# Probability

## Statistical Methods in Political Research I

Yuki Shiraito

University of Michigan

Fall 2021

# Example: Survey in Nigeria

Interviewer says to a respondent:

*For this question, I want you to answer yes or no. But I want you to consider the number of your dice throw. If 1 shows on the dice, tell me no. If 6 shows, tell me yes. But if another number, like 2 or 3 or 4 or 5 shows, tell me your own opinion about the question that I will ask you after you throw the dice. [TURN AWAY FROM THE RESPONDENT] Now you throw the dice so that I cannot see what comes out. Please do not forget the number that comes out. [ WAIT TO TURN AROUND UNTIL RES- PONDENT SAYS YES TO: ] Have you thrown the dice? Have you picked it up?*

*Now, during the height of the conflict in 2007 and 2008, did you know any militants, like a family member, a friend, or someone you talked to on a regular basis. Please, before you answer, take note of the number you rolled on the dice.*

(Blair, Imai, and Zhou 2015)

# Outcomes and Events in the Example

- Outcomes
  1. Respondent's true answer is yes, and 1 shows
  2. Respondent's true answer is yes, and 2 shows
  3. Respondent's true answer is yes, and 3 shows
  4. Respondent's true answer is yes, and 4 shows
  5. Respondent's true answer is yes, and 5 shows
  6. Respondent's true answer is yes, and 6 shows
  7. Respondent's true answer is no, and 1 shows
  8. Respondent's true answer is no, and 2 shows
  9. Respondent's true answer is no, and 3 shows
  10. Respondent's true answer is no, and 4 shows
  11. Respondent's true answer is no, and 5 shows
  12. Respondent's true answer is no, and 6 shows
- Events
  1. Respondent answers "yes"
  2. Respondent answers "no"
  3. Respondent answers "yes" or "no"
  4. Null

# Generalization: Axiomatic Approach

- Probability Space $(\Omega, \mathcal{F}, \mathbb{P})$:
  1. Sample space, $\Omega$: A non-empty set
  2. Set of events, $\mathcal{F}$: A set of subsets of $\Omega$ such that
     - $\Omega \in \mathcal{F}$
     - $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
     - $A_i \in \mathcal{F}$ for $i = 1, 2, \cdots \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$
  3. A probability measure, $\mathbb{P}$: A function such that
     - $\mathbb{P} : \mathcal{F} \to [0, 1]$
     - $\mathbb{P}(\Omega) = 1$
     - $A_i \in \mathcal{F}$ for $i = 1, 2, \ldots$ and $A_i \cap A_j = \emptyset$ for any $i \neq j$ $\Rightarrow \mathbb{P}(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

- The set of the events in the previous slide satisfies Condition 2 **Proof in the next slide**

- $\mathbb{P}$ has to satisfy $\mathbb{P}($ "yes" $) = p$ and $\mathbb{P}($ "no" $) = 1 - p$, where $0 \leq p \leq 1$.

# Proof

- $\Omega \in \mathcal{F}$
  - Event "Respondent answers 'yes' or 'no"'
- $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
  1. (Respondent answers "yes") and (Respondent answers "no")
  2. (Respondent answers "yes" or "no") and Null
- $A_i \in \mathcal{F}$ for $i = 1, 2, \cdots \Rightarrow \cup_{i=1}^{\infty} A_i \in \mathcal{F}$
  1. Union of (Respondent answers "yes") and any others:
     1. (Respondent answers "no") $\Rightarrow$ (Respondent answers "yes" or "no")
     2. NULL $\Rightarrow$ (Respondent answers "yes")
     3. (Respondent answers "yes" or "no") $\Rightarrow$ (Respondent answers "yes" or "no")

  ⋮

# Important Properties

- Set of events:
  1. If $A_1, A_2 \in \mathcal{F}$ then $A_1 \cap A_2 \in \mathcal{F}$ and $A_1 \setminus A_2 \equiv A_1 \cap A_2^c \in \mathcal{F}$
     **Proof:**
     1. The fact that $A_1, A_2 \in \mathcal{F}$ implies that $A_1^c, A_2^c \in \mathcal{F}$.
     2. Define $A_i = \Omega$ for $i \geq 3$. Then, $\cup_{i=1}^{\infty} A_i^c = A_1^c \cup A_2^c$ and so $A_1^c \cup A_2^c \in \mathcal{F}$.
     3. Therefore, $(A_1^c \cup A_2^c)^c = A_1 \cap A_2 \in \mathcal{F}$.
     4. The same proof applies to $A_1 \cap A_2^c$ because $A_2^c \in \mathcal{F}$.

- Probability: Let $A_1, A_2 \in \mathcal{F}$. Then,
  1. $\mathbb{P}(A_1^c) = 1 - \mathbb{P}(A_1)$
  2. $A_1 \subset A_2 \Rightarrow \mathbb{P}(A_1) \leq \mathbb{P}(A_2)$
  3. $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap A_2)$
     **Proofs:** Left to Fabricio's section and problem sets

# Modeling the Survey using Probability

- The set of events describing the survey response—useful?
- No: we want to know the respondent's true answer

- Need to modify the model:
  - Include "Respondent's true answer is yes" in the set of events
  - Expand the set of events so it satisfies the condition
- Events
  1. Respondent answers "yes"
  2. Respondent answers "no"
  3. Respondent answers "yes" or "no"
  4. Null
  5. Respondent's true answer is yes
  6. Respondent's true answer is no
  7. Union of event 1 and 5
  8. Union of event 1 and 6
  9. Union of event 2 and 5
  10. Union of event 2 and 6
  11. doesn't stop here...

# Countable Sample Spaces

- The "size" of a sample space may be:
  1. Countable: Discrete outcomes
     1. Finite: One-to-one correspondence to $\{1, 2, \ldots, n\}$ for some $n$
     2. Infinite: One-to-one correspondence to $\{1, 2, \ldots\}$
  2. Uncountable: Continuum of outcomes
     1. Always infinite

- For a countable space $\Omega$, we can have:
  - $\mathcal{F} = 2^{\Omega}$: The set of events is the power set of $\Omega$
  - $p : \Omega \to [0, 1]$ such that $\sum_{\omega \in \Omega} p(\omega) = 1$: Probability mass function
  - For any $A \subset \mathcal{F}$, $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$

- Intuitive case–e.g.:
  - Assume all outcomes in the Nigeria survey example have probability mass $1/12$
  - You can calculate the probability of any event

# Example: Court

- 9 justices vote on a case:
  1. Roberts
  2. Thomas
  3. Breyer
  4. Alito
  5. Sotomayor
  6. Kagan
  7. Gorsuch
  8. Kavanaugh
  9. Barret

- Question: What is the probability of the plaintiff winning?
- Steps:
  1. How many possible outcomes are there?
  2. What are the events where the plaintiff wins?
  3. How many outcomes are in those events?

# Number of Outcomes and Events

- Assuming no abstentions, how many outcomes are there?
- Multiplication Rule: Multiples of # of choices
- Two choices for each of 9 Justices: $2^9 = 512$
- Example of sampling with replacement: any choice is not precluded by others' choices
- What if only Justice Roberts abstains?

- Events where the plaintiff wins:
  - 5 justices vote for the plaintiff
  - 6 justices vote for the plaintiff
  - 7 justices vote for the plaintiff
  - 8 justices vote for the plaintiff
  - 9 justices vote for the plaintiff

# Permutations

- How many outcomes in each of the events?
- Which justices vote for the plaintiff?

- Sampling without replacement:
  - Choices are precluded after they are chosen
  - Once you choose Justice Roberts, you cannot choose him again
  - $9 \times 8 \times 7 \times 6 \times 5 = 15120$ ways to sample 5 out of 9 justices
  - More generally, $n(n-1)(n-2)\ldots(n-k+1)$ ways to sample $k$ out of $n$ objects

- Overcounting:
  1. Roberts, Thomas, Barret, Breyer, Alito
  2. Barret, Alito, Thomas, Breyer, Robers

  $\vdots$

- How many times we count the same set of justices?
- Permutations:
  - Special case of sampling without replacement
  - Sample and exhaust all 5 justices: $5 \times 4 \times 3 \times 2 \times 1 = 5! = 120$
  - More generally, $k!$ ways to permute $k$ objects

# Combinations and Binomial Theorem

- Adjust overcounting: Combinations

$$\frac{\text{number of ways to sample 5 justices without replacement}}{\text{number of ways to permute 5 justices}}$$
$$= \frac{9 \times 8 \times 7 \times 6 \times 5}{5!}$$

- General formula: Binomial coefficient

$$\binom{n}{k} \equiv \frac{n(n-1)(n-2)\dots(n-k+1)}{k!} = \frac{n!}{(n-k)!k!}$$

- Application: Binomial theorem

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{(n-k)}$$

# Example: Survey in Nigeria, Modified

The interviewer gives the respondent a stick, instead of a dice, and says:

*For this question, I want you to answer yes or no. But I want you to consider in which direction your stick falls. Once I turn away from you, please put your stick upright on the ground and then take your hands off. If the stick falls in between the true north and azimuth 60 degrees, tell me no. If the stick falls in between the true north and azimuth 300 degrees, tell me yes. Otherwise, tell me your own opinion about the question that I will ask. [TURN AWAY FROM THE RESPONDENT] ...*

# Uncountable Sample Spaces

- Outcomes are continuum: Stick may fall in any direction
- Outcomes:
  1. Respondent's true answer is yes, and the stick falls azimuth $x$ degrees
  2. Respondent's true answer is no, and the stick falls azimuth $x$ degrees

  for $x \in [0, 360)$
- Substantively, the situation is exactly identical

- Sample space: $\Omega = \{\text{yes}, \text{no}\} \times [0, 360)$
- Event
  1. "Respondent answers 'yes"':
     $A = (\{\text{yes}, \text{no}\} \times [300, 360)) \cup (\{\text{yes}\} \times [60, 300))$
- Probability: $\mathbb{P}(A) = p$ where $0 \leq p \leq 1$
- Probability mass function cannot be consistent:
  1. If any, $p(\omega)$ is constant for all $\omega$ with the same true answer
  2. If $p(\omega) > 0$ then $\sum_{\omega \in A} p(\omega) = \infty$
  3. If $p(\omega) = 0$ then $\sum_{\omega \in A} p(\omega) = 0$
- Any $\{\omega\}$ where $\omega \in \Omega$ is not included in $\mathcal{F}$

# Rolling a Dice and the True Answer

- Formalize the Nigeria survey with a dice roll example:
  1. Sample space $\Omega$ contains 12 outcomes
  2. Set of events $\mathcal{F}$ is the power set of $\Omega$, $2^\Omega$
  3. Probability mass function is denoted by $p(\omega_i) = p_i$ where
     - true answer is yes and $i$ shows on the dice for $i = 1, \ldots, 6$
     - true answer is no and $i - 6$ shows on the dice for $i = 7, \ldots, 12$
  4. Dice is fair: $p_1 = \cdots = p_6 = p_Y$ and $p_7 = \cdots = p_{12} = p_N$

- Events
  1. Respondent's true answer is yes: $G \equiv \{\omega_1, \ldots, \omega_6\}$
  2. 1 shows on the dice: $B_1 \equiv \{\omega_1, \omega_7\}$
  3. More generally, $k$ shows on the dice: $B_k$

- Does the dice have any information about respondent's true answer?

# Independence of Events

- Independence of two events: Two events $A_1$ and $A_2$ are independent if and only if
$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1)\mathbb{P}(A_2).$$

- $G$ and $B_k, k = 1, \ldots, 6$ are independent:
$$\mathbb{P}(G \cap B_k) = p_y$$
$$\mathbb{P}(G)\mathbb{P}(B_k) = 6p_y \times \frac{1}{6} = p_y$$

- Independent events are "unrelated"
- Can create artificially indenpendent events:
    - $\Omega$ is independent of any event
    - $\emptyset$ is independent of any event
    - $B_2 \cup B_3$ and $B_1 \cup B_3 \cup B_5$ are independent

# Mutually Independent

- What if there are more than two events?
- Independence of more than two events: Events $A_1, \dots, A_K$ are mutually independent if and only if for every subset $A_{i_1}, \dots, A_{i_J}$ of $J$ of these events ($J = 2, \dots, K$),
$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_J}) = \mathbb{P}(A_{i_1}) \dots \mathbb{P}(A_{i_J})$$
- Are $G, B_2 \cup B_3$, and $B_1 \cup B_3 \cup B_5$ are independent?
  1. $G$ and $B_2 \cup B_3$ are indepedent
  2. $G$ and $B_1 \cup B_3 \cup B_5$ are independent
  3. $B_2 \cup B_3$ and $B_1 \cup B_3 \cup B_5$ are independent
  4. $G, B_2 \cup B_3$, and $B_1 \cup B_3 \cup B_5$:
$$\mathbb{P}\big(G \cap (B_2 \cup B_3) \cap (B_1 \cup B_3 \cup B_5)\big) = p(\omega_3) = p_Y$$
$$\mathbb{P}(G)\mathbb{P}(B_2 \cup B_3)\mathbb{P}(B_1 \cup B_3 \cup B_5) = 6p_Y \times \frac{1}{3} \times \frac{1}{2} = p_Y$$
- However, pairwise independence does not necessarily imply mutual independence (**Fabricio's section**)

# Conditioning

- If you know an event has occurred, can you say anything about the probability of another event?
  = Given that the realized outcome is included in event $A$, what is the probability that the realized outcome is also included in event $B$?

- Intuition:
  - You happen to know 1 shows on the dice
  - The outcome must be either $\omega_1$ or $\omega_7$
  - What is the probability of $G$?
  - If $G$ is occurring, $\omega_1$ has to be the outcome
  - So, the probability should be:

$$\frac{p(\omega_1)}{p(\omega_1) + p(\omega_7)}$$

# Conditional Probability

- Conditional probability: For event $A_2$ with $\mathbb{P}(A_2) > 0$, the conditional probability of event $A_1$ given $A_2$, denoted by $\mathbb{P}(A_1 \mid A_2)$, is defined as:
$$\mathbb{P}(A_1 \mid A_2) = \frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)}$$
- $\mathbb{P}(G \mid B_1) = \mathbb{P}(G \cap B_1)/\mathbb{P}(B_1) = p_Y/(p_Y + p_N)$
- Independence and conditional probability: If events $A_1$ and $A_2$ are independent and $\mathbb{P}(A_2) > 0$, then
$$\mathbb{P}(A_1) = \mathbb{P}(A_1 \mid A_2)$$
- $G$ and $B_1$ are independent:
  - $\mathbb{P}(G) = 6p_Y$
  - $\mathbb{P}(G \mid B_1) = p_Y/(p_Y + p_N) = 6p_Y/6(p_Y + p_N) = 6p_Y$
- Independence: Does not change your belief

# Conditional Probability and Probability Axioms

- Conditional probabilities are probabilities
- For a fixed event $A_2$, $\mathbb{P}(\cdot \mid A_2)$ satisfies the axioms of probability

- For any $A_1$, $0 \leq \mathbb{P}(A_1 \mid A_2)$.
  **Proof:** $0 \leq \mathbb{P}(A_1 \cap A_2)$ implies that $\frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_2)} = \mathbb{P}(A_1 \mid A_2) \geq 0$.

- $\mathbb{P}(\Omega \mid A_2) = 1$.
  **Proof:** $\mathbb{P}(\Omega \mid A_2) = \frac{\mathbb{P}(\Omega \cap A_2)}{\mathbb{P}(A_2)} = \frac{\mathbb{P}(A_2)}{\mathbb{P}(A_2)} = 1$.

- $\widetilde{A}_i \in \mathcal{F}$ for $i = 1, 2, \ldots$ and $\widetilde{A}_i \cap \widetilde{A}_j = \emptyset$ for any $i \neq j$
  $\Rightarrow \mathbb{P}(\cup_{i=1}^{\infty} \widetilde{A}_i \mid A_2) = \sum_{i=1}^{\infty} \mathbb{P}(\widetilde{A}_i \mid A_2)$
  **Proof:**
  $$\mathbb{P}(\cup_{i=1}^{\infty} \widetilde{A}_i \mid A_2) = \frac{\mathbb{P}(\cup_{i=1}^{\infty} \widetilde{A}_i \cap A_2)}{\mathbb{P}(A_2)} = \frac{\sum_{i=1}^{\infty} \mathbb{P}(\widetilde{A}_i \cap A_2)}{\mathbb{P}(A_2)} = \sum_{i=1}^{\infty} \underbrace{\frac{\mathbb{P}(\widetilde{A}_i \cap A_2)}{\mathbb{P}(A_2)}}_{\mathbb{P}(\widetilde{A}_i \mid A_2)}.$$

# Important Properties of Conditional Probability

- For events $A_1$ and $A_2$, if $\mathbb{P}(A_2) > 0$ then
$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2)\mathbb{P}(A_1 \mid A_2).$$

- Law of total probability: Let $A_1, \ldots, A_K$ be a partition of the sample space $\Omega$ and $\mathbb{P}(A_k) > 0$ for all $k = 1, \ldots, K$. Then, for any event $E$,
$$\mathbb{P}(E) = \sum_{k=1}^{K} \mathbb{P}(A_k)\mathbb{P}(E \mid A_k)$$

- Nigeria survey example: Conditional on each number on the dice,
$$\mathbb{P}(G) = 6p_Y$$
$$\sum_{k=1}^{6} \mathbb{P}(B_k)\mathbb{P}(G \mid B_k) = \sum_{k=1}^{6} \frac{1}{6}\frac{p_Y}{p_Y + p_N} = \frac{6p_Y}{6(p_Y + p_N)} = 6p_Y$$

# Independence and Conditional Independence

- Conditional probabilities are probabilities
- So, we can also think of conditional independence
- Two events $A_1$ and $A_2$ are conditionally independent given $A_3$ with $\mathbb{P}(A_3) > 0$ if and only if
$$\mathbb{P}(A_1 \cap A_2 \mid A_3) = \mathbb{P}(A_1 \mid A_3)\mathbb{P}(A_2 \mid A_3)$$
- Unconditional independence neither implies nor is implied by conditional independence

# Bayes' Rule

- Model: True answer being yes is probability $6p_Y$.
- "Nature" (a.k.a. "data generating process") knows $p_Y$
- Data are generated according to:

$$\mathbb{P}\big(\text{True answer is } \cdot, \text{Dice shows } \cdot\big)$$

$$\mathbb{P}\big(\text{Response is "yes"} \mid \text{True answer is } \cdot, \text{Dice shows } \cdot\big)$$

- Analysts don't know $p_Y$: We want to estimate it from data
- Question: Given an observed response, what is the probability of the true answer?
- Posterior (probability): $\mathbb{P}(\text{True answer is } \cdot \mid \text{Response is } \cdot)$

- Bayes' Rule: For events $A_1$ and $A_2$ with non-zero probability measures,

$$\mathbb{P}(A_1 \mid A_2) = \frac{\mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_1)}{\mathbb{P}(A_2)}$$

# Bayes' Theorem

- Bayes' rule + Law of total probability: Let $\widetilde{A}_1, \ldots, \widetilde{A}_K$ be a partition of $\Omega$ with $\mathbb{P}(\widetilde{A}_k) > 0$ and $\mathbb{P}(A_2) > 0$. Then
$$\mathbb{P}(\widetilde{A}_k \mid A_2) = \frac{\mathbb{P}(A_2 \mid \widetilde{A}_k)\mathbb{P}(\widetilde{A}_k)}{\sum_{k'=1}^{K} \mathbb{P}(A_2 \mid \widetilde{A}_{k'})\mathbb{P}(\widetilde{A}_{k'})}$$

- Use of Bayes' theorem in statistics: Posterior of hidden truth
- Posterior of truth can be computed from:
  1. Conditional probability of data given truth = model
  2. Marginal probability of truth = prior

- Can compute what's unknown from what's known: Estimation

# Application to Nigeria Survey Example

- Given that a response is "yes," what is the posterior that the true answer is yes?

- $\mathbb{P}(\text{Response is "yes"} \mid \text{True answer is yes}) = \frac{5}{6}$
- $\mathbb{P}(\text{Response is "yes"} \mid \text{True answer is no}) = \frac{1}{6}$
- Prior belief: $\mathbb{P}(\text{True answer is yes}) = \rho$
- Applying Bayes' theorem:

$$\frac{\frac{5}{6}\rho}{\frac{5}{6}\rho + \frac{1}{6}(1-\rho)} = \frac{5\rho}{1+4\rho}$$

- If $\rho = \frac{1}{2}$, then $\frac{5}{6}$
- What if $\rho = \frac{1}{1000}$?

# Role of Prior and Data

- Posterior may be sensitive to prior:
  1. If $\rho = \frac{1}{2}$, posterior is $\frac{5}{6}$
  2. If $\rho = \frac{1}{3}$, posterior is $\frac{5}{7}$
  3. If $\rho = \frac{1}{1000}$, posterior is $\frac{5}{1004}$
- In extreme cases:
  1. $\rho = 0$: Posterior is always 0
  2. $\rho = 1$: Posterior is always 1
- If you know truth *a priori*, you never update your belief
- Prior determines how surprising data are:
  1. Surprising $\Rightarrow$ bigger change in belief
  2. Too surprising $\Rightarrow$ ignore data

# Multiple Responses

- Remember: Conditional probabilities are probabilities
- Conditional version of Bayes' theorem:

$$\mathbb{P}(\widetilde{A}_k \mid A_2, A_3) = \frac{\mathbb{P}(A_2 \mid \widetilde{A}_k, A_3)\mathbb{P}(\widetilde{A}_k \mid A_3)}{\mathbb{P}(A_2 \mid A_3)}$$
$$= \frac{\mathbb{P}(A_2 \mid \widetilde{A}_k, A_3)\mathbb{P}(\widetilde{A}_k \mid A_3)}{\sum_{k'=1}^{K} \mathbb{P}(A_2 \mid \widetilde{A}_{k'}, A_3)\mathbb{P}(\widetilde{A}_{k'} \mid A_3)}$$

- What if a respondent answers the question twice?
- Posterior:

$$\mathbb{P}(\text{True answer is yes} \mid \text{Response 1 and 2 are "yes"})$$
$$= \mathbb{P}(\text{Response 2 is "yes"} \mid \text{True answer is yes})$$
$$\times \mathbb{P}(\text{Response 1 is "yes"} \mid \text{True answer is yes})$$
$$\times \mathbb{P}(\text{True answer is yes}) \Big/ \mathbb{P}(\text{Response 1 and 2 are "yes"})$$

# Key Points

- Probability spaces:
  - Distinction between outcomes and events:
    1. One outcome realizes from one trial
    2. Events are sets of outcomes
  - Probability is defined for events, not outcomes
  - Countable outcomes: Probability can be defined for outcomes

- Counting:
  - Be aware of overcounting and adjust it

- Independence and Conditional Probabilities:
  - Conditional probabilities are probabilities
  - Independence: Do not change your belief

- Bayes:
  - Compute posterior using model and prior
  - Prior determines how informative data are