Statistical Methods for the Biological, Environmental, and Health Sciences

STAT 007

Probability

Chapter 4

Basic Concepts of Probability

Section 4-1

- In this section we will:
  - Discuss how to compute the probability of something not happening (the complement).

Complementary Events

- Sometimes we need to find the probability that an event A *does not* occur.
- The **complement** of event A, denoted by $\overline{A}$, consists of all outcomes in which event A does not occur.
- The following relationship between A and $\overline{A}$ holds:

$$P(\overline{A}) = 1 - P(A).$$

Example

In recent years, there were about 3 million skydiving jumps and 21 of them resulted in deaths.

Use the frequentist approach to find the probability of *not* dying when making a skydiving jump.

Practice

Look at the exercises at the end of Section 4-1 in page 128

Specially, look at exercises:
1, 2, 3, 4, 5, 6, 7, 8, 13-20, 25, 26, 27, 29, 30, 31, 33, 34, 35, 36

Addition Rule and Multiplication Rule

Section 4-2

- In this section we will:
  - Introduce the addition rule to compute $P(A \text{ or } B)$.
  - Introduce the multiplication rule to compute $P(A \text{ and } B)$.
  - Discuss disjoint and independent events.

- Consider a procedure with some sample space, and consider A and B two events.
- In this section we will discuss how to compute the probability that A occurs, or B occurs, or A and B occur. This is the probability of the "new" event A *or* B: $P(A \textit{ or } B)$.
- Be cautious to not *double count* events!
- We will also discuss how to compute the probability that A and B occur. This is, the probability of the "new" event A *and* B: $P(A \textit{ and } B)$.
- Be cautious to identify if event A somehow affects the probability of event B.

Addition Rule: **A or B**

- The *Addition Rule* is used to compute the probability of the event *A or B*. This is $P(A \text{ or } B)$.
- Intuitively: To find **P(A or B) add** the probability of event *A* and the probability of event *B*, and if there is any overlap that causes double-counting, subtract the probability of outcomes that are included twice.
- Formally: **P(A or B)** $= P(A) + P(B) - P(A \text{ and } B)$.

Example

TABLE 4-1 Results from Drug Tests of Job Applicants

	Positive Test Result (Test shows drug use.)	Negative Test Result (Test shows no drug use.)
Subject Uses Drugs	45 (True Positive)	5 (False Negative)
Subject Does Not Use Drugs	25 (False Positive)	480 (True Negative)

- **Drug Testing of Job Applicants**: The table includes results form 555 adults in the U.S..
- Assume that one subject is randomly chosen, find the probability of selecting a subject who had a positive test result or uses drugs.

Addition Rule for events that are disjoint

- **Definition**: Events A and B are **disjoint** (or mutually exclusive) if they cannot occur at the same time. (Disjoint events do not overlap.)
- If events A and B are disjoint, then $P(A \text{ and } B) = 0$, therefore $P(A \text{ or } B) = P(A) + P(B)$.

Example

Discuss whether the following events are disjoint or not and how does this translate into a mathematical formula.

a) Event A: randomly select a subject for a clinical trial who is a male.
   Event B: randomly select a subject for a clinical trial who is a female.

b) Event A: randomly select a subject taking a statistics course.
   Event B: randomly select a subject who is a female.

Addition Rule and complementary events

- Principle: we are certain that either and event occurs or it does not occur.
- Use the addition rule to translate the previous principle into a mathematical expression. And show that
  - $P(A) + P(\bar{A}) = 1$.
  - $P(\bar{A}) = 1 - P(A)$.
  - $P(A) = 1 - P(\bar{A})$.

Example

Based on a journal article, the probability of randomly selecting someone who has sleepwalked is 0.292 (based on data from "Prevalence and Comorbidity of Nocturnal Wandering in the U.S. General Population," by Ohayon et al., *Neurology*, Vol. 78, No. 20).

If a person is randomly selected, find the probability of getting someone who has not sleepwalked.

Multiplication Multiplication: **A and B**

- The *Multiplication Rule* is used to compute the probability of the event A *and* B, where events A and B occur in different trials. Notation: $P(A \text{ and } B)$.
- Intuitively: To find **P(A and B) multiply** the probability of event A by the probability of event B, but be sure that the probability of event B is found by assuming that event A has already occurred.
- Formally: **P(A and B)** $= P(A)P(B \mid A)$.
- Events A and B are **independent** if the occurrence of one does not affect the probability of the occurrence of the other.
- If A and B are independent events, then $P(A \text{ and } B) = P(A)P(B)$
- If events A and B are not independent, they are said to be **dependent**.

Multiplication Multiplication: **A and B**

Example

Drug Testing of Job Applicants: Consider only the 50 test results from the subjects who use drugs. The number of subjects that had a positive test result is 45 and the number of subjects that had a negative test result is 5.

a) If 2 of these 50 subjects are randomly selected *with replacement*, find the probability that the first selected person had a positive test result and the second selected person had a negative test result.

b) Repeat part a) by assuming that the two subjects are randomly selected *without replacement*.

Practice

Look at the exercises at the end of Section 421 in page 143

Specially, look at exercises:
1, 2, 4, 5, 6, 7, 9-20, 21-24, 27

Complements, Conditional Probabilities and Bayes' Theorem

Section 4-3

- In this section we will:
  - Define conditional probabilities.
  - Introduce Bayes' Theorem.
  - Discuss disjoint and independent events.
  - Extend the use of the addition rule ($P(A \text{ or } B)$).

- Conditional probabilities are use to find the probability of an event occurring, giving that another event has already occurred.
- Bayes' theorem is also used to compute conditional probabilities.
- Conditional probabilities can be understood as an update of uncertainty of an event, given that another event has already occur.
- Finally, we will extend the use of the addition rule and use to compute probabilities for events happening "at least once". For finding this probability we will use complements.

Conditional Probabilities

- The probability of an event can be affected by knowledge that some other event has occurred.
- If this is the case, we compute conditional probabilities.
- Intuition: The **conditional probability** of event B occurring given that event A has occurred can be found by assuming that event A has occurred and then calculating the probability that event B will occur.
- Notation $P(B \mid A)$.
- Formal: The probability $P(B \mid A)$ can be found by dividing the probability of events A and B both occurring by the probability of event A:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}.$$

Conditional Probabilities

Example

TABLE 4-1 Results from Drug Tests of Job Applicants

	Positive Test Result (Test shows drug use.)	Negative Test Result (Test shows no drug use.)
Subject Uses Drugs	45 (True Positive)	5 (False Negative)
Subject Does Not Use Drugs	25 (False Positive)	480 (True Negative)

- **Drug Testing of Job Applicants**: The table includes results form 555 adults in the U.S..
- Assume that one subject is randomly chosen.

a) Find the probability that the subject had a positive test result, given that the subject actually uses drugs.

b) Find the probability that the subject actually uses drugs, given that he or she had a positive test result.

Bayes' Theorem

- Bayes' theorem is used to compute conditional probabilities.
- Bayes' theorem can be used with sequential events. In this case, new information is used to *update* the probability of the initial event.
- Formal definition:

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)} = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid \bar{A})P(\bar{A})}.$$

Example

Assume that subjects are applying to a Job and they are tested for use of drugs. Assume that the probability of a subject having a positive test result is $70/555$, that the probability of a subject using drug is $50/555$, and that the probability of a subject having a positive test results given that the subject uses drug is $45/50$.

Compute the probability that the subject uses drugs given that the subject tested positive.

Complements: The Probability of "At Least One"

- First, note that the probability of some event occurring "at least once" has the same meaning as some event occurring "one or more" times.
- Second, note that the complement of getting "at least one" particular event is to get no occurrences of that event.
- So, $P(\text{event } A \text{ ocurring at least once}) = 1 - P(\text{no ocurrences of event } A)$.

Example

Consider the procedure of *sequentially* sampling at random three students from a class and noting their gender. Let "g" denote a girl and let "b" denote a boy. Assume that sampling a boy or a girl is equally likely. This implies that all simple events are equally likely.

Find the probability of sampling at least one girl when sampling three students from the class.

Practice

Look at the exercises at the end of Section 4-3 in page 150.

Specially, look at exercises:
3, 13-16, 17-20.

Extra Exercises

Example

According to the pattern on its back a beetle can be from a rare subspecies or from a common subspecies.

Assume that the probability of a beetle being from a rare subspecies is 0.001, the probability of a beetle having the pattern is 0.05093, and the probability of a beetle having the pattern given that it is from the rare subspecies is 0.98.

What is the probability of a beetle being from a rare subspecies given that it has the pattern?

Extra Exercises

Example

Consider a test for a disease that is 90% reliable in the sense that if a person has the disease, there is a 0.9 probability that the test will give a positive response; whereas, if a person does not have the disease, there is a probability of only 0.1 that the test will give a positive response.

Additionally, data indicate that your chances of having the disease are only 1 in 10,000.

What is the probability that you have the disease given that you had a positive response to the test?