Probability (part 1)

Ryan Miller

Outline

1. Sources of uncertainty
2. Definitions
3. Unions and intersections

Introduction

In high-quality research, it's common to encounter *randomness* in two areas:

1) *Random sampling* helps ensure that our sample data are representative of the population we're studying
2) *Random assignment* protects us against confounding variables

An important consequence of randomness is that the trends in our sample data might not identically mirror those within the broader population

Example

- The Sampling Distribution for a Mean section of StatKey comes pre-loaded with data on the contracts of all 877 MLB players during the 2019 season
- If we only wanted to draw conclusions about this season, we might consider these data to be a population
- When using a *random sample* to study this population, notice how the *sample mean* is unlikely to be identical to the population's mean due to *random chance*

Probability Definitions

- A **trial** is a single instance of a **random process**
  - Every trial results in an **outcome**

Probability Definitions

- A **trial** is a single instance of a **random process**
  - Every trial results in an **outcome**
- For example, consider the selection of one member of the Xavier faculty (population) into a sample
  - This selection represents one trial of a random process
  - A possible outcome would be “Ryan Miller”
  - Note that we could measure things differently, so another outcome might be “Math Department”

Probability Definitions

- A **trial** is a single instance of a **random process**
  - Every trial results in an **outcome**
- For example, consider the selection of one member of the Xavier faculty (population) into a sample
  - This selection represents one trial of a random process
  - A possible outcome would be "Ryan Miller"
  - Note that we could measure things differently, so another outcome might be "Math Department"
- The collection of *all possible outcomes* is called the **sample space**
  - For example, the sample space of Xavier faculty would be a list of hundreds of names
  - In the special case of random sampling, each outcome in the sample space is *equally likely*

Probability Definitions

- Statisticians generally to focus on **events**, which are *combinations of one or more observed outcomes*
- Below are a couple examples of events for a single trial in the Xavier faculty example:
  - The faculty member is younger than 40 and teaches math
  - The faculty member teaches math or teaches computer science

Probability Definitions

- **Probability** is way of quantifying the likelihood of observing a particular event
  - To begin, consider a single coin flip. Everyone agrees the probability of a fair coin landing "heads" is 1/2, but why?

Probability Definitions

- **Probability** is way of quantifying the likelihood of observing a particular event
  - To begin, consider a single coin flip. Everyone agrees the probability of a fair coin landing "heads" is $1/2$, but why?
- **Frequentist** statisticians define probability as the *long-run proportion of an event occurring*
  - Thus, $P(\text{Heads}) = 0.5$ means that if we conducted many *trials* (different coin flips) we'd expect the *event* "Heads" to be observed in half of them
  - Alternatively, this random process has two outcomes that are equally likely, so $1/2 = 0.5$ must be the probability of "Heads"

Valid Probabilities

- Probability is a long-run proportion, so the probability of any event *must* be between 0 and 1
  - Please do not ever report a probability outside of this range on an exam or homework assignment (if necessary, leave your incorrect work and don't provide a final answer)

Empirical Probabilities

- Because probabilities are just long-run relative frequencies (proportions), it's possible to estimate them using data
  - For example, Steph Curry's career free through percentage is approximately 90%, so we might estimate $P(\text{Make}) = 0.9$ for his next attempt

Empirical Probabilities

- Because probabilities are just long-run relative frequencies (proportions), it's possible to estimate them using data
  - For example, Steph Curry's career free through percentage is approximately 90%, so we might estimate $P(\text{Make}) = 0.9$ for his next attempt
- This is called an *empirical probability*, it is different from a *theoretical probability* like $P(\text{Heads}) = 0.5$
  - Empirical probabilities are *estimated* using a finite amount of data
  - Theoretical probabilities are *governed* by the nature of the random process

Example #1

According to the US Census, the current racial composition of the United States is 61.5% Non-Hispanic White, 17.6% Hispanic (of any race), 12.3% Black, 5.3% Asian, 0.7% Native American, and 2.6% other races. For the questions that follow, consider the race of a *randomly selected individual* from this population.

1) In words, briefly describe what a *trial* refers to in this example.
2) What is the *sample space* of the trial you described in #1?
3) What is $P(\text{Asian})$?

Example #1 (solutions)

1) A trial is the random selection of an individual
2) The space is the collection of all possible racial/ethnic categories (Non-Hispanic White, Hispanic, Black, Asian, Native American, and Other)
3) $P(\text{Asian}) = 0.053$

Unions and Intersections

- As previously mentioned, an *event* is a combinations of *one or more outcomes*
  - There are *two ways* to combine multiple outcomes, *unions* and *intersections*

Unions and Intersections

- As previously mentioned, an *event* is a combinations of *one or more outcomes*
  - There are *two ways* to combine multiple outcomes, *unions* and *intersections*
- **Intersection** refers to two (or more) outcomes simultaneously occurring
  - Intersections are expressed using "and" or the symbol $\cap$
  - Consider rolling a six-sided die, $P(\text{Five and Six}) = P(\text{Five} \cap \text{Six}) = 0$
  - Alternatively, $P(\text{Five and Odd Number}) = P(\text{Five} \cap \text{Odd Number}) = 1/6$

Unions and Intersections

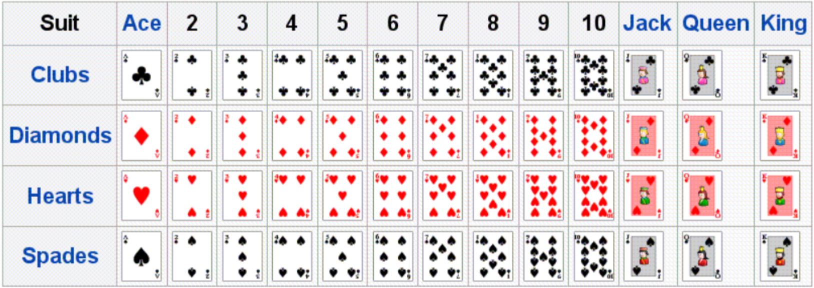
- As previously mentioned, an *event* is a combinations of *one or more outcomes*
  - There are *two ways* to combine multiple outcomes, *unions* and *intersections*
- **Intersection** refers to two (or more) outcomes simultaneously occurring
  - Intersections are expressed using "and" or the symbol $\cap$
  - Consider rolling a six-sided die, $P(\text{Five and Six}) = P(\text{Five} \cap \text{Six}) = 0$
  - Alternatively, $P(\text{Five and Odd Number}) = P(\text{Five} \cap \text{Odd Number}) = 1/6$
- **Union** refers to *at least one* of the specified outcomes occurring
  - Union are expressed using "or" or the symbol $\cup$
  - Consider rolling a six-sided die, $P(\text{Five or Six}) = P(\text{Five} \cup \text{Six}) = 2/6 = 1/3$
  - Alternatively, $P(\text{Five or Odd Number}) = P(\text{Five} \cup \text{Odd Number}) = 3/6 = 1/2$

Sample Spaces, Unions, and Intersections

- The probability of the *union of all outcomes* in a sample space is 1
  - If we flip a single coin: $P(\text{Heads or Tails}) = 1$
  - If we randomly sample letter grades on an exam: $P(\text{A or B or C or D or F}) = 1$

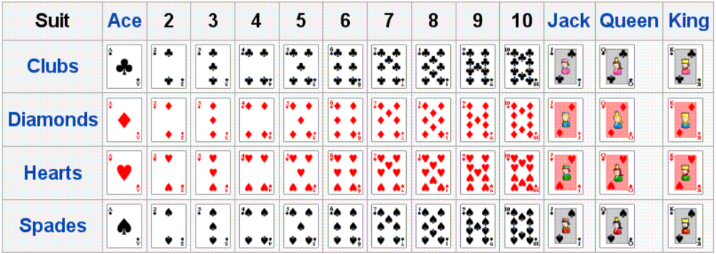
Example #2

When all outcomes are equally likely, a strategy for finding the probability of a union or an intersection is to write out the entire sample space and count the outcomes that satisfy the event of interest. We'll do an example using a standard deck of 52 cards:



Example #2 (cont)

For the questions that follow, consider a randomly selected card from a standard deck:



1) What is $P(\text{Heart})$?
2) What is $P(\text{Heart} \cap \text{Queen})$?
3) What is $P(\text{Heart} \cup \text{Queen})$?

Example #2

To begin, realize that all 52 cards in the sample space are equally likely.

1) Since $1/4$ of the cards are hearts, $P(\text{Heart}) = 0.25$
2) Since only 1 of the 52 cards is the queen of hearts, $P(\text{Heart} \cap \text{Queen}) = 1/52 = 0.019$
3) There are 13 hearts (including the queen of hearts) and an additional 3 questions, so the probability is $16/52 = 0.308$

Conclusion

- This presentation introduced basic definitions and concepts related to probability
  - random process -> trial -> outcome -> event -> probability -> union or intersection
- The next presentation will cover more advanced probability, focusing on rules used to calculate the probabilities of my complicated events