Statistical Methods for the Biological, Environmental, and Health Sciences

STAT 007

Probability

Chapter 4

Complements, Conditional Probabilities and Bayes' Theorem

Section 4-3

- In this section we will:
  - Define conditional probabilities.
  - Introduce Bayes' Theorem.

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.

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?

Practice

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

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

Discrete Probability Distributions

Chapter 5

Probability Distributions

Section 5-1

- In this section we will:
  - Define random variables and their probability distributions.
  - Introduce parameters, such as mean, variance and standard deviations, of probability distributions.

- In Chapters 2 and 3 we described data that were collected from a procedure (experiment or observational study).
- In Chapter 4 we introduced the concept of probability and used it to compute probabilities of events. We also learned some rules and definitions (addition and multiplication rules; conditional probabilities and Bayes' theorem).
- Now we will introduce the concept of *random variable* to (theoretically) describe an uncertain procedure of interest.
- We will assign probabilities to each possible numerical value of the random variable and discuss some theoretical properties of the uncertain procedure of interest (the random variable).
- The data sets that we collect (analysis in Ch. 2 and 3) are the values we observed for the random variable of interest. The frequencies of each value are related to the probability distribution of the random variable.

Random Variables and Probability Distributions

- A **random variable** is a variable that has a numerical value (denoted x), determined by chance, for each outcome of a procedure.
  - A **discrete random variable** has a collection of values that if finite or countable.
  - A **continuous random variable** has infinitely many values, and the collection of value is not countable.
- A **probability distribution** is a description that gives the probability for each value of the random variable, notation $P(x)$. It is often expressed in the format of a table, formula, or graph. It must satisfy the following:
  a) There is a random variable x, and its number values are associated with corresponding probabilities.
  b) The sum of the probabilities of all values of the random variable must be one, this is, $\sum P(x) = 1$.
  c) The probability of each value of the random variable must be between 0 and 1, inclusive, this is $0 \leq P(x) \leq 1$.

Random Variables and Probability Distributions

Example

Consider a procedure, in which it is of interest to describe the number of girls that are born in two births. Consider a random variable that describes the number of girls in two births, this is

$$x = \text{number of girls in two births.}$$

The probability the number of girls being zero is 0.25 and the probability of the number of girls being 1 is 0.5.

- What possible values can the random variable take?
- Write a table with the probability distribution of the number of girls in two births.
- Is the previous table a probability distribution?
- Find the probability of the number of girls being 3.
- Draw a probability histogram of the number of girls in two births.

Parameters of a Probability Distribution

- The **Mean** (also known as Expected Value) of a discrete random variable is a parameter denoted μ and is the theoretical mean outcome for infinitely many trials. It is a measure of center of the random variable.

$$\mu = \sum [x * P(x)].$$

- The **Variance** of a discrete random variable is a parameter denoted σ^2.

$$\sigma^2 = \sum [(x - \mu)^2 * P(x)] = \sum [x^2 * P(x)] - \mu^2.$$

- The **Standard Deviation** of a discrete random variable is a parameter denoted σ.

$$\sigma = \sqrt{\sigma^2} = \sqrt{\sum [x^2 * P(x)] - \mu^2}.$$

- We can identify significantly low or high values of the random variable with the following Rule of Thumb:
  — **Significantly low** values are $(\mu - 2\sigma)$ or lower.
  — **Significantly high** values are $(\mu + 2\sigma)$ or higher.
  — **Not Significant** values are between $(\mu - 2\sigma)$ and $(\mu + 2\sigma)$.

Parameters of a Probability Distribution

Example

Consider a procedure, in which it is of interest to describe the number of girls that are born in two births. Consider a random variable that describes the number of girls in two births, this is

$$x = \text{number of girls in two births}$$

and the following probability distribution $P(x) = \frac{1}{2(2-x)!x!}$.

- Find the mean and the variance of the number of girls in two births.
- Is the value 2 a significantly high value?

Practice

Look at the exercises at the end of Section 5-1 in page 710.

Specially, look at exercises:
1, 2, 3, 5, 6, 7-12, 13, 14, 15, 16, 17, 18.

Binomial Probability Distributions

Sections 5-2

- In this section we will:
  - Introduce the Binomial probability distribution to measure uncertainty in procedures involving number of successes in a finite number of trial.
  - Identify parameters such as mean, variance, and standard deviation.

Binomial Distribution

Definition

A binomial probability distribution can be used when

1. *The procedure has a fixed number of trials. (A trial is a single success/failure observation.)*
2. *The trials are independent.*
3. *Each trial must have all outcomes classified into exactly two categories.*
4. *The probability of success remains the same in all trials.*

*The probability distribution formula of the Binomial distribution with parameters **n** and **p** is given by*

$$P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \dots, n,$$

where

$$\begin{aligned} x &= \textit{number of successes in n trials}, \\ p &= \textit{probability of succes in ech trial.} \end{aligned}$$

Parameters: $\mu = np, \quad \sigma^2 = np(1-p), \quad \sigma = \sqrt{np(1-p)}.$

Example

When Gregor Mendel conducted his famous hybridization experiments, he used peas with green pods and peas with yellow pods. Because green is dominant and yellow is recessive, when crossing two parents with the green/yellow pair of genes, we expect that 3/4 of the offspring peas should have green pods. That is, $P(\text{green pod}) = 3/4 = 0.75$. Assume that all parents have the green/yellow combination of genes.

a) Would the Binomial probability distribution be a fit if we want to find the probability that exactly three of five offspring peas have green pods?

b) Find the probability of getting exactly 3 peas with green pods when 5 offspring peas are generated.

c) In an actual experiment, Mendel generated 580 offspring peas. He claimed that 75%, of them would have green pods. The actual experiment resulted in 428 peas with green pods.
   - c1) Find the mean and standard deviation for the numbers of peas with green pods.
   - c2) Would you conclude that Mendels actual result of 428 peas with green pods is significantly low or significantly high? Does this suggest that Mendels value of 75% is wrong?

 Discrete Probability Distributions

Practice

Look at the exercises at the end of Section 5-2 in page 201.

Specially, look at exercises:
1, 2, 10, 11, 12, 15-20, 21-24, 25, 26.

Bayes' Theorem: 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?