Probability and Conditional Probability

Bret Hanlon and Bret Larget

Department of Statistics
University of Wisconsin—Madison

September 27–29, 2011

Parasitic Fish

Case Study

Example 9.3 beginning on page 213 of the text describes an experiment in which fish are placed in a large tank for a period of time and some are eaten by large birds of prey. The fish are categorized by their level of parasitic infection, either uninfected, lightly infected, or highly infected. It is to the parasites advantage to be in a fish that is eaten, as this provides an opportunity to infect the bird in the parasites next stage of life. The observed proportions of fish eaten are quite different among the categories.

	Uninfected	Lightly Infected	Highly Infected	Total
Eaten	1	10	37	48
Not eaten	49	35	9	93
Total	50	45	46	141

The proportions of eaten fish are, respectively, $1/50 = 0.02$, $10/45 = 0.222$, and $37/46 = 0.804$.

Questions

- There are three *conditional probabilities* of interest, each the probability of being eaten by a bird given a particular infection level.
- How do we test if these are the same?
- How do we estimate differences between the probability of being eaten in different groups?
- Is there a relationship between infection level in the fish and bird predation?

Vampire Bats

Case Study

Example 9.4 on page 220 describes an experiment. In Costa Rica, the vampire bat *Desmodus rotundus* feeds on the blood of domestic cattle. If the bats respond to a hormonal signal, cows in estrous (in heat) may be bitten with a different probability than cows not in estrous. (The researcher could tell the difference by harnessing painted sponges to the undersides of bulls who would leave their mark during the night.)

	In estrous	Not in estrous	Total
Bitten by a bat	15	6	21
Not bitten by a bat	7	322	329
Total	22	328	350

The proportion of bitten cows among those in estrous is $15/22 = 0.682$ while the proportion of bitten cows among those not in estrous is $6/328 = 0.018$.

Questions

- Are the probabilities of being bitten different for cows in estrous or not?
- How do we estimate the difference in probabilities of being bitten?
- How do we estimate the odds ratio?
- Here, the odds of a cow in estrous being bitten are roughly 2 to 1, while the odds of a cow not in estrous being bitten are roughly 2 to 100, so the odds ratio is about 100 times larger to be bitten for cows in estrous compared to those not.
- How do we quantify uncertainty in this estimate?

The Big Picture

- When comparing two categorical variables, it is useful to summarize the data in tables.
- Data in the tables can be used to calculate observed proportions sampled from different populations.
- We may have interest in estimating differences between population probabilities.
- We may wish to test if population proportions are different.
- We may wish to test if two categorical variables are independent.

More Probability

- To understand the methods for comparing probabilities in different populations and analyzing categorical data, we need to develop notions of:
  - *conditional probability*; and
  - *independence*;
- We will also more formally introduce some probability ideas we have been using informally.

Running Example

Example

Bucket 1 contains colored balls in the following proportions:

- 10% red;
- 60% white; and
- 30% black.

Bucket 2 has colored balls in different proportions:

- 10% red;
- 40% white; and
- 50% black.

A bucket is selected at random with equal probabilities and a single ball is selected at random from that bucket.

- Think of the buckets as two biological populations and the colors as traits.

Outcome Space

Definition

- A *random experiment* is a setting where something happens by chance.
- An *outcome space* is the set of all possible elementary outcomes.
- An *elementary outcome* is a complete description of a single result from the random experiment (which, in fact, might be rather complicated, but is called elementary because it cannot be divided any further).

Example

- In the example, one elementary outcome is $(1, W)$ meaning Bucket 1 is selected and a white ball is drawn.
- The outcome space is the set of six possible elementary outcomes:

$$\Omega = \{(1, R), (1, W), (1, B), (2, R), (2, W), (2, B)\}$$

Probability

Definition

- A *probability* is a number between 0 and 1 that represents the chance of an outcome.
- Each elementary outcome has an associated probability.
- The sum of probabilities over all outcomes in the outcome space is 1.

Example

$$\begin{array}{lll} P((1,R)) = 0.05 & P((1,W)) = 0.30 & P((1,B)) = 0.15 \\ P((2,R)) = 0.05 & P((2,W)) = 0.20 & P((1,B)) = 0.25 \end{array}$$

Events

Definition

- An *event* is a subset (possible empty, possibly complete) of elementary outcomes from the outcome space.
- The probability of an event is the sum of probabilities of the outcomes it contains.

Example

- $P(\text{Bucket 1}) = P(\{(1, R), (1, W), (1, B)\}) = 0.05 + 0.30 + 0.15 = 0.5$.
- $P(\text{Red Ball}) = P(\{(1, R), (2, R)\}) = 0.05 + 0.05 = 0.1$.
- $P(\text{Bucket 1 and Red Ball}) = P(\{(1, R)\}) = 0.05$.
- $P(\Omega) = 1$.

Combining Events

Definition

Consider events A and B.

- The *union* (or) of two events is the set of all outcomes in either or both events. Notation: $A \cup B$.
- The *intersection* (and) of two events is the set of all outcomes in both events. Notation: $A \cap B$.
- The *complement* (not) of an event is the set of everything in Ω not in the event. Notation: A^c.

Example

Let $A = \{\text{Bucket 1}\} = \{(1, R), (1, W), (1, B)\}$ and $B = \{\text{Red Ball}\} = \{(1, R), (2, R)\}$.

- $A \cup B = \{\text{Bucket 1 or Red Ball}\} = \{(1, R), (1, W), (1, B), (2, R)\}$.
- $A \cap B = \{\text{Bucket 1 and Red Ball}\} = \{(1, R)\}$.
- $A^c = \{\text{not Bucket 1}\} = \{(2, R), (2, W), (2, B)\}$.