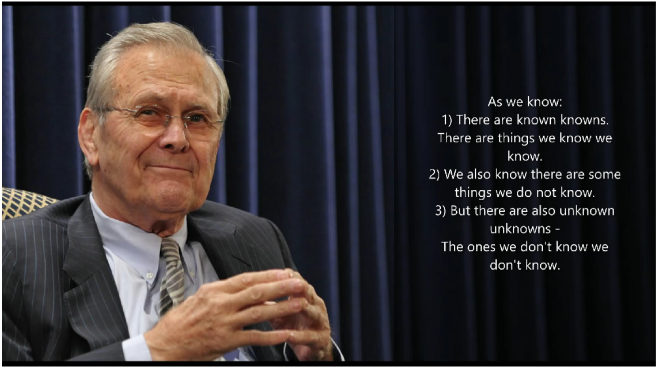
Probability

Ye Wang
University of North Carolina at Chapel Hill

Mathematics and Statistics For Political Research
POLI783

From certainties to uncertainties

- Calculus describes a world with certainties.
- But we live in a world full of uncertainties.
- Even some uncertainties can be quantified.



From certainties to uncertainties

- The result of a coin flip is uncertain.
- But the head and the tail will appear with equal chances.
- The effect of democracy on growth is uncertain.
- We can infer the probability for the effect to be non-zero.
- This branch of mathematics is known as probability.

Classic probability

- We use probabilities in daily life.
- E.g., what is the probability of raining tomorrow?
- Probability is intuitive when the possibilities are finite.
- E.g., the probability of rolling any face of a fair die is $1/6$.
- We can think it as the frequency for a consequence to occur if we repeat the process.
- This is how Pascal and Fermat studied probability.
- But classic probability runs into difficulties.
- What if there are infinite possibilities?
- What is the probability for candidate A's vote share to exceed 50%?

Modern probability

- Modern probability is founded by Kolmogorov in the early 20th century.
- It is built upon measure theory developed by Henri Lebesgue in the early 19th century.
- Measure theory: What is the length of a random set of real numbers? Can we define the length for any set of real numbers?
- Modern probability: What is the probability for a random outcome to occur? Can we define probability for any random outcome?

Probability and social science

- Probability has numerous applications in social science.
- Most social phenomena are random events.
- Sampling: each individual in the population has a probability to be drawn.
- How to set the probabilities if you are interested in Asian Americans who voted for Trump?
- Statistical inference: a point estimate is meaningless.
- We need to know how certain the finding is.
- Prediction: who will win the 2026 Election?

Definition of probability

- We need three components to define probability.
- The sample space Ω, the σ-algebra $\mathcal{F}$, and a probability function $\mathbb{P}$.
- A sample space consists of all possible outcomes of an "experiment."
- E.g., coin flip, tomorrow's weather, result of an election, etc.
- The outcomes must be observable ex post.
- What are the elements in Ω for a coin flip? How about for vote shares in an election?
- What about for two coin flips?

Definition of probability (*)

- Remember that the collection of all subsets in Ω is 2^{Ω}.
- Each element in 2^{Ω} is called an "event."
- We oftentimes cannot define a probability for any event in 2^{Ω}.
- We use the σ-algebra $\mathcal{F}$ to depict events with a probability.
- An "algebra" is a collection of sets that satisfies some properties.
  1. $\Omega \in \mathcal{F}$;
  2. if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$;
  3. if $A_1, A_2, \ldots, \in \mathcal{F}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{F}$.

Definition of probability (*)

- The smallest $\mathcal{F}$: $\{\Omega, \emptyset\}$.
- We can construct $\mathcal{F}$ from basic elements in Ω.
- E.g., for a coin flip, $\mathcal{F}$ can be $\{\emptyset, \{H\}, \{T\}, \{H, T\}\}$.
- For real numbers, we start from all open intervals and include their unions, intersections, and complements in $\mathcal{F}$.
- This is known as the "Borel sets."
- Only on Borel sets can we define basic measures such as length.
- In practice, $\mathcal{F}$ is just a technicality.

Definition of probability

- Finally, $\mathbb{P}$ assigns a number to each event in $\mathcal{F}$.
- This number is the probability of this event.
- We have some restrictions on $\mathbb{P}$'s value.
  1. $\mathbb{P}(\Omega) = 1$;
  2. $\mathbb{P}(S) \geq 0$ for any $S \in \mathcal{F}$;
  3. if $S_1, S_2, \ldots$ are disjoint events, then $\mathbb{P}(\cup_{k=1}^{\infty} S_k) = \sum_{k=1}^{\infty} \mathbb{P}(S_k)$.
- These are known as "axioms of probability."
- They are imposed to ensure that the defined probabilities are consistent with our intuition.

Properties of probability

- Using the axioms and properties of set operations, we can see:
- $\mathbb{P}(S^c) = 1 - \mathbb{P}(S)$, $\mathbb{P}(\emptyset) = 0$, $\mathbb{P}(S) \leq 1$.
- If $A \subset B$, then $\mathbb{P}(A) \leq \mathbb{P}(B)$.
- $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$.
- $\max\{0, \mathbb{P}(A) + \mathbb{P}(B) - 1\} \leq \mathbb{P}(A \cap B) \leq \min\{\mathbb{P}(A), \mathbb{P}(B)\}$ (Fréchet–Hoeffding bounds).

Compute probabilities

- In many cases, we assume all outcomes in the sample space are equally likely.
- With N possible outcomes in Ω, each outcome occurs with probability $\frac{1}{N}$.
- For any event $E \subset \Omega$, $\mathbb{P}(E) = \frac{|E|}{N}$.
- Computing probabilities reduces to counting outcomes.
- If two dice are rolled, what is the probability that the sum of the upturned faces will equal 7?
- There are totally 36 possible outcomes.
- Consider those with a sum of 7: $(1, 6), (6, 1)$, $(2, 5), (5, 2)$, $(3, 4), (4, 3)$.
- The probability is $\frac{6}{36} = \frac{1}{6}$.

Compute probabilities

- If $n < 365$ people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year?
- Count all possible combinations of birthdays: 365^n.
- The number of combinations that everyone has a unique birthday: $365 * 364 * ... * (365 - n + 1)$.
- The probability is $\frac{365*364*...*(365-n+1)}{365^n}$.
- In poker, what is the probability of being dealt a full house?
- Remember that a standard deck has 52 cards, 13 denominations, and 4 suits.
- A full house consists of 5 cards: 3 cards of the same denomination (e.g., three Kings) and 2 cards of a different but matching denomination (e.g., two 7s).
- Count all possible combinations of 5 cards: $\binom{52}{5}$.
- Count all possibilities of a full house: $\binom{13}{1}\binom{12}{1}\binom{4}{3}\binom{4}{2} \approx 0.014$.

Conditional probabilities

- The probability for one event to occur is often affected by things that have happened.
- What is the probability for Donald Trump to win if he was not shot?
- We use conditional probability to describe the probability for event A conditional on event B's outcome: $\mathbb{P}(A \mid B)$.
- E.g., $\mathbb{P}(\text{Trump won} \mid \text{Shooting})$.
- We define conditional probability using concepts we have learned:
$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$
- We can see that it satisfies the axioms of probability.
- All probabilities are normalized to event B.

An example

	Democrats	Republicans	Independents	Total
Men	33	40	2	75
Women	15	9	1	25
Total	48	49	3	100

- Choose one senator at random from this population.
- What is the probability that a randomly selected **Republican** is a woman?

$$\mathbb{P}(\text{Woman} \mid \text{Republican}) = \frac{9}{49} \approx 0.1837.$$

- Choose two senators at random (without replacement):
  - $\mathbb{P}(2 \text{ women} \mid \text{at least one woman})$
  - $\mathbb{P}(2 \text{ women} \mid \text{one draw is Liz Warren})$

Joint probabilities

- $\mathbb{P}(A \cap B)$ is known as the joint probability for the two events.
- It can be generalized to multiple events: $\mathbb{P}(\cap_{k=1}^{n} A_k)$.
- We sometimes write it as $\mathbb{P}(A_1, A_2, \ldots, A_n)$
- From the definition of conditional probability, we know that

$$\mathbb{P}(\cap_{k=1}^{n} A_k) = \\ \mathbb{P}(A_n \mid A_1, \ldots, A_{n-1})\mathbb{P}(A_{n-1} \mid A_1, \ldots, A_{n-2}) \ldots \mathbb{P}(A_1).$$

- We have been using it implicitly when calculating probabilities.
- E.g., the probability that no two people have the same birthday:

$$\mathbb{P}(A_{n-k} \mid A_1, \ldots, A_{n-k-1}) = \frac{\binom{365-n+k+1}{1}}{365^n} = \frac{365 - n + k + 1}{365^n}.$$

Law of total probability

- Probability and conditional probability are connected.
- For any events A and B, we have

$$A = (A \cap B) \cup (A \cap B^c).$$

- Since $A \cap B$ and $A \cap B^c$ are disjoint events, we can use the axioms of probability and obtain

$$\mathbb{P}(A) = \mathbb{P}(A \cap B) + \mathbb{P}(A \cap B^c) = \mathbb{P}(A \mid B)\mathbb{P}(B) + \mathbb{P}(A \mid B^c)\mathbb{P}(B^c).$$

- We call $\{B_k\}_{k=1}^K$ a partition of S if any B_k and B_l are disjoint and $\cup_{k=1}^K B_k = S$.
- Using the same logic, we have

$$\mathbb{P}(A) = \sum_{k=1}^K \mathbb{P}(A \cap B_k) = \sum_{k=1}^K \mathbb{P}(A \mid B_k)\mathbb{P}(B_k).$$

- This is known as the law of total probability.

Law of total probability

- This law gives us an approach to calculate the probability for A to occur.
- We can first calculate the conditional probability for A to occur in each of the disjoint categories (B_k).
- The probability equals the sum over these conditional probabilities multiplied by the probability for each category to occur.

Law of total probability

- Consider the relationship between **education** and **support for a climate policy**.
- The data are summarized below:

Education Level	$\mathbb{P}(E)$	$\mathbb{P}(S \mid E)$
High school or less	0.40	0.30
Bachelor's degree	0.35	0.60
Graduate or above	0.25	0.80

- What is the probability that a randomly selected person supports the policy?

$$\mathbb{P}(S) = \sum_i \mathbb{P}(S \mid E_i) \cdot \mathbb{P}(E_i)$$

$$\mathbb{P}(S) = 0.3 \times 0.4 + 0.6 \times 0.35 + 0.8 \times 0.25 = 0.54$$

Bayes' rule

- Usually, $\mathbb{P}(A \mid B) \neq \mathbb{P}(B \mid A)$.
- E.g., A: the person is a dissident; B: the person posted anti-government content online
- $\mathbb{P}(A \mid B)$: What is the chance that someone who posts anti-government content is a dissident?
- $\mathbb{P}(B \mid A)$: What is the chance that a dissident posts anti-government content?
- These two are connected through Bayes' Rule:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B)}.$$

- Bayes' Rule lets us compute one conditional probability from the other, given $\mathbb{P}(A)$ and $\mathbb{P}(B)$.

Bayes' rule

- Suppose 1% of the population are dissidents.
- A dissident posts anti-government content with a probability of 60%.
- A non-dissident posts such content with a probability of 2%.
- You are a police officer and observe someone posting anti-government content.
- What is the probability this person is a true dissident?
- In this example, $\mathbb{P}(A) = 0.01$, $\mathbb{P}(B \mid A) = 0.6$, and $\mathbb{P}(B \mid A^c) = 0.02$.
- First step: $\mathbb{P}(B) = \mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid A^c)\mathbb{P}(A^c) = 0.006 + 0.0198 = 0.0258$.
- Second step: $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)} = \frac{0.6 * 0.01}{0.0258} = 0.233$.
- Catching dissidents is hard!

From Bayes' rule to Bayesian statistics

- In practice, A is often a parameter of interest, while B represents data we have collected.
- E.g., A: Putin's popularity in Russia; B: responses from a survey on Russians.
- $\mathbb{P}(A)$: our current estimate of Putin's popularity in Russia (the prior).
- $\mathbb{P}(B \mid A)$: the probability for any response to occur given Putin's popularity (the likelihood).
- $\mathbb{P}(A \mid B)$: the updated estimate of Putin's popularity given the responses (the posterior).
- Basic steps of Bayesian statistics: 1. specify the prior and likelihood functions; 2. collect data; 3. estimate the posterior.

Independence

- If B has no impact on the probability for A to happen, $\mathbb{P}(A \mid B) = \mathbb{P}(A)$, and we say the two events are independent.
- We denote it as $A \perp\!\!\!\perp B$.
- $A \perp\!\!\!\perp B$ implies $B \perp\!\!\!\perp A$ and vice versa.
- Using the definition of conditional probability, we can see that

$$\mathbb{P}(A) = \mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

- Therefore

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

- The probability for both A and B to happen equals the probability for A to happen multiples that for B to happen.
- B contains no information for us to learn about A.

Independence

- The independence of multiple events, $\{A_k\}_{k=1}^K$, means that

$$\mathbb{P}\left(\cap_{k=1}^K A_k\right) = \prod_{k=1}^K \mathbb{P}(A_k).$$

- Each event A_k contains no information of other events.
- Therefore, $A_k \perp\!\!\!\perp A_l$ for any k and l.
- The converse is not true!
- Pairwise independence does not imply joint independence.
- E.g., the result of tossing a coin twice: $\Omega = \{HH, TH, HT, TT\}$.
- A: first toss is heads ($\{HH, HT\}$); B: second toss is heads ($\{TH, HH\}$); C: even number of heads ($\{HH, TT\}$).
- $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{2}$, and $\mathbb{P}(C) = \frac{1}{2}$.
- $\mathbb{P}(A \cap B) = \frac{1}{4}$, $\mathbb{P}(A \cap C) = \frac{1}{4}$, and $\mathbb{P}(B \cap C) = \frac{1}{4}$.
- But $\mathbb{P}(A \cap B \cap C) = \frac{1}{4} \neq \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C)$.

Conditional independence

- Sometimes, two events are independent only given additional information:

$$\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid C)\mathbb{P}(B \mid C).$$

- We denote it as $A \perp\!\!\!\perp B \mid C$.
- $A \perp\!\!\!\perp B \mid C \not\Rightarrow A \perp\!\!\!\perp B$, and $A \perp\!\!\!\perp B \not\Rightarrow A \perp\!\!\!\perp B \mid C$.
- A: being a liberal; B: support climate policies; C: being a Democrat.
- A: good at writing; B: good at math; C: being admitted to college.
- Conditional independence plays a central role in causal inference.