

2

# Probability

# Experiment

An “**experiment**” or “**random variable**” is any activity whose outcome is unknown or random upfront:

- tossing a coin
- selecting a card from a deck
- measuring the commuting time on a particular morning

**Experiment:** controlled laboratory conditions (hopefully)

**Study:** more general, observational settings

Examples of each? Does one have an advantage over the other?

This distinction is at the core of **causal vs non-causal relationships between variables**. Statistics based on observational studies can only give non-causal interpretations.

# Sample Space

## Definition

The **sample space**, denoted by  $\mathcal{S}$ , is the set of all possible outcomes of an experiment or study.

## Example – continuous outcomes:

Study: Measuring the birth weight of babies born in 2000.

Sample space?

# Examples of sample spaces

The simplest experiment (study) is one with two possible outcomes:

- tossing a coin
- examining a single fuse to see whether it is defective

Sample space for coin toss example?

Sample space for fuse example?

# Events

An **event** is any collection (subset) of outcomes from the sample space  $\mathcal{S}$ .

An event is **simple** if it consists of exactly one outcome and **compound** if it consists of more than one outcome.

When an experiment is performed, a particular event (let's call that event "A") is said to occur if the resulting experimental outcome is contained in A. Examples?

# Events - example

Study: a birth weight for a baby born in 2000

Sample space: set of real numbers

Event A = small for gestational age (SGA)

**How are some ways this event be clearly defined?**

# Events - example

You and your friends (in 3 cars) are off to ski to a remote mountain lodge. The directions you have only take you to the correct highway exit, but after that you're on your own.

So, each of the three vehicles can take a turn left ( $L$ ) or right ( $R$ ) at the end of the exit ramp.

What is the sample space?

# Events - example

cont' d

Some compound events include

$A$  = the event that exactly one of the three vehicles turns right

$B$  = the event that at most one of the vehicles turns right

$C$  = the event that all three vehicles turn in the same direction

How many ways can  $A$  happen? What is the probability  $A$  happens? What about  $B$ ?  $C$ ?



# Set Theory

An **event is a set**, so relationships and results from set theory can be used to study events.

## Definitions

1. The **complement** of an event  $A$ , denoted by  $A'$ , is the set of all outcomes in  $\mathcal{S}$  that are not contained in  $A$ .

1. The **union** of two events  $A$  and  $B$  (denoted by  $A \cup B$ ) is the event consisting of all outcomes that are *either in  $A$  or in  $B$  or in both*. What is a common graphic for this?

1. The **intersection** of two events  $A$  and  $B$  (denoted by  $A \cap B$ ), is the event consisting of all outcomes that are in *both  $A$  and  $B$* .

# Set Theory

Sometimes  $A$  and  $B$  have no outcomes in common, so that the intersection of  $A$  and  $B$  contains no outcomes.

## Definition

Let  $\emptyset$  denote the *null event* (the event consisting of no outcomes whatsoever).

When  $A \cap B = \emptyset$ ,  $A$  and  $B$  are said to be **mutually exclusive** or **disjoint** events.

# Set Theory

The operations of union and intersection can be extended to more than two events.

For any three events  $A$ ,  $B$ , and  $C$ :

-  $A \cup B \cup C = ?$

-  $A \cap B \cap C = ?$

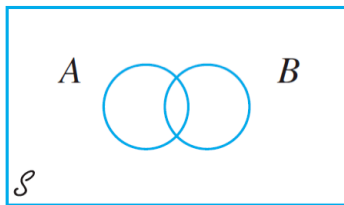
Events are said to be **mutually exclusive** (or pairwise disjoint) if no two events have any outcomes in common.

# Some Relations from Set Theory

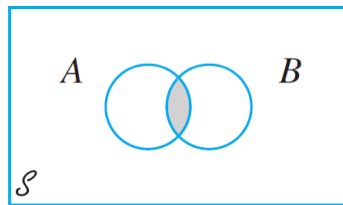
A pictorial representation of events and manipulations with events is obtained by using **Venn diagrams**.

To construct a Venn diagram, draw a rectangle whose interior will represent the sample space  $\mathcal{S}$ .

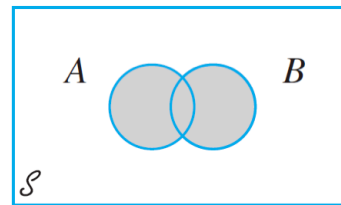
Then any event  $A$  is represented as the interior of a closed curve (simplest: circle) contained in  $\mathcal{S}$  :



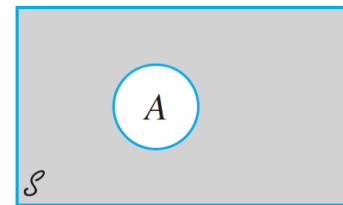
(a) Venn diagram of events  $A$  and  $B$



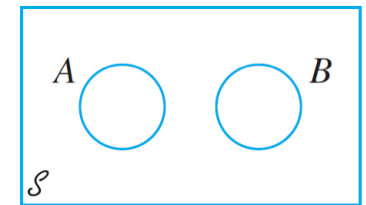
(b) Shaded region is  $A \cap B$



(c) Shaded region is  $A \cup B$



(d) Shaded region is  $A'$



(e) Mutually exclusive events

# Properties of Probability

Given an experiment and a sample space  $\mathcal{S}$ , *the objective of probability is to assign to each event  $A$  a number that is the probability of the event  $A$ , which quantifies how likely is that  $A$  will occur.*

**Axiom 1:** For any event  $A$ ,  $P(A) \geq 0$ .

**Axiom 2:**  $P(\mathcal{S}) = 1$ .

**Axiom 3:**

If  $A_1, A_2, A_3, \dots$  is an (infinite) collection of *disjoint* events:

$$P(A_1 \cup A_2 \cup A_3 \cup \dots) = \sum_{i=1}^{\infty} P(A_i)$$

# More Probability Properties

## Propositions

1. **Law of Total Probability:** For any event  $A$ ,

$$P(A) + P(A') = 1, \text{ and thus } P(A) = 1 - P(A')$$

2. **Inclusion-Exclusion Principle:** For any sets  $A$  and  $B$ ,

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

3.  $P(A \cup B \cup C) = P(A) + P(B) + P(C)$

$$- P(A \cap B) - P(A \cap C) - P(B \cap C)$$

$$+ P(A \cap B \cap C)$$

# What is Probability?

Probability = relative frequencies (simple context)

Consider an experiment that can be repeatedly performed, in an identical and independent fashion, and let  $A$  be an event consisting of a set of outcomes of the experiment.

Simple examples of such repeatable experiments?

# Interpreting Probability

If the experiment is performed  $n$  times, in independent and identical fashion, the event  $A$  will occur (the outcome will be in the set  $A$ ) in some of the replications. In other replications,  $A$  will not occur.

Let  $n(A)$  denote the number of replications on which  $A$  does occur.

Then the ratio  $n(A)/n$  is called the ***relative frequency*** of occurrence of the event  $A$  in the sequence of  $n$  replications of the experiment.



# Example

Let  $A$  be the event that

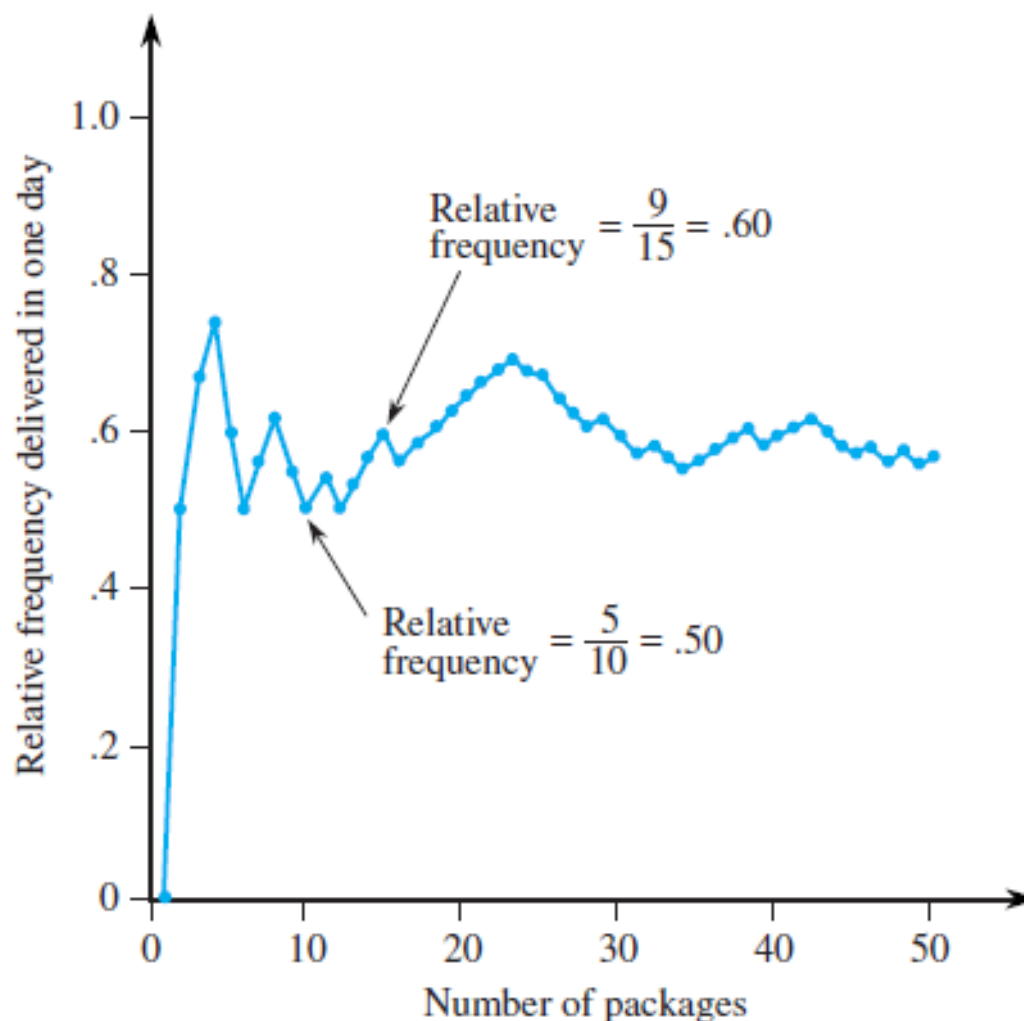
**“a package sent within the state of California for 2<sup>nd</sup> day delivery actually arrives within one day”**

The results from sending 10 such packages (the first 10 replications of the experiment) are as follows:

Package #	1	2	3	4	5	6	7	8	9	10
Did $A$ occur?	N	Y	Y	Y	N	N	Y	Y	N	N
Relative frequency of $A$	0	.5	.667	.75	.6	.5	.571	.625	.556	.5

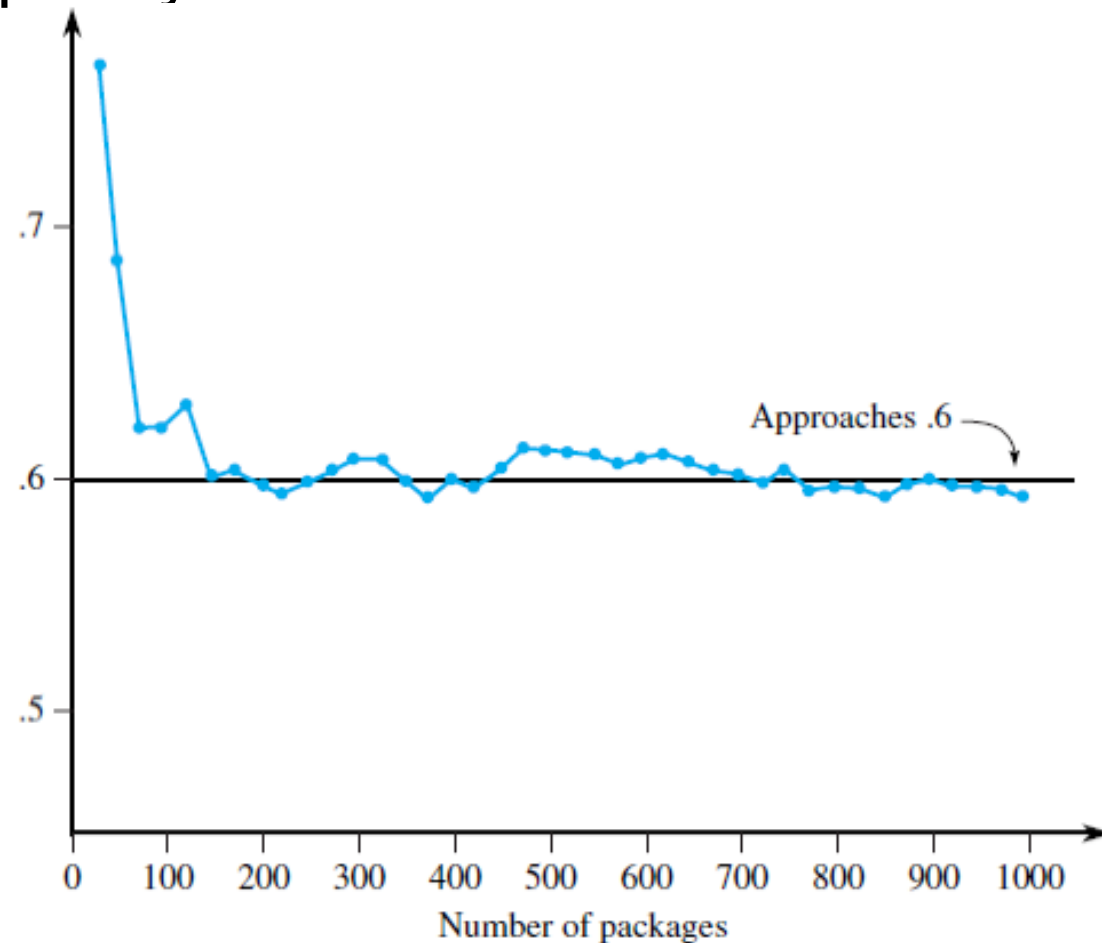
# Example

The relative frequency  $n(A)/n$  fluctuates rather substantially over the course of the first 50 measurements.



# Example

But as the number of replications continues to increase, the relative frequency stabilizes.



Behavior of relative frequency -- Long-run stabilization

# Interpreting Probability

Empirical evidence (based on the results of many such repeatable experiments) indicates that the relative frequency of this sort will stabilize as the number of replications  $n$  increases.

That is, as  $n$  gets very large,  $n(\mathbf{A}) / n$  approaches a limiting value referred to as the limiting (or long-run) relative frequency of the event  $A$ .

The objective interpretation of probability states that this **limiting relative frequency of an event  $A$  is the probability of that event,  $P(A)$ .**

# Interpreting Probability

What does the statement “the probability of a package being delivered within one day of mailing is 0.6” really mean?

How does this relate to tossing a “fair coin”?

# R CODE FOR DELIVERY EXAMPLE

# Assigning Probabilities to events

There are several ways we can learn about probabilities of events:

- 1) Empirically – ie, assess probabilities based on experience
- 2) Analytically, when we know more about the problem
- 3) Computationally:
  - 1) Using counting techniques
  - 2) Using combinatorial tricks
  - 3) Using computer simulations (ie, simulating repeated experiments, or put differently, simulating “experience”)

## Computing probabilities: analytical computation

A train has five cars. Suppose a commuter is twice as likely to select the middle car (#3) as to select either adjacent car (#2 or #4), and is twice as likely to select any of the adjacent cars as to select any of the end cars (#1 or #5).

What is the probability of the commuter selecting each car?



## Counting: enumeration for equally likely outcomes

Think of drawing a card at random from a deck of cards.

If there are  $N$  equally likely outcomes, the probability for each is  $1/N$ .

How do we figure out the probability of an even such as  $A = \{\text{ace}\}$ ?

## Counting – enumeration without order

A family requires the services of an obstetrician and a pediatrician. There are two medical clinics, each having 2 obstetricians and 3 pediatricians.

The family wishes to select both doctors from the same clinic. In how many ways can this be done?

## Counting – enumeration with order

If a six-sided die is tossed five times, then the outcome is an ordered collection of five numbers -- a “5-tuple”.

We will call an ordered collection of  $k$  objects a ***k-tuple***

There are  $n_1 n_2 \cdot \cdot \cdot \cdot n_k$  possible  $k$ -tuples.

**In the dice example, how many possible outcomes are there for the 5-tuple?**

# Permutations and Combinations

A subset where order matters is called a **permutation**: The number of permutations of size  $k$  that can be formed from  $n$  objects will be denoted by  $P_{k,n}$ .

An unordered subset is called a **combination**: Denoted as  $C_{k,n}$  or  $\binom{n}{k}$  (pronounced: “ $n$  choose  $k$ ”).

# Permutations and Combinations

Example: A college of engineering has 7 departments. Each has one representative on the student council.

From these 7 representatives, one is to be chosen chair, another vice-chair, and one secretary.

Is this a permutation or a combination? How many ways are there to select these three officers?

# Combinations

Now suppose that 3 of the 7 representatives are to be selected to attend a convention.

Why is this a combination and not a permutation?

How do we calculate the number of ways this can happen?

# Combinations

Which number is larger: the number of combinations or permutations?

Why?

# Combinations

Generalizing ....

## Definition

$$\binom{n}{k} = \frac{P_{k,n}}{k!} = \frac{n!}{k!(n-k)!}$$

Notice that  $\binom{n}{n} = 1$  and  $\binom{n}{0} = 1$  since there is only one way to choose a set of (all)  $n$  elements or of no elements, and  $\binom{n}{1} = n$  since there are  $n$  subsets of size 1.



## Using permutations and combinations

A particular iPod playlist contains 100 songs, 10 of which are by the Beatles.

Suppose the shuffle feature is used to play the songs in random order, without repetition. What is the probability that the 1st Beatles song heard is the 5th song played?

# Conditional Probability

With conditional probability, we examine how the information “an event  $B$  has occurred” affects the probability assigned to  $A$ .

For example,  $A$  might refer to an individual having a particular disease in the presence of certain symptoms.

If a blood test is performed on the individual and the result is negative, then the updated probability of disease will be different than if the test result was positive.

We will use the notation  $P(A | B)$  to represent the **conditional probability of event  $A$  given that the event  $B$  has occurred**.  $B$  is the “conditioning event.”

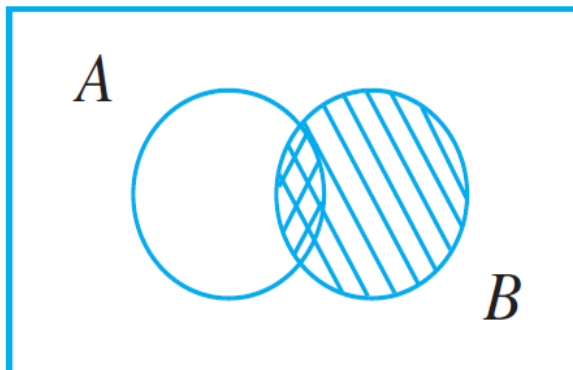
# Conditional Probability

The conditional probability is expressed as a ratio of unconditional probabilities --

probability of the intersection of the two events

-----  
probability of the conditioning event  $B$

Given that  $B$  has occurred, the relevant sample space is no longer  $S$  but it boils down to only the outcomes in  $B$ .



# Example – conditional probability

Specific computer parts are assembled in a plant that uses two different assembly lines,  $A$  and  $A'$ . Line  $A$  uses older equipment than  $A'$ , so it is somewhat slower and less reliable.

Suppose on a given day line  $A$  has assembled 8 parts, whereas  $A'$  has produced 10.

From the 8 parts from  $A$ , 2 were defective and 6 as nondefective. From the 10 parts from  $A'$ , 1 was defective and 9 nondefective.

# Example, cont

cont' d

Say for example, that the manager chose a part that turned out to be defective – ie, the event  $D=\{defective\ part\}$  has occurred.

What is the chance that it was made by the line A? How is this different than the probability that line A makes a defective part?

# The Multiplication Rule for $P(A \cap B)$

The definition of conditional probability yields the following result:

## The Multiplication Rule

$$P(A \cap B) = P(A|B) \cdot P(B)$$

# Independence

Previous examples:

The **condition probability**,  $P(A | B)$  differed from the **unconditional probability**  $P(A)$ . The information “ $B$  has occurred” resulted in a change in the chance of  $A$  occurring.

If events are independent:

The chance that  $A$  will occur is not affected by knowledge that  $B$  has occurred.

How can we write this in mathematical notation?

# Independence

## Definition

Two events  $A$  and  $B$  are **independent** if  $P(A | B) = P(A)$  and are **dependent** otherwise.

**IMPORTANT:** The definition of independence is “unsymmetric.”



# Independence

However, using the definition of conditional probability and the multiplication rule,

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)P(B)}{P(A)} \quad (2.7)$$

The right-hand side of Equation (2.7) is  $P(B)$  if and only if  $P(A|B) = P(A)$  (independence), so the equality in the definition implies the other equality (and vice versa).

If  $A$  and  $B$  are independent, then so are:

(1)  $A'$  and  $B$ , (2)  $A$  and  $B'$ , and (3)  $A'$  and  $B'$ .

# The Multiplication Rule for $P(A \cap B)$

## Proposition

$A$  and  $B$  are independent if and only if (iff)

$$P(A \cap B) = P(A) \cdot P(B) \quad (2.8)$$

How can this be verified using the multiplication rule?

# Independence of More Than Two Events

## Definition

Events  $A_1, \dots, A_n$  are **mutually independent** if for every  $k$  ( $k = 2, 3, \dots, n$ ) and every subset of indices  $i_1, i_2, \dots, i_k$ ,

$$P(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \dots \cdot P(A_{i_k})$$

# Example

cont' d

The multiplication rule is most useful when the experiment consists of several stages in succession.

The conditioning event  $B$  then describes the outcome of the first stage and  $A$  the outcome of the second, so that  $P(A | B)$  —conditioning on what occurs first—will often be known.

The rule is easily extended to experiments involving more than two stages.

# Bayes' Theorem

The computation of a posterior probability  $P(A_j|B)$  from given prior probabilities  $P(A_i)$  and conditional probabilities  $P(B|A_i)$  occupies a central position in elementary probability.

The general rule for such computations, which is really just a simple application of the multiplication rule, goes back to Reverend Thomas Bayes, who lived in the eighteenth century.

To state it we first need another result. Recall that events  $A_1, \dots, A_k$  are mutually exclusive if no two have any common outcomes. The events are *exhaustive* if one  $A_i$  must occur, so that  $A_1 \cup \dots \cup A_k = \mathcal{S}$ .

# The Law of Total Probability

Let  $A_1, \dots, A_k$  be mutually exclusive and exhaustive events. Then for any other event  $B$ ,

$$P(B) = P(B|A_1)P(A_1) + \dots + P(B|A_k)P(A_k)$$

$$= \sum_{i=1}^k P(B|A_i)P(A_i) \quad (2.5)$$

# Example

An individual has 3 different email accounts. Most of the messages, in fact 70%, come into account #1, whereas 20% come into account #2 and the remaining 10% into account #3.

Of the messages into account #1, only 1% are spam, whereas the corresponding percentages for accounts #2 and #3 are 2% and 5%, respectively.

What is the probability that a randomly selected message is spam?

Say she randomly selected a message and it was indeed spam. What is the probability that it came from account #1?

# Bayes' Theorem

Let  $A_1, A_2, \dots, A_k$  be a collection of  $k$  mutually exclusive and exhaustive events with *prior* probabilities  $P(A_i)$

Then for any other event  $B$  for which  $P(B) > 0$ , the *posterior* probability of  $A_j$  given that  $B$  has occurred is

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^k P(B|A_i) \cdot P(A_i)} \quad j = 1, \dots, k \quad (2.6)$$