Probability 1: The Basic

Howard Liu

2024-09-25

Probability

What is a reasonably safe gathering size in the age of COVID?

Probability and assumptions

- If we have a gathering of size n drawn randomly from population of SC, what's the probability someone in attendance is infected?

Probability and assumptions

- If we have a gathering of size n drawn randomly from population of SC, what's the probability someone in attendance is infected?
  - Imagine current incidence rate of COVID in SC: 1.37%
  - Will we get 1.37% from our samples?

Probability and assumptions

- If we have a gathering of size n drawn randomly from population of SC, what's the probability someone in attendance is infected?
  - Imagine current incidence rate of COVID in SC: 1.37%
  - Will we get 1.37% from our samples?
- Calculating probabilities relies on assumptions
  - Gathering is a random sample from population
  - Covid outbreaks are equally likely across all counties
  - One county's outbreak likelihood is independent of one another

Probability and assumptions

- If we have a gathering of size n drawn randomly from population of SC, what's the probability someone in attendance is infected?
  - Imagine current incidence rate of COVID in SC: 1.37%
  - Will we get 1.37% from our samples?
- Calculating probabilities relies on assumptions
  - Gathering is a random sample from population
  - Covid outbreaks are equally likely across all counties
  - One county's outbreak likelihood is independent of one another
  - Wrong assumptions $\rightsquigarrow$ wrong probabilities
- Why do we care? Enables learning from data!
  - Are the data consistent with the probabilities in theory?
  - If not, casts doubt on our assumptions $\rightsquigarrow$ we've learned something!

Probability and assumptions

- If we have a gathering of size n drawn randomly from population of SC, what's the probability someone in attendance is infected?
  - Imagine current incidence rate of COVID in SC: 1.37%
  - Will we get 1.37% from our samples?
- Calculating probabilities relies on assumptions
  - Gathering is a random sample from population
  - Covid outbreaks are equally likely across all counties
  - One county's outbreak likelihood is independent of one another
  - Wrong assumptions $\rightsquigarrow$ wrong probabilities
- Why do we care? Enables learning from data!
  - Are the data consistent with the probabilities in theory?
  - If not, casts doubt on our assumptions $\rightsquigarrow$ we've learned something!
  - The heart of **statistical inference**

Conjunction fallacy

- Why mathematical probability?
  - Because our intuitions about probability are often **terrible**

Conjunction fallacy

- Why mathematical probability?
  - Because our intuitions about probability are often **terrible**

Example

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations

- What is more probable?
  1. Linda is a bank teller?
  2. Linda is a bank teller and is active in the feminist movement?

Conjunction fallacy

- Why mathematical probability?
  - Because our intuitions about probability are often **terrible**

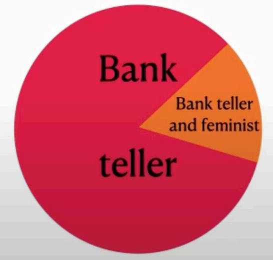
Example

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations

- What is more probable?
  1. Linda is a bank teller?
  2. Linda is a bank teller and is active in the feminist movement?

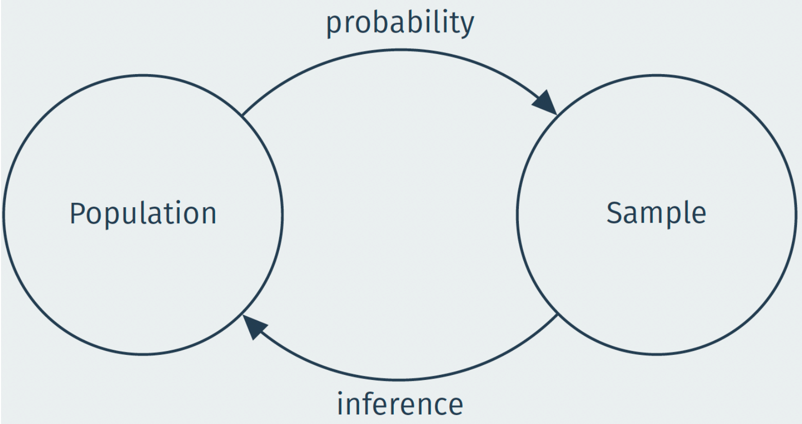
Conjunction fallacy

- Famous example of the **conjunction fallacy** called the Linda problem
  - Majority of respondents chose 2, but this is impossible!



- Learning mathematical probability avoids these mistakes!

Learning about populations



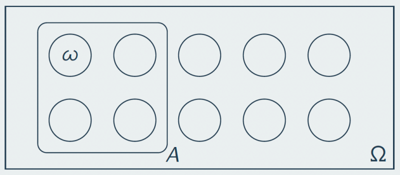
- **Probability**: formalize the uncertainty about how our data came to be.
- **Inference**: learning about the population from a set of data.

Roadmap

1. Naive Definition of Probability
2. Non-naive Definition of Probability

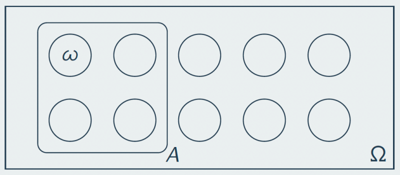
1. Naive Definition of Probability

Sample spaces & events



- Probability formalizes chance variation or uncertainty in outcomes.
  - It might rain or be sunny tomorrow, we don't know which.
  - To formalize, we need to define the set of possible outcomes.

Sample spaces & events



- Probability formalizes chance variation or uncertainty in outcomes.
  - It might rain or be sunny tomorrow, we don't know which.
  - To formalize, we need to define the set of possible outcomes.
- A sample space Ω is the set of possible outcomes.
  - Can be finite, countably infinite, or uncountably infinite.
- $\omega \in \Omega$ is one particular outcome.
- A subset of Ω is an event and we write this as $A \subset \Omega$

Naive definition of probability

- Naive definition of probability: **all outcomes equally likely**, so:

$$P(A) = \frac{number\ of\ elements\ in\ A}{number\ of\ elements\ in\ \Omega}$$

- Often wrong, why?

Naive definition of probability

- Naive definition of probability: **all outcomes equally likely**, so:

$$P(A) = \frac{number\ of\ elements\ in\ A}{number\ of\ elements\ in\ \Omega}$$

- Often wrong, why?
- Often wrong, but justified under a few assumptions:
  - **Symmetry**: fair coin, shuffled cards, dice, etc.
  - **Simple random sample** from a population.
- Example: `randomly select` a card from a standard deck of playing cards
  - Each of the 52 card has equal probability

$$P(10\clubsuit) = P(8\heartsuit) = \frac{1}{52}$$

Counting

- **Multiplication rule**: if you have two sub-experiments, A with a possible outcomes and B with b possible outcomes, then in the combined experiment there are ab possible outcomes.
- Example: what to watch and where to watch?
  - What to watch? Netflix or Hulu.
  - Where to watch? TV, tablet, or phone.

Counting

- **Multiplication rule**: if you have two sub-experiments, A with a possible outcomes and B with b possible outcomes, then in the combined experiment there are ab possible outcomes.
- Example: what to watch and where to watch?
  - What to watch? Netflix or Hulu.
  - Where to watch? TV, tablet, or phone.
  - $2 \times 3 = 6$ possible outcomes (Netflix on TV, Hulu on phone, ...)
- Assumes number of outcomes in one experiment *does not* depend on the outcome of the other experiment $\rightarrow$ big assumption, often wrong in real world

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there?

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4
  - What if you can choose k $= 2$ objects. How many possible outcomes are there?

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4
  - What if you can choose k $= 2$ objects. How many possible outcomes are there? Ans:4x4

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4
  - What if you can choose k $= 2$ objects. How many possible outcomes are there? Ans:4x4
  - There are n^k possible outcomes when order matters (multiplication rule)
- **Sampling without replacement**: Choose k objects from a set of n one at a time without replacement.
  - Chosen object can't be chosen again.
  - Number of possible outcomes?

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4
  - What if you can choose k $= 2$ objects. How many possible outcomes are there? Ans:4x4
  - There are n^k possible outcomes when order matters (multiplication rule)
- **Sampling without replacement**: Choose k objects from a set of n one at a time without replacement.
  - Chosen object can't be chosen again.
  - Number of possible outcomes? $\mathsf{n}(\mathsf{n}-1)\cdots(\mathsf{n}-\mathsf{k}+1)$

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4
  - What if you can choose k $= 2$ objects. How many possible outcomes are there? Ans:4x4
  - There are n^k possible outcomes when order matters (multiplication rule)
- **Sampling without replacement**: Choose k objects from a set of n one at a time without replacement.
  - Chosen object can't be chosen again.
  - Number of possible outcomes? $n(n-1)\cdots(n-k+1)$
- Ex: 2024 best grad student paper awards has 11 candidates. How many top-3 award possibilities?

Sampling objects

- **Sample with replacement**: Choose k objects from a set of n, one at a time with replacement.
  - Any object may be selected multiple times
  - For example: $\Omega = \{a, b, c, d\}$, and you choose k $= 1$ object and order matters. How many possible outcomes are there? Ans:4
  - What if you can choose k $= 2$ objects. How many possible outcomes are there? Ans:4x4
  - There are n^k possible outcomes when order matters (multiplication rule)
- **Sampling without replacement**: Choose k objects from a set of n one at a time without replacement.
  - Chosen object can't be chosen again.
  - Number of possible outcomes? $\text{n}(\text{n}-1)\cdots(\text{n}-\text{k}+1)$
- Ex: 2024 best grad student paper awards has 11 candidates. How many top-3 award possibilities?
  - 11 first-place choices
  - 10 second-place choices among the remaining candidates
  - 9 third-place choices
  - Total: $11 \cdot 10 \cdot 9 = 990$ possibilities

2. Non-Naive Definition of Probability

Probability

- A probability space consists of:
  - Sample space Ω
  - Probability function $\mathbb{P}$ mapping events $\mathrm{A} \subseteq \Omega$ onto the real line.
- **The function $\mathbb{P}$ must satisfy the following axioms**:
  1. (Non-negativity) $\mathbb{P}(\mathrm{A}) \geq 0$ for every event
  2. (Normalization) $\mathbb{P}(\Omega) = 1$
  3. (Additivity) If a series of events, $\mathrm{A}_1, \mathrm{A}_2, \ldots$, are **disjoint**, then
  - Events A and B are disjoint if their intersection is zero: $P(A \cap B) = 0$
  - E.g. being early to class and being late to the same class

$$\mathbb{P}(\cup_{n=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

- Union of mutually exclusive events $\rightsquigarrow$ use additivity

Interpretation of probabilities

- How do we interpret $\mathbb{P}(A)$? Huge debate about this in stats literature
  1. **Frequentist**: $\mathbb{P}()$ reflect relative frequency in a large number of trials.
     - Repeat a coin flip many times frequency of head ≈ 0.5

Interpretation of probabilities

- How do we interpret $\mathbb{P}(A)$? Huge debate about this in stats literature
  1. **Frequentist**: $\mathbb{P}()$ reflect relative frequency in a large number of trials.
     - Repeat a coin flip many times frequency of head ≈ 0.5
  2. **Bayesian**: $\mathbb{P}()$ are subjective beliefs about outcomes
     - How likely I think a particular event will be

Interpretation of probabilities

- How do we interpret $\mathbb{P}(A)$? Huge debate about this in stats literature
  1. **Frequentist**: $\mathbb{P}()$ reflect relative frequency in a large number of trials.
     - Repeat a coin flip many times frequency of head ≈ 0.5
  2. **Bayesian**: $\mathbb{P}()$ are subjective beliefs about outcomes
     - How likely I think a particular event will be
- Both viewpoints are helpful in different contexts
  - Properties of probabilities exactly the same in either approach
  - This class: focus on frequentist perspectives. But we will talk a bit about Bayes next week, so you know how it works given that it's becoming more popular.

Gambling 101

- What's the probability of selecting a 8 card from a well-shuffled deck?
  - "Well-shuffled" $\rightsquigarrow$ "randomly selected" $\rightsquigarrow$ all cards have a chance

Gambling 101

- What's the probability of selecting a 8 card from a well-shuffled deck?
  - "Well-shuffled" $\rightsquigarrow$ "randomly selected" $\rightsquigarrow$ all cards have a chance
- "8 card" event $= \{8\clubsuit \cup 8\spadesuit \cup 8\heartsuit \cup 8\diamondsuit\}$
- Union of mutually exclusive events $\rightsquigarrow$ use additivity

$$\mathbb{P}\text{ (a 8 card)} = \mathbb{P}(8\clubsuit) + \mathbb{P}(8\spadesuit) + \mathbb{P}(8\heartsuit) + \mathbb{P}(8\diamondsuit) = \frac{4}{52}$$

Some properties of probabilities (refresher)

1. $\mathbb{P}(A') = 1 - \mathbb{P}(A)$
   - Probability of not A is 1 minus the probability of A.
   - Follows from $A \cup A' = \Omega$ and $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A')$

Some properties of probabilities (refresher)

1. $\mathbb{P}(A') = 1 - \mathbb{P}(A)$
   - Probability of not A is 1 minus the probability of A.
   - Follows from $A \cup A' = \Omega$ and $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A')$
2. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
   - Subsets of events have lower probability than the event.

Some properties of probabilities (refresher)

1. $\mathbb{P}(A') = 1 - \mathbb{P}(A)$
   - Probability of not A is 1 minus the probability of A.
   - Follows from $A \cup A' = \Omega$ and $1 = \mathbb{P}(\Omega) = \mathbb{P}(A) + \mathbb{P}(A')$
2. If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$
   - Subsets of events have lower probability than the event.
3. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
   - Avoid "double-counting" the part where A and B overlap.
   - **Inclusion-exclusion**