Probability I

Day 6 AM

Sarah Moore & J. Seawright

Northwestern University

Math Camp 2023

Day 6 Agenda

- What are probability and statistics?
- Concepts in probability
- Random variables-- discrete and continuous

Probability

- Probability theory is foundational to quantitative social science, as it is a means to derive the uncertainty of an outcome given a set of potential outcomes.
- Most of your coursework will cover frequentist statistics, which is based on the notion of objective probability. Other frameworks, like rational choice theory and Bayesian statistics, incorporate subjective probability.

Probability Concepts

Outcome

- The result of an experiment.

Example: What are the possible outcomes when rolling a 6-sided die?

Probability Concepts

Event

A specified outcome or subset of outcomes.

Example: Define the number of outcomes of the event that a 6-sided die lands on any number $\leq \mathbf{2}$.

$\rightarrow$ The defined event has 2 possible outcomes.

Probability Concepts

Sample Space

Ω: A list of ALL possible outcomes of an event.

What is the sample space of rolling a 6-sided die?

$\rightarrow \Omega = \{1, 2, 3, 4, 5, 6\}$

Notation

Probability and statistics will rely on some of the set notation we talked about in Day 1.

$\cup$ Union

$\cap$ Intersection

A^c or A' Complement

An important note is that we referenced the OR operator in R as |. However when you see this operator in probability, it is meant to indicate conditionality. For example, $Pr(A|B)$ is the probability of A given B.

Calculating Probability

It's likely you have learned the following formula to calculate simple probability:

$$Pr(\text{event}) = \frac{\#\text{ outcomes in event}}{\#\text{ total outcomes in sample space}}$$

Calculating Probability

$$Pr(\text{event}) = \frac{\#\text{ outcomes in event}}{\#\text{ total outcomes in sample space}}$$

For example, in the case of the 6-sided die, what is the simple probability that the die lands on a number ≤ 2?

First, we can define the sample space as before, where $S = \{1, 2, 3, 4, 5, 6\}$. There are 6 total outcomes in the sample space.

Then, we can determine the number of pertinent outcomes are in the event, here it is 2, such that $E = 1, 2$.

Therefore, we are interested in the $Pr(\text{Die} = 1, 2)$. This is 2 outcomes out of 6, which simplifies to $\frac{1}{3}$.

BUT

Independence, mutual exclusivity, and collective exhaustiveness are all important properties to ensure you are calculating probability correctly.

Probability Axioms and some SOP

Suppose we are interested in event A. The probability of event is denoted $Pr(A)$.

- Probability will always lie in the range $[0, 1]$, where 0 indicates that A cannot happen (i.e., is improbable) and 1 indicates the event will deterministically happen.

- Furthermore, the *law of total probability* states that the total probability of a sample space is 1. Therefore, when we sum together probabilities of all events in a sample space, the total probability equals 1.

Probability Axioms and some SOP

- If $Pr(A)$ is notated alone, we can also assume that we are calculating the probability *independently* of any other event OR that we have already considered any other conditional event.
- We are usually interested in the probability of a concurrence of events, that is, one or more events happening together, but not necessarily conditionally. We can calculate the joint probability. There are different types of joint probability, based on if we want to know the intersection ∩ or the union ∪ of events.
- We can also account for the conditional probability of A given some other event B. This is annotated $Pr(A|B)$ and read as the probability of A given or conditional on B, or probability of A given B.

Independence

Two or more events are independent if the probability that any one of them may occur does not depend on the occurrence of another.

Example: The probability that it rains today is independent from the probability that it is anyone's birthday in this room.

Formally, statistical independence between two events, A and B, is given by:

1. $P(A|B) = P(A)$
2. $P(B|A) = P(B)$
3. $P(A \cap B) = P(A)P(B)$

This can be generalized to all events $A_1, \ldots, A_k$ that are independent: $P(\cap_{i=1}^{k} A_i) = \prod_{i=1}^{K} P(A_i)$.

Mutual Exclusivity

Two or more events are mutually exclusive if they do not coincide. You may also hear that two events are mutually exclusive if the events are disjoint.

Example: The event that you were born in a given country is mutually exclusive of the event that you were born in another country. You cannot have been born in two countries.

Therefore, the probability of two mutually exclusive events, A and B is 0:

$$P(A \cap B) = 0$$

Collectively Exhaustive

All events cover the probability sample space, in other words, at least one of them is going to happen.

Example: When rolling a 6-sided die, the events are collectively exhaustive as the probability that any of the events happens is 1. We know that a roll guarantees that the die will land on one of the numbers.

Events are thus collectively exhaustive if the union of all possible event outcomes covers the sample space:

$$A \cup B = S$$

https://www.statisticshowto.com/collectively-exhaustive/

Joint Probability

Sometimes we want to know the probability of the concurrence of events.

In the case that two concurrent events are independent, calculating the probability is simple.

The joint probability of independent events is merely the product of the probabilities:

$$P(A \cap B) = P(A)P(B)$$

Joint Probability, Example

Example: Let's say you are drawing 2 widgets from 2 different urns.

Each of the urns has 2 red widgets, 2 white widgets, and 4 black widgets.

Assuming that the draws from each urn are independent of one another, what is the probability that you choose a black ball from each of the urns?

$$P(B_{\text{Urn 1}}) = \frac{4}{8}; P(B_{\text{Urn 2}}) = \frac{4}{8}$$

$$P(B_{\text{Urn 1}} \cap B_{\text{Urn 2}}) = P(B_{\text{Urn 1}}) \times P(B_{\text{Urn 2}}) = \frac{4}{8} \times \frac{4}{8} = \frac{1}{4}$$

Joint Probability, cont'd.

In other cases, we might want to know about the concurrence of events that are mutually exclusive and independent. This is the case that we want to know about Event A *or* Event B happening. The joint probability of mutually exclusive events is the sum of the probabilities:

$$P(A \cup B) = P(A) + P(B)$$

Example: Let's say you want to calculate the probability that a randomly chosen adult in Chicago voted for Trump or Biden in the last election. Let's say that the probabilities are respectively P(Trump)= 0.35, P(Biden)=0.60. What is the probability that a randomly chosen person voted for Trump OR Biden?

$$P(\text{Trump}) = 0.35, P(\text{Biden}) = 0.60$$

$$P(\text{Trump} \cup \text{Biden}) = 0.35 + 0.60 = 0.95$$

Why are these events mutually exclusive? Are they collectively exhaustive?

Joint Probability, cont'd.

For events that are independent but NOT mutually exclusive, we still want to take into account the union of the probabilities we need to somehow account for the overlap of the events.

In this case, the joint probability is the sum of the probabilities minus their intersection:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Joint Probability, Example II

Returning to the example of the urns, let's say you are drawing 2 widgets from 2 different urns.

Each of the urns has 2 red widgets, 2 white widgets, and 4 black widgets What is the probability that you draw a black ball from urn 1 OR urn 2?

$$P(B_{\text{Urn 1}}) = \frac{4}{8}; P(B_{\text{Urn 2}}) = \frac{4}{8}$$

$$P(B_{\text{Urn 1}} \cup B_{\text{Urn 2}}) = P(B_{\text{Urn 1}}) + P(B_{\text{Urn 2}}) - P(B_{\text{Urn 1}} \cap B_{\text{Urn 2}}) = \frac{4}{8} + \frac{4}{8} - \frac{1}{4} = \frac{3}{4}$$

Conditional Probability

- Conditional probability allows for us to consider the likelihood of one event given another.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Conditional Probability

Consider the following frequency table:

```
##      b
##  a    1  2
##     1  8  9
##     2  2  6
```

1. $P(A)$
2. $P(B)$
3. $P(A \cap B)$
4. $P(A|B) = \frac{P(A \cap B)}{P(B)}$
5. $P(B|A) = \frac{P(B \cap A)}{P(A)}$

Bayes Rule

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} =$$

$$= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A')P(A')}$$

Counting Methods

Permutations

How many ways are there to arrange a set of objects?

For example, the

- Sampling without replacement: permutations of n elements taken k at a time:

$$ {}_nP_k = n(n-1)\dots(n-k+1) = \frac{n!}{(n-k)!} $$

Where ! indicates a factorial.

Permutations

- Number of different arrangements, order is important.

$$ {}_nP_k = n(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!} $$

Combinations

- Unordered selection of objects.

$$ {}_nC_k = \binom{n}{k} \frac{{}_nP_k}{k!} = \frac{n!}{k!(n-k)!} $$

Suppose we are filling a committee of 6 total people, 3 graduate students and 3 faculty members. There are 5 graduate students and 3 faculty members to choose from.

How many ways are there to choose the graduate students to fill their spots?

$$ {}_5C_3 = \frac{5!}{3!(5-3)!} = 10 $$

How many ways are there to choose the faculty to fill their spots?

$$ {}_3C_3 = \frac{3!}{3!(3-3)!} = 1 $$

How many total ways are there to choose the committee?

$$ {}_5C_3 \times_3 C_3 = 10 $$

Combinations

Closed list party-list systems entail that people vote for parties, and the parties themselves pick the representatives that will represent them in government.

Suppose a party won 4 seats in an election and the party leadership want to know how many different combinations there are of choosing 4 representatives out of a list of 15 party members.

```
choose(15, 4)
```

```
## [1] 1365
```

Combinations

Let's say the party leadership wanted to ensure that there are no fewer than 2 women among the representatives. If there are 6 party members that identify as women, how many possible combinations of *women representatives* are there?

> - Note that there can be more than 2 women representatives.

```
choose(6, 2) + choose(6, 3) + choose(6, 4)
```

```
## [1] 50
```

Must account for:

- All combinations of two women.
- All combinations of three women.
- All combinations of four women.