Events and Probabilities

School of Information Studies
Syracuse University

Learning Topics for This Week

Define events and trials.

Demonstrate using outcome tables to reason about random events.

Convert event counts into probabilities.

Using R, generate and plot random distributions to reason about random events.

Define histogram; use the table() and barplot() functions in R.

Calculate cumulative probabilities; interpret a bar chart.

Use contingency tables to reason about a more complex set of events.

Compute and interpret marginal totals in a contingency table.

Isolate and normalize a row or column of a contingency table to reason about the probabilities of events when new information is learned.

Connect with Bayesian reasoning about prior and posterior probability.

Define Events and Trials

Events and Trials

A “trial” contains a specific number of events, where each event may have a particular outcome; examples include:

- One trial of coin tosses where each trial consists of 10 coin flips (0 to 10 heads)
- One trial of dice throwing that includes two dice (the sum of the two dice may range from 2 to 12)
- One trial of playoffs (for one team) that comprises four games (the team may have zero to four wins)
- One trial of five defendants in a bank robbery case (zero to five convictions)

To find out the **likelihood of the various events within a trial,** we run multiple trials

- 100 trials of throwing two dice, taking the sum of the two dice each time, and tabulating how often each sum occurs

Toast, Events, and Trials

Imagine a reality TV show about a restaurant where the waiters always drop the food on the way to the table. Each order contains six pieces of toast—what we will call **six events**—and each piece of toast may fall with the topping down or topping up. When one waiter drops one order of (six pieces of) toast, we call that one **trial.** Here's an example of data that might come from 100 trials:

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1

Which Event Is Most Frequent?

In the data we collected from our imaginary TV show, we have a range of scenarios involving different numbers of jelly-down toast. The event that occurs the most is three pieces of toast with the jelly side down. Why do you think that three jelly sides down is the most likely event?

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1

Each Piece of Toast Is 50/50

As each piece of toast falls from the tray to the floor, its fate is independent of the other pieces. We can also surmise that there is a 50/50 chance of jelly side down for any particular piece of toast. Having all of the toast fall jelly side down (or up) seems unlikely. In any given trial, about half of the events (toast landings) will be jelly down.

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1

Demonstrate Using Outcome Tables to Reason About Random Events

Common vs. Rare Events

People who play cards or dice often have an intuitive feel for the likelihood of various outcomes. Drawing four kings in a row from the top of a shuffled deck is rare. Throwing “snake eyes” (two dice showing 1 on top) only happens once in a while. In the table below, it is rare for a tray of toast to land where all of the pieces land jelly side down or all of them land jelly side up. In the trials below, those extreme events occurred only 5 out of 100 times.

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1

Why Seven Outcomes?

In a trial that involves six pieces of toast, there are seven possible outcomes because zero jelly sides down is one of the possibilities. You can think of each event as having the value of zero or one. Jelly side up gets a zero and jelly side down gets a one. The result of each trial is therefore simply the sum of the component events. When events can only have one of two different values, we call this “binomial” (two names). The binomial distribution is a mathematical construct that is valuable in reasoning about coin tosses, toast drops, and other events.

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1

Convert Event Counts Into Probabilities

Proportions and Probability

Simply divide the number of events in each category by the total number of trials (in this case, 100) to arrive at the probabilities of each event. The probabilities should sum to 1. A probability is therefore simply a proportion of events.

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1
Probability of that count	0.04	0.09	0.2	0.34	0.21	0.11	0.01

Probability Theory

The first known scientific work on probability, *On Reasoning in Games of Chance*, was published by Dutch scientist Christiaan Huygens in 1657.

Using the terminology introduced earlier, we define probability as the relative frequency of a particular event in a "universe" of trials.

Additionally, there is a notion of **subjective probability** that we will use later in the class that refers to human beliefs about the likelihood of an event.

Portrait of Christiaan Huygens. Picture credit: Wikipedia. Public domain.

Using R, Generate and Plot Random Distributions to Reason About Random Events

Using rbinom()

R has the capability to generate random numbers in a variety of configurations. Each configuration of random numbers reflects a specific underlying **distribution.**

When working with events that have two states, such as heads and tails, the best distribution to use is the **binomial distribution.**

In R, the rbinom() command generates random trials of the binomial distribution.

Try this command:

```
rbinom(n=20, size=6, prob=0.5)
```

Make Your Own Toast Drop Trials

Use the table command to organize your results. This avoids listing the result of each toast-drop event on the console.

The table command will tabulate the number of events in each category and provide a compact overview of the results of your trials.

Try this command to generate a table of 100 trials with six toast drops each:

```
table( rbinom(n=100,size=6,prob=0.5) )
```

Define Histogram and Use the Table() and Barplot() Functions in R

Histograms

A histogram is a frequency plot that summarizes numeric data.

Numeric data are sorted and binned into several categories. Each category generally contains a range of values.

An algorithm decides how many bins are needed, based on the amount and range of the data.

The shape of a histogram—whether it is symmetric, tall, flat, or lumpy—can help guide the choice of how to analyze data.

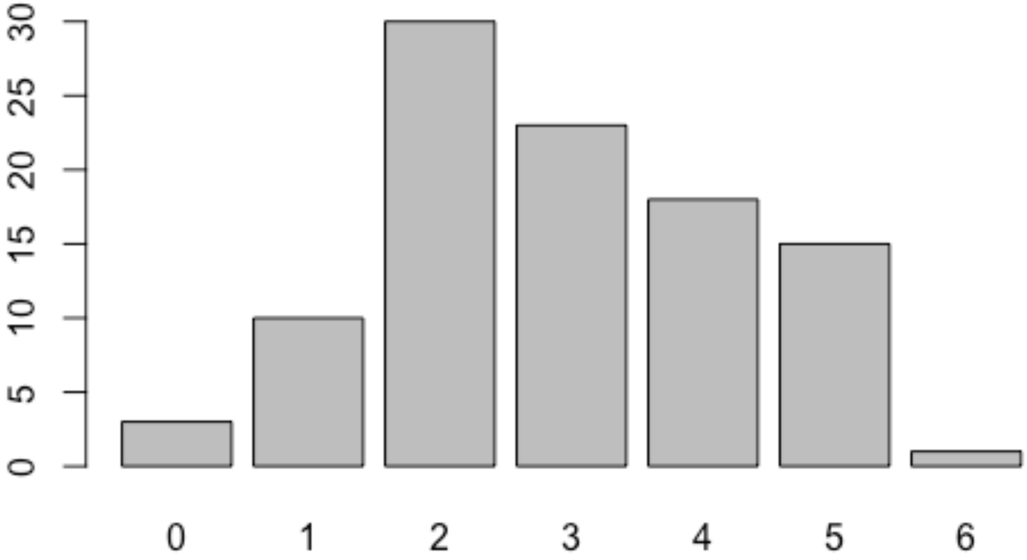
Try this command to show a simple histogram:

```
hist(c(1,1, 2,2,2))
```

Use Barplot() to Show Histogram of Trials

The barplot() command allows us to plot the frequency of occurrence of various discrete events. We can use barplot() when we don't want hist() to sort the data into bins:

```
barplot( table( rbinom(n=100,size=6,prob=0.5) ) )
```



What happens if you run this command several times?

Why Does the Barplot Change?

Why does the barplot change each time we run the **barplot(table(rbinom(n=100,size=6,prob=0.5)))** command?

Results will look slightly different each time because rbinom() generates random numbers. Occasionally, results may not even show anything in the zero category or the six category. Running 10,000 trials (or more) will get a better looking graph. Raising the number of trials will eventually begin to show how the binomial distribution approximates the normal curve.

Calculate Cumulative Probabilities and Interpret a Bar Chart of Cumulative Probabilities

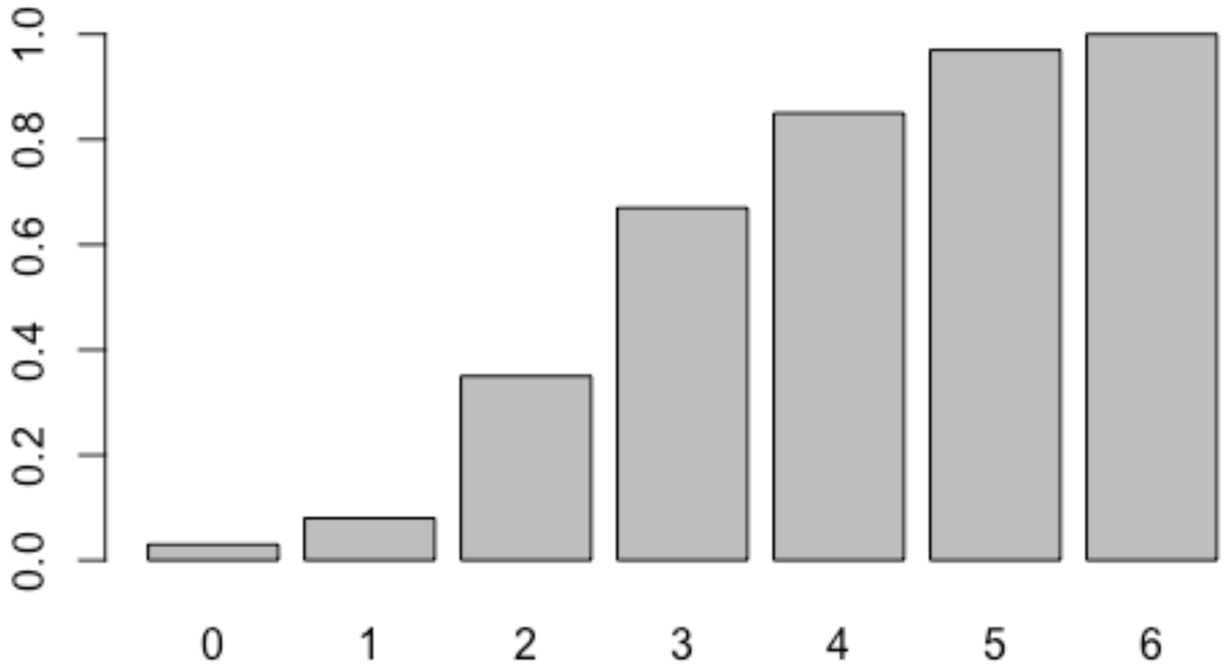
Reasoning About a Range of Event Outcomes

As we progress in making sense out of probabilities, it will help to be able to reason about ranges of event outcomes, rather than just individual events. For example, in the table below, what is the probability of either five or six jelly-down events?

Jelly-down count	0	1	2	3	4	5	6
Number of trials with that count	4	9	20	34	21	11	1
Probability of that count	0.04	0.09	0.2	0.34	0.21	0.11	0.01

Cumulative Probabilities

```
probTable <- table( rbinom(n=100, size=6, prob=0.5) )/100
barplot( cumsum(probTable) )
```



Demonstrate Using Contingency Tables to Reason About a More Complex Set of Events

Contingency Tables: Two Characteristics for Each Event

In our examples so far, each event has two mutually exclusive outcomes: jelly side up or jelly side down.

We can add a new twist by putting two different toppings on our toast: jelly or butter.

Now each event (a piece of toast falling down) has two classifications attached to it. Each piece of toast
may fall topping down or topping up, but we also now have two kinds of topping (jelly or butter).

These two characteristics are "fully crossed," which means that we have four possible outcomes for each event.

Contingency Tables: Linked Outcomes

Now each event (a piece of toast falling down) has two classifications attached to it. Each piece of toast may fall topping down or topping up, but we also now have two kinds of topping (jelly or butter). This table shows one possibility for what might happen with 10 events (10 dropped pieces of toast).

	Down	Up
Jelly	2	1
Butter	3	4

Compute and Interpret Marginal Totals in a Contingency Table

Learn More With Marginal Totals

What is the total number of “toast drop” events that have occurred in this trial?

How many times did the toast land with the topping down?

How many times did the toast land with the topping up?

Is it more likely that the toast lands with the topping down or up?

Across all of our dropped toast, which is more likely/popular—butter or jelly?

	Down	Up	Row Totals
Jelly	2	1	3
Butter	3	4	7
Column Totals	5	5	10

Does the Topping Matter?

For jelly, the ratio of down:up is 2:1 (2.0). For butter, the ratio of down:up is 3:4 (0.75).

So for this particular dataset, jelly toast is much more likely to fall topping side down than is butter toast.

Don't forget that this is just a single trial with only 10 events.

	Down	Up	Row Totals
Jelly	2	1	3
Butter	3	4	7
Column Totals	5	5	10

A Larger Contingency Table

Work Hours	Anne	Bill	Calvin	Devon	Marginal
Monday	6	0	5	8	19
Tuesday	0	6	6	0	12
Wednesday	8	5	0	8	21
Thursday	4	6	5	8	23
Friday	8	8	5	8	29
Saturday	8	10	8	8	34
Sunday	0	3	6	0	9
Marginal	34	38	35	40	147

Same Table Converted to Proportions

Two interesting questions:

1. You sit down in the restaurant and you notice that Bill is your server. What day is it most likely to be?
2. If you go to the restaurant today, knowing that it is Monday, who is most likely to be your server?

	Anne	Bill	Calvin	Devon	Marginal
Monday	0.0408	0.0000	0.0340	0.0544	0.1290
Tuesday	0.0000	0.0408	0.0408	0.0000	0.0820
Wednesday	0.0544	0.0340	0.0000	0.0544	0.1430
Thursday	0.0272	0.0408	0.0340	0.0544	0.1560
Friday	0.0544	0.0544	0.0340	0.0544	0.1970
Saturday	0.0544	0.0680	0.0544	0.0544	0.2310
Sunday	0.0000	0.0204	0.0408	0.0000	0.0610
Marginal	0.2310	0.2590	0.2380	0.2720	1

Isolate and Normalize a Row or Column of a Contingency Table to Reason About the Probabilities of Events When New Information Is Learned

Introducing New Information: Bill Is Your Server

You sit down in the restaurant and you notice that Bill is your server. What day is it most likely to be?

	Anne	Bill	Calvin	Devon	Marginal
Monday	0.0408	0.0000	0.0340	0.0544	0.1290
Tuesday	0.0000	0.0408	0.0408	0.0000	0.0820
Wednesday	0.0544	0.0340	0.0000	0.0544	0.1430
Thursday	0.0272	0.0408	0.0340	0.0544	0.1560
Friday	0.0544	0.0544	0.0340	0.0544	0.1970
Saturday	0.0544	0.0680	0.0544	0.0544	0.2310
Sunday	0.0000	0.0204	0.0408	0.0000	0.0610
Marginal	0.2310	0.2590	0.2380	0.2720	1

Isolating and Normalizing a Column

Saturday is the most likely day, given the knowledge that Bill is our server. The probability is 0.2632.

Bill is our server: What day is it?	Bill's raw work probabilities	Bill's normalized work probabilities
Monday	0.0000	0.0000
Tuesday	0.0408	0.1579
Wednesday	0.0340	0.1316
Thursday	0.0408	0.1579
Friday	0.0544	0.2105
Saturday	0.0680	0.2632
Sunday	0.0204	0.0789
Marginal	0.2590	1

Introducing New Information: It’s Monday

If you go to the restaurant today, knowing that it is Monday, who is most likely to be your server?

	Anne	Bill	Calvin	Devon	Marginal
Monday	0.0408	0.0000	0.0340	0.0544	0.1290
Normalized probabilities for Monday	0.3163	0.0000	0.2636	0.4217	1

If you know that it is Monday, you are most likely to have Devon as a server. The probability is 0.4217.

Connect These Concepts With Bayesian Reasoning About Prior and Posterior Probability

Bayesian Thinking

Thomas Bayes, the 19th-century mathematician, documented the idea that we can update our beliefs about a set of outcomes when presented with new data.

In the two examples above, we began with a **prior** understanding of the weekly work hours and probabilities of four servers. For any combination of server and day, we could pick the raw probability out of the contingency table.

But when we updated our scenario with new knowledge—either about the server we got or what day it was—we suddenly had a new and more refined understanding of the **posterior** probabilities.

Prior and posterior probabilities are the heart of Bayesian thinking.