Maths Day Three: Probability

Princeton Sociology Methods Camp

Varun Satish[1]

[1]These slides are the collective effort of everyone who has contributed to the Princeton Sociology Methods Camp.

Outline

1. Why do we need probability?
2. Sets and operations on sets
   - Subsets, unions, intersections, complements
3. Experiments, events, and sample spaces
4. An introduction to probability
5. The law of total probability, Bayes' rule, and independence

Why do we need probability?

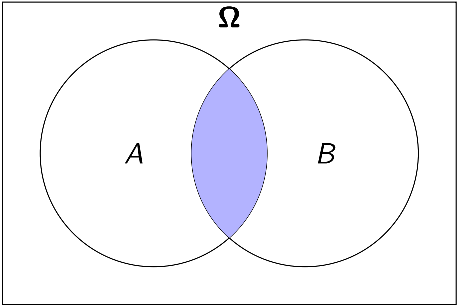
- Social processes are not deterministic
- The "effects" of social causes are difficult to isolate and estimate
- We need a framework for communicating our uncertainty about the inferences we draw in our empirical work
- **Probability theory is the root of social statistics**
- Being able to think rigorously about uncertainty is also helpful for case selection and small-n inference in qualitative research

Set and Operations

- **Set:** A collection of **elements** usually denoted by curly braces
- **Subset**: A set that is composed entirely of elements of another set (e.g. Set A is subset of Set B if every element of A was also an element of B – Set B *contains* Set A)
- **Union**: The union of two sets contains all the elements that belong to either sets (in programming, the OR operation)
- **Intersection**: The intersection of two sets contains only those elements found in both sets (in programming, the AND operation)
- **Complement**: The complement of a given set is the set that contains all elements not in the original set

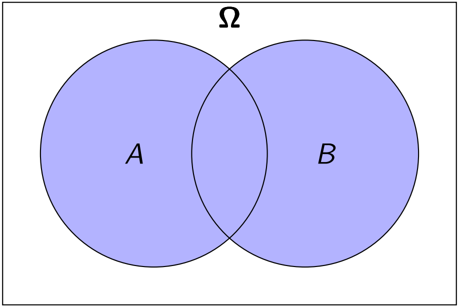
Intersection

Intersection: The intersection of two sets contains only those elements found in both sets ($\{A \cap B = \{X | X \in A \textbf{ and } X \in B\}$)



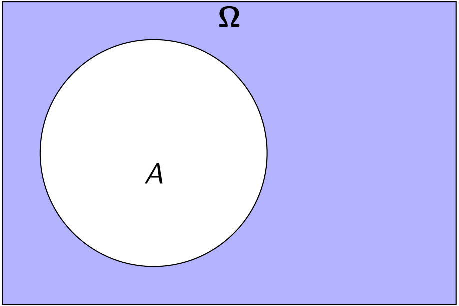
Union

Union: The union of two sets contains all the elements that belong to either sets ($A \cup B = \{X | X \in A \textbf{ or } X \in B\}$)



Complement

Complement: The complement of a given set is the set that contains all elements not in the original set ($A^C = \{X \in \Omega | X \notin A\}$)



Set and Operations

Characteristics of multiple sets

- **Disjoint**: Two sets are disjoint when their intersection is empty (*note*: complements are by definition disjoint)
- **Mutually Exclusive**: When k sets are all pairwise disjoint with each other

Let's check our understanding

Answer the following in your worksheet...

Consider the universal set $U = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. Given the sets $A = \{2, 4, 6, 8, 10\}$ and $B = \{3, 6, 9\}$, answer the following questions:

- Find the complement of set A.
- Find the complement of set B.
- Calculate the intersection of sets A and B.
- Find the union of the complements of sets A and B (i.e. $A^C \cup B^C$).
- Determine the intersection of the complement of set A and set B (i.e. $A^C \cap B$)

Experiments, events and sample spaces

- An **experiment** is an activity with an observable result (e.g. tossing a fair coin once)
- An **outcome** is the result of an experiment
- The **sample space** (sometimes denoted Ω) is the set of all outcomes of an experiment (e.g. $\{H, T\}$ – how many elements are in this set?)
- An **event** is a subset of the sample space (e.g. $\{\mathrm{H}\}$)

Let's check our understanding

Please go to your worksheet – suppose the experiment is tossing two coins in a row. What is the sample space (i.e. the set of possible outcomes)? What is an event that you might be interested in?

Basic Probability

What does probability do?

- Probability gives us a expression of how likely an event is
- In the case of a "simple" experiment, we can do this pretty straight forwardly by counting how many element of the sample space correspond with the event we are interested in
- For example, if we were interested in like probability of getting one head and one tail in an experiment where we are tossing a fair coin twice....
- $\Omega = \big\{\{H, H\}, \{H, T\}, \{T, H\}, \{T, T\}\big\}$
- How many of the possible outcomes correspond with tossing one head and one tail?

What does probability do?

- In reality the world is a lot more complicated and it is often hard (or impossible) to count the number of possible outcomes
- For example, what if we were interested in how much income someone earns per year?
- This motivates a more technical definition of something called a probability function which gives us the math to work with probabilities
- **You don't need to commit these to memory**
- But it is important to know that probability isn't just magic. We can build a lot of probability theory off the following three axioms (called Kolmogorov's axioms)

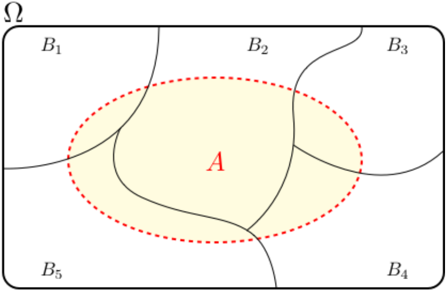
Axioms for a Probability Function

A probability function maps defined event(s) onto a metric in the interval of [0:1]. It enables us to discuss various *degrees of likelihood of occurrence* in a systematic way. It satisfies three axioms:

1. **Nonnegativity**: for any event A, $P(A) \geq 0$
2. **Normalization**: $P(\mathbf{S}) = 1$
3. **Additivity**: the probability of unions of n mutually exclusive events is the sum of their individual probabilities: $P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

Law of Total Probability

These axioms give us the math to
The probability of event A can be decomposed into parts: one part that intersects with B_1, another part that intersects with B_2, and so on.



We can also just think of it as only two parts: B_1 and everything else in the sample space that isn't B_1, or just B and B^C.

Law of Total Probability

Formally:

$$P(A) = P(A \cap B) + P(A \cap B^C)$$

Colloquially:

If break up any event into mutually exclusive parts, then the probability that the event occurs is the sum of the probabilities). For example, the probability of the event "tossing both a heads and a tails" is the sum of the probabilities of "tossing heads and then tails" and "tossing tails and then heads" since they are mutually exclusive.

Conditional Probability, Bayes' Rule, and Independence

Conditional Probability

We think about outcomes of experiments/processes differently if we have partial information about them versus no information at all. Conditional probability statements recognize that some prior information bears on the determination of subsequent probabilities.

For example: Incumbent senator Debbie Stabenow (D) is up for re-election next year in Michigan. Her probability of reelection can be expressed as $P(A)$. That probability depends on the winner of the GOP primary. One of the GOP candidates in the race so far is musician Kid Rock. Let's express the event of Kid Rock winning the GOP primary as B. So $P(A|B)$ would express the probability of Debbie Stabenow being reelected **conditional** on Kid Rock winning the GOP primary.

Conditional Probability

The probability of a Debbie Stabenow reelection given a Kid Rock victory can be calculated as the probability of Debbie Stabenow winning the reelection AND Kid Rock winning the primary divided by the probability that Kid Rock wins the primary:

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

which can be rearranged as:

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

Conditional Probability

We can also use conditional probability statements to express the Law of Total Probability:

$$P(A) = P(A|B) * P(B) + P(A|B^C) * P(B^C)$$

Bayes' Rule

Note that with conditional probability statements, order matters!
$P(A|B) \neq P(B|A)$

So what if we want to know $P(B|A)$ but we only have $P(A|B)$...

Bayes' Rule

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

This is Bayes' Rule.

Too bad we still don't know $P(B)$...

Bayes' Rule

But wait! The Law of Total Probability tells us that

$$P(B) = (P(B|A) * P(A)) + (P(B|A^C) * P(A^C))$$

So plugging that back into Bayes' Rule gives us:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B|A) * P(A) + (P(B|A^C) * P(A^C)}$$

Let's check our understanding

Suppose you work in a building that has a fire alarm system. The fire alarm is designed to go off when there is a fire, and it's also known that sometimes the alarm can go off due to smoke from a malfunctioning HVAC system. You want to calculate the probability that there is an actual fire when the fire alarm goes off

- There is a 1% chance that there is a fire: $P(Fire) = 0.01$
- The alarm system works pretty well and there is a 95% chance it goes off when there is an actual fire: $P(Alarm\ goes\ off | Fire) = 0.95$.
- But, there is a 10% chance that the alarm goes off due to smoke without a fire: $P(Alarm\ goes\ off | No\ Fire) = 0.1$

Now solve on worksheet – what's the probability of actually there being a dangerous fire given that there is smoke. In other words, find $P(Fire | Alarm\ goes\ off)$.

Independence

Sometimes we have information about the outcome of event A but it doesn't change the probability of event B happening (e.g. – knowing that today is Thursday does not help predict whether I will get heads or tails when I flip a coin). This is the intuitive description of **independence.**

Independence

Formally, A and B are independent if $P(A \cap B) = P(A) * P(B)$.

From that, we can also deduce that when A and B are independent:

$$P(A) = P(A|B)$$

and

$$P(B) = P(B|A)$$

Summing up

- We introduced methods of counting
- We introduced the concepts of sets and defined some concepts—subsets, unions, intersections, and complements—associated with them
- We walked through how to think of experiments, events, and sample spaces, which are ways to formalize events that have uncertain outcomes
- We showed how this framework is useful for defining probabilities
- We generalized these concepts and defined the law of total probability
- We further defined other key concepts such as Bayes' rule and independence