PS 2010: 4. Probability

Qing Chang

Department of Political Science
University of Pittsburgh

Fall 2023

- Homework 2 Due today
- Homework 3 is out
  Due Sep 26th 8:59:59am

What we have learned

- Instantaneous rate of change
- Limit
- Numerical differentiation and derivatives
- Integral: area problem

Today's Agenda

- Probability
- Counting Rules
- Conditional Probability
- Probability mass/density function

What is probability

- **Probability:** is a measure of uncertainty
- Measures the likelihood of events or outcomes occurring
- Frequentist: the limit of relative frequency.

$$P = \frac{\#\ \textit{of event occurs}}{\#\ \textit{of trials in repeated trials under the same conditions}}$$

  - Same conditions?
  - Repeated trials?
- Bayesian: a measure of one's subjective belief about the likelihood of an event occurring
  - Too subjective?

Key Concepts

- We need to three concepts to define probability
  - Experiment: actions that produce stochastic events
  - Sample space: a set of all possible outcomes
  - Event: a subset of the sample space
- Example:
  - Flipping a coin
  - Voting

Probability Axioms

- $P(A) >= 0$
- $P(All\ possible\ outcomes) = 1$
- If events A and B are mutually exclusive, then $P(A \cup B) = P(A) + P(B)$
- Voting example

Venn Diagram

Figure: Event A and B are mutually exclusive

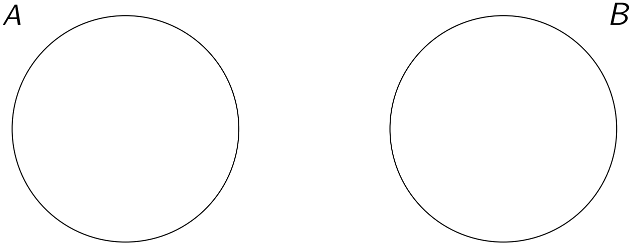
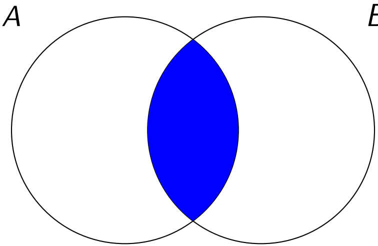


Figure: Event A and B are not mutually exclusive



Additional Rule

- For any events A and B,

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

- *Complement of a set* consist all elements in the sample space except those in the set
  - $P(A^c) = 1 - P(A)$
  - $P(A \cup B) = P(A \cap B^c) + P(B \cap A^c) + P(A \cap B)$
- Law of total probability

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

mn Rule

- With m elements and n elements, it is possible to form $m * n$ pairs containing one element from each group
- Examples:
  - Define # of sample points for tossing two dice
  - What is the probability of different birthday of 20 people from 365 days?

Permutation

- Permutation refers to an arrangement of objects in a specific order
- Without repetition
  - Arrange items A,B,C
- Position matters
- When arranging k objects out of n unique objects:

$$P(n,k) = \frac{n!}{(n-k)!}$$

  where ! is factorial operator,
  $n! = n * (n-1) * (n-2) \cdots * 2 * 1$, and $0! = 1$

Example

- In a political organization, you need to form a committee of three members out of a group of six individuals. How many different ways can you choose the committee members?
- 20 faculty for a committee of 3 people: chair, m1, m2
- Award 2 people from 10 people

Sampling With/Without Replacement

- Sample with replacement: same unit can be repeatedly sampled
- Sample without replacement: each unit can be sampled at most once
- The above example again

Combinations

- When choosing k distinct elements from n elements:

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

- Position does not matter
- Examples
  - 20 faculty for a committee of 3 people
  - Award 2 people from 10 people

Example

- A group of three undergraduate and five graduate students are available to fill certain student government posts. If four students are to be randomly selected from this group, find the probability that exactly two undergraduates will be among the four chosen.

Conditional Probability

- Conditional probability: $P(A \mid B)$

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

- $P(A \cap B)$ is the **joint probability**
- $P(B)$ is the **marginal probability**
- We can also have

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

- Conditional probability brings in new information
  - $P(Trump) = 0.49$
  - $P(Trump \mid D)$

Examples

- In a survey of voters, 60% of respondents support Candidate A, and 40% support Candidate B. Among those who support Candidate A, 75% are female. Among those who support Candidate B, 60% are female. What is the probability that a randomly selected voter is female, given that they support Candidate A?
- In a study on policy support, it is found that 70% of liberals and 40% of conservatives support a specific policy. Assume we have 40% liberals and 60% of conservatives in the total population. If a randomly selected individual supports the policy, what is the probability that they are a liberal?

Example

Recent polls suggest that
83% of Democrats support impeachment of Trump
44% of Independents support impeachment of Trump
14% of Republicans support impeachment of Trump
Meanwhile, $P(D) = 31\%$, $P(I) = 40\%$, $P(R) = 29\%$

- What is the average probability of supporting for impeachment?
- Is this true for all Americans?
- Find the probability that a randomly selected American is a Democrat who supports impeachment.
- Now suppose that the randomly selected American supports impeachment. How does this information change the probability that the selected American belongs to a particular political party?

Independence

- Knowledge of one event does not provide any more information about the other event

$$P(A \mid B) = P(A) \text{ and } P(B \mid A) = P(B)$$

- If two events A and B are independent:

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

- Example: randomization

Bayes Rule

- Update our beliefs after we observe the data
- **Definition:**

$$P(A \mid B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{P(B|A)P(A)+P(B|A^c)P(A^c)}$$

  $P(A)$ is prior probability, $P(A \mid B)$ is posterior probability
- Recall law of total probability is

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

  we know
- $P(A \cap B) = P(B \mid A)P(A)$
- Therefore, $P(B) = P(B \mid A)P(A) + P(B \mid A^c)\, P(A^c)$

Example

The fraction of persons in the population who have a particular disease is .01. The probability that a particular diagnostic test gives a positive results is .2. The probability that the test gives a positive result given that an individual actually has the disease is .8. If an individual has the test done, and it comes back positive, what is the probability that she or he has the disease? Why is this probability not equal to one?

Example

In a certain election campaign, three campaign strategies, Strategy A, Strategy B, and Strategy C, are employed by a political party to target different voter demographics, contributing to 30%, 45%, and 25% of their outreach efforts, respectively. It is known from past elections that 2%, 3%, and 2% of the voters influenced by each strategy, respectively, eventually became disengaged from the political process.

- Suppose that a vote is randomly selected. What is the probability that this voter is disengaged?
- If a voter is chosen randomly and found to be disengaged, what is the probability that it was due to strategy C?