Introduction

- Up to now, we have focused on data and the question: What is the message in this data?
- We also need to address questions like:
  1. If this data is a sample from a larger population, how well does the "message in this data" apply to the parent population?
  2. How to account for the "luck of the draw"?

These questions involve the following basic question:

- If this data comes from a scientific experiment, will the "message in this data" be the same if the experiment is repeated and different numbers come out? How to account for other factors influencing the outcome of the experiment?

Answer: Probability Modeling

- As the next 3 pages show, everyone talks about "luck" and "chance". It is not a new idea.
- Our goal is to see how to use "probability" as a tool.
HEISENBERG
DEPT. OF PHYSICS

YOU ARE
PROBABLY
HERE?

chas
"I've got two chances at it this year."
"I tend to agree with you—especially since $6 \cdot 10^{-9} \sqrt{c}$ is my lucky number."
"Probability modeling" had its origin in the context of games of chance. Gamblers came to mathematicians with questions about the games they were playing. This began seriously in the 17th century.

Our goal here is not to become better at games involving chance, like poker, or roulette, or monopoly, etc. (although it is ok if this happens as a byproduct of our study of probability).

Rather, our goal is to become better at other types of "gamble": interpretation of data, decision-making using data. In other words, statistical inference based on chance data is a kind of gamble, and we use statistical science to guide us. Probability tools are used for such purposes.

For example, if a blood test for a certain disease suggests that we have the disease (but not conclusively), what is the probability that we do have the disease? Or, if the blood test suggests that we do not have the disease, what is the probability that we really do not have it? Given the result of the test, a patient needs to use one of these probabilities in order to make an informed decision. Also relevant are probabilities such as the probability that the test detects the disease when it is present.

We want to be able to use "probability" as a tool.
Some basic concepts about "chance" (probability)

The chance of something gives the percentage of time it is expected to happen, when a given process or experiment is carried out repeatedly and independently and under the same conditions.

Chances are between 0% and 100%. Equivalently, probabilities are between 0 and 1.

The chance of something equals 100% minus the chance of the opposite thing.

Example: Tossing a "fair" coin

(i) In 100 tosses, about how many heads are expected?
   (A) 100  (B) 50  (C) 1

(ii) For any toss, what is the "chance" of Heads?
   (A) 1  (B) 1/2  (C) 0

Q. What does "probability" model?
   (A) A physical property inherent in the "experiment"
   (B) Our uncertainty about the outcome of the experiment
   (C) Both (A) and (B)
Example: Drawing tickets out of a box

\[\begin{array}{c}
1 \\
2 \\
3 \\
4 \\
5
\end{array}\] ← box with 5 tickets

When the drawing is done at random, each ticket in the box has equal chance to be selected.

With Replacement

- Each selection is returned to the box and can be selected again

Without Replacement

- A selected item is not returned to the box.

Example:

- 1st selection: 3
- 2nd selection: 1
- 3rd selection: 4
- 4th selection: 3 again

Example:

- 1st selection: 3
- 2nd selection: 1
- 3rd selection: 4
- 4th selection: 5
- 5th selection: 2

- What is the chance of getting 2 for 1st selection?
  - (A) 20%  (B) 50%  (C) 100%

- What is the chance of getting 2 for 2nd selection?
  - (A) 20%  (B) 50%  (C) 100%

- With replacement: the 2nd selection follows same chances as 1st
- Without replacement: Each ticket has equal chance of being the one selected in the 2nd drawing.
Conditional Probabilities

Example: Drawing without replacement from

\[
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
\end{array}
\]

a. If the 1st selection is 3, what is the chance of 2 for the 2nd selection?

(A) 20%  (B) 50%  (C) 100%  (D) 25%

Solution: If we have that 1st is 3, then the experiment for the 2nd becomes a new box:

\[
\begin{array}{ccccc}
1 & 2 & 4 & 5 \\
\end{array}
\]

Each choice has \(\frac{1}{3}\%\).

Terminology:

The conditional probability that the 2nd selection is 2, given that the 1st is 3, is \(\frac{1}{3}\)%.

Useful notation:

\[
P(\text{2nd selection is 2} | \text{1st selection is 3}) = \frac{1}{3}\%
\]

Recall the 2-coin-toss demonstration – Conditional probability is used to change the probability model for our uncertainty about an outcome, when some additional information is acquired that changes our uncertainty.
The Multiplication Rule

Question: Suppose that there are two chance events of interest. What is the chance that they both occur?

Example. Drawing without replacement from

\[ \{1, 2, 3, 4, 5\} \]

What is the chance of getting 2 on 1st draw and 4 on the 2nd?

Solution:

(i) Getting 2 on 1st draw has chance 20%, or probability \( \frac{1}{5} \).

(ii) Given 2 on 1st draw, the experiment now becomes a selection from

\[ \{1, 3, 4, 5\} \]

and the chance of getting 4 is 25%, or probability \( \frac{1}{4} \).

(iii) We multiply these together to get the answer:

\[ \frac{1}{5} \times \frac{1}{4} = \frac{1}{20}. \]

Multiplication Rule. The chance that two things will both happen equals the chance that one of them happens multiplied by the conditional chance that the other will happen, given that the first has happened.
INDEPENDENCE

Two possible chance events are independent if the conditional chance of the second given the first is not different from the chance of the second not given information about the first. Otherwise, the two chance events are dependent.

Example. Drawing two items from

```
1  2  3  4  5
```

Consider the two chance events

- 5 on 1st draw
- 3 on 2nd draw.

(i) Drawing with replacement

When drawing at random with replacement the draws are all independent.

After 5 on 1st draw, the experiment for the 2nd draw is unchanged. So the chance of 3 on 2nd draw is the same as in the original experiment: 3/5. Then these chance events are independent.

(ii) Drawing without replacement

When drawing at random without replacement the draws are dependent.

After 5 on 1st draw, the experiment for the 2nd draw becomes 4 2 3 4, and the chance of 3 is 4/4. Then these chance events are dependent.
Example. The 148 students in a certain class are grouped by gender and grade level as follows:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Senior</td>
<td>42</td>
<td>48</td>
</tr>
<tr>
<td>Junior</td>
<td>38</td>
<td>20</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td></td>
<td>148</td>
</tr>
</tbody>
</table>

A student is selected at random.

(i) The chance of getting a male is
(A) 80%  
(B) \( \frac{80}{148} \)  
(C) \( \frac{1}{148} \)

(ii) The chance of getting a junior is
(A) 58%  
(B) \( \frac{58}{148} \)  
(C) \( \frac{1}{148} \)

(iii) Given that the selected student is a junior, the chance that the student is male is
(A) \( \frac{38}{58} \)  
(B) \( \frac{80}{148} \)  
(C) \( \frac{1}{148} \)

(iv) The chance events "male" and "junior" are
(A) independent  
(B) dependent

(v) The probability that the student is a male junior is
(A) \( \frac{58}{148} \times \frac{38}{148} \)  
(B) \( \frac{1}{4} \)  
(C) 0

(vi) The probability that the student is a female sophomore is
(A) \( \frac{90}{148} \times \frac{48}{148} \)  
(B) \( \frac{1}{4} \)  
(C) 0
Listing the Ways

It is sometimes helpful to make a list of all possible outcomes in an "equally-likely" model. Then one can obtain the chance of an event by counting up the relevant outcomes and dividing by the total.

Example. Two random selections from

\[
\begin{array}{ccccc}
1 & 2 & 3 & 4 & 5 \\
\end{array}
\]

drawing with replacement.

Q. What is the probability of getting 1 at least once?

Solution.

(i) List all possibilities.

\[
\begin{array}{cccccc}
1,1 & 2,1 & 3,1 & 4,1 & 5,1 \\
1,2 & 2,2 & 3,2 & 4,2 & 5,2 \\
1,3 & 2,3 & 3,3 & 4,3 & 5,3 \\
1,4 & 2,4 & 3,4 & 4,4 & 5,4 \\
1,5 & 2,5 & 3,5 & 4,5 & 5,5 \\
\end{array}
\]

(25 in all)

(ii) Count how many have 1 at least once.

\[
\begin{array}{cccccc}
(9 \text{ in all}) \\
1,1 & 2,1 & 3,1 & 4,1 & 5,1 \\
1,2 & 2,2 & 3,2 & 4,2 & 5,2 \\
1,3 & 2,3 & 3,3 & 4,3 & 5,3 \\
1,4 & 2,4 & 3,4 & 4,4 & 5,4 \\
1,5 & 2,5 & 3,5 & 4,5 & 5,5 \\
\end{array}
\]

(iii) Answer: \[\frac{9}{25}\]
The Addition Rule

Two chance events are mutually exclusive if they cannot both occur at the same time.

Example. Two random selections from 1 2 3 4 5, drawing with replacement. Outcomes (25 in all):

<table>
<thead>
<tr>
<th></th>
<th>1,1</th>
<th>2,1</th>
<th>3,1</th>
<th>4,1</th>
<th>5,1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,2</td>
<td>2,2</td>
<td>3,2</td>
<td>4,2</td>
<td>5,2</td>
<td></td>
</tr>
<tr>
<td>1,3</td>
<td>2,3</td>
<td>3,3</td>
<td>4,3</td>
<td>5,3</td>
<td></td>
</tr>
<tr>
<td>1,4</td>
<td>2,4</td>
<td>3,4</td>
<td>4,4</td>
<td>5,4</td>
<td></td>
</tr>
<tr>
<td>1,5</td>
<td>2,5</td>
<td>3,5</td>
<td>4,5</td>
<td>5,5</td>
<td></td>
</tr>
</tbody>
</table>

Consider 3 chance events:

A = "getting 1 at least once"
B = "getting (2,1) or (2,2) or (2,3) or (2,4) or (2,5)"
C = "getting (2,2) or (2,3) or (2,4) or (2,5)"

Then:
- A and B are mutually exclusive True False
- A and C are mutually exclusive True False
- B and C are mutually exclusive True False

Addition Rule. For mutually exclusive chance events, the probability of at least one occurring is the sum of their individual probabilities.

Example, continued

\[ P(\text{either } A \text{ or } C \text{ or both}) = P(A) + P(C) = \frac{9}{25} + \frac{4}{25} = \frac{13}{25}. \]