rate at which the cumulative probability F is increasing at x. That is, it is the slope (derivative) of $F(x)$ at x.
- In a similar fashion, $F(x)$ is the area under the probability density curve. It is the integral of $f(x)$.
- Although the probability of a single value x cannot be defined with continuous random variables, the probability of an interval can be. To compute the probability of an interval, we simply take the difference of the two cumulative probabilities. That is,

$$\Pr(a \leq X \leq b) = F(b) - F(a)$$

Summing it Up

- Although the preceding sections may have seemed very abstract, the rules for computing probability from probability distributions are really very simple, *if you have a distribution calculator.*

Summing it Up
The Rules

- For *discrete distributions*:
  1. You can compute the probability that $X = x$, i.e., that a discrete random variable takes on precisely the value x.
  2. The probability that X falls in some interval is obtained by summing up the probabilities for all the discrete values in that interval. Note that *it can make a big difference whether the interval is open or closed*, that is, whether you are asked to compute $\Pr(a \leq X \leq b)$ or $\Pr(a < X < b)$
  3. Because only discrete values occur, $\Pr(X > a) = 1 - \Pr(X \leq a)$.

Summing it Up
The Rules

- For *discrete distributions*:
  1. You can compute the probability that $X = x$, i.e., that a discrete random variable takes on precisely the value x.
  2. The probability that X falls in some interval is obtained by summing up the probabilities for all the discrete values in that interval. Note that *it can make a big difference whether the interval is open or closed*, that is, whether you are asked to compute $\Pr(a \leq X \leq b)$ or $\Pr(a < X < b)$
  3. Because only discrete values occur, $\Pr(X > a) = 1 - \Pr(X \leq a)$.

Summing it Up

The Rules

- For *discrete distributions*:
  1. You can compute the probability that $X = x$, i.e., that a discrete random variable takes on precisely the value x.
  2. The probability that X falls in some interval is obtained by summing up the probabilities for all the discrete values in that interval. Note that *it can make a big difference whether the interval is open or closed*, that is, whether you are asked to compute $\Pr(a \leq X \leq b)$ or $\Pr(a < X < b)$
  3. Because only discrete values occur, $\Pr(X > a) = 1 - \Pr(X \leq a)$.

Summing it Up
The Rules

- For *discrete distributions*:
  1. You can compute the probability that $X = x$, i.e., that a discrete random variable takes on precisely the value x.
  2. The probability that X falls in some interval is obtained by summing up the probabilities for all the discrete values in that interval. Note that *it can make a big difference whether the interval is open or closed*, that is, whether you are asked to compute $\Pr(a \leq X \leq b)$ or $\Pr(a < X < b)$
  3. Because only discrete values occur, $\Pr(X > a) = 1 - \Pr(X \leq a)$.

Summing it Up

The Rules

- For *continuous distributions*:
  1. You *cannot compute the probability that X takes on a specific value.*
  2. You can compute the probability that X falls in an interval. Because the distribution is continuous, there is no practical distinction between $\Pr(a < X < b)$ and $\Pr(a \leq X \leq b)$.
  3. The probability for an interval from a to b is the area under the probability density curve between a and b, i.e., assuming $b > a$, $F(b) - F(a)$.
  4. Since the total area under the curve is 1, if $\Pr(X \leq a) = F(a)$, then $\Pr(X > a) = 1 - F(a)$

Summing it Up
The Rules

- For *continuous distributions*:
  1. You *cannot compute the probability that X takes on a specific value.*
  2. You can compute the probability that X falls in an interval. Because the distribution is continuous, there is no practical distinction between $\Pr(a < X < b)$ and $\Pr(a \leq X \leq b)$.
  3. The probability for an interval from a to b is the area under the probability density curve between a and b, i.e., assuming $b > a$, $F(b) - F(a)$.
  4. Since the total area under the curve is 1, if $\Pr(X \leq a) = F(a)$, then $\Pr(X > a) = 1 - F(a)$

Summing it Up
The Rules

- For *continuous distributions*:
  1. You *cannot compute the probability that X takes on a specific value.*
  2. You can compute the probability that X falls in an interval. Because the distribution is continuous, there is no practical distinction between $\Pr(a < X < b)$ and $\Pr(a \leq X \leq b)$.
  3. The probability for an interval from a to b is the area under the probability density curve between a and b, i.e., assuming $b > a$, $F(b) - F(a)$.
  4. Since the total area under the curve is 1, if $\Pr(X \leq a) = F(a)$, then $\Pr(X > a) = 1 - F(a)$

Summing it Up
The Rules

- For *continuous distributions*:
  1. You *cannot compute the probability that X takes on a specific value.*
  2. You can compute the probability that X falls in an interval. Because the distribution is continuous, there is no practical distinction between $\Pr(a < X < b)$ and $\Pr(a \leq X \leq b)$.
  3. The probability for an interval from a to b is the area under the probability density curve between a and b, i.e., assuming $b > a$, $F(b) - F(a)$.
  4. Since the total area under the curve is 1, if $\Pr(X \leq a) = F(a)$, then $\Pr(X > a) = 1 - F(a)$

Summing it Up
The Rules

- For *continuous distributions*:
  1. You *cannot compute the probability that X takes on a specific value.*
  2. You can compute the probability that X falls in an interval. Because the distribution is continuous, there is no practical distinction between $\Pr(a < X < b)$ and $\Pr(a \leq X \leq b)$.
  3. The probability for an interval from a to b is the area under the probability density curve between a and b, i.e., assuming $b > a$, $F(b) - F(a)$.
  4. Since the total area under the curve is 1, if $\Pr(X \leq a) = F(a)$, then $\Pr(X > a) = 1 - F(a)$

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  - *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  - *p*, for calculating cumulative probability.
  - *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  - *norm*. The normal distribution.
  - *t*. Student's t distribution.
  - *f*. The F distribution.
  - *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's *t* distribution.
  3. *f*. The *F* distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
    1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
    2. *p*, for calculating cumulative probability.
    3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
    1. *norm*. The normal distribution.
    2. *t*. Student's t distribution.
    3. *f*. The F distribution.
    4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. *d*, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. *p*, for calculating cumulative probability.
  3. *q*, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- R provides a set of functions for performing probability calculations for many of the most important distributions.
- The key function types are
  1. d, for calculating probability (discrete r.v.) or probability density (continuous r.v.)
  2. p, for calculating cumulative probability.
  3. q, for calculating inverse cumulative probabilities, or quantiles.
- To perform these calculations, you need to know the code for the distribution you are working with. Some common codes are:
  1. *norm*. The normal distribution.
  2. *t*. Student's t distribution.
  3. *f*. The F distribution.
  4. *chisq*. The chi-square distribution.
- You combine the function type and the distribution name to obtain the name of the function you need to perform a particular calculation.

Calculating Probability with R

- Each distribution function has certain rules you need to follow when using it.
- At first these may seem difficult, but before long they will seem easy.
- For a normal distribution, if the mean is 0 and the standard deviation 1 (i.e., the distribution is in Z score form), you only need to supply one value.
- `qnorm(p)` gives the value at a particular quantile

Calculating Probability with R

- What value of the standard normal distribution is at the 95th percentile?

```
> qnorm(.95)
[1] 1.644854
```

- What proportion of the standard normal distribution is between -1 and $+1$.

```
> pnorm(1) - pnorm(-1)
[1] 0.6826895
```

Calculating Probability with R

- To work with any other distribution besides the standard normal, you have to provide the metric (mean and standard deviation).
- What is the probability that an observation from a normal distribution with mean 500 and standard deviation 100 will be between 600 and 700?

```
> pnorm(700,500,100)-pnorm(600,500,100)
[1] 0.1359051
```

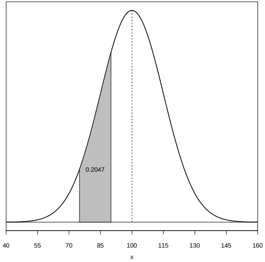
Calculating Probability with R

- I created some routines for you that can draw normal curves and fill in the area, so you can visualize the probability.
- Here's a question. IQ scores have a mean of 100 and a standard deviation of 15 in the general population. What proportion of the population has an IQ between 75 and 90?

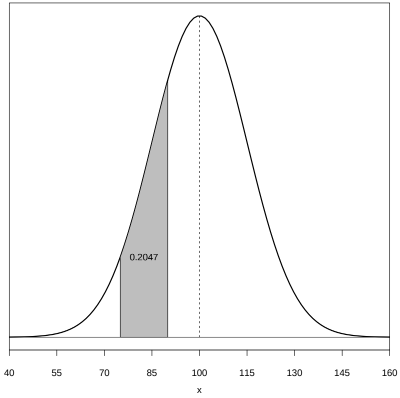
Calculating Probability with R

Look at the picture. The total area to the left of center is 0.50, and you can see that the shaded area is somewhat less than half the left side.

```
> source("http://www.statpower.net/R2101/DrawNormalCurve.R")
> nc(mu=100,sigma=15) ## draw the normal distribution with mu=100,sigma=15
> cn(lo=75,hi=90,mu=100,sigma=15) #shade from 75 to 90 (default is gray)
```



Calculating Probability with R



Calculating Probability with R

- If you just wanted the probability, you could get it directly as follows

```
> pnorm(90,100,15) - pnorm(75,100,15)
[1] 0.2047022
```

Calculating Probability with R

- If SAT scores have a mean of 500 and a standard deviation of 100, what proportion of people have scores above 600?
- If you just wanted the probability, you could get it directly as follows

```
> 1 - pnorm(600,500,100)

[1] 0.1586553
```

- Here is a picture.

Calculating Probability with R

```
> nc(500,100)
> cn(600,1000,500,100,color="green")
```