Lecture 18 – Foundations of Probability (continued)

DSC 40A, Fall 2022 @ UC San Diego

Dr. Truong Son Hy, with help from **many others**

Announcements

- Look at the readings linked on the course website!
- Groupwork Relsease Day: Thursday afternoon
  Groupwork Submission Day: Monday midnight
  Homework Release Day: Friday after lecture
  Homework Submission Day: Friday before lecture
- See `dsc40a.com/calendar` for the Office Hours schedule.
- We graded the midterm and will release the grades soon.

Agenda

- Complement and addition rules for probability
- Principle of inclusion-exclusion
- Multiplication rules
- Conditional probability

Probability: Complement & addition

Sets

- A **set** is an unordered collection of items.
- A **sample space**, S, is the set of all possible outcomes of an experiment.
- An **event**, A is a subset of the sample space. In other words, an event is a set of outcomes.
  - Notationally: $A \subseteq S$.
- $|A|$ denotes the number of elements (i.e. cardinality) in set A.

Equally-likely outcomes

- If S is a sample space with n possible outcomes, and all outcomes are equally-likely, then the probability of any one outcome occurring is $\frac{1}{n}$.

- The probability of an event A, then, is

$$P(A) = \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} = \frac{\text{\# of outcomes in A}}{\text{\# of outcomes in S}} = \frac{|A|}{|S|}$$

- **Example:** Flip a (fair) coin three times and we record head/tail each time. How large is the sample space S?

Equally-likely outcomes

- If S is a sample space with n possible outcomes, and all outcomes are equally-likely, then the probability of any one outcome occurring is $\frac{1}{n}$.

- The probability of an event A, then, is

$$P(A) = \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} = \frac{\text{\# of outcomes in A}}{\text{\# of outcomes in S}} = \frac{|A|}{|S|}$$

- **Example:** Flip a (fair) coin three times and we record head/tail each time. How large is the sample space S?
  $|S| = 8$.
  If event A is when head appears exactly once during 3 times, what are $|A|$ and $P(A)$?

Equally-likely outcomes

- If S is a sample space with n possible outcomes, and all outcomes are equally-likely, then the probability of any one outcome occurring is $\frac{1}{n}$.

- The probability of an event A, then, is

$$P(A) = \frac{1}{n} + \frac{1}{n} + \ldots + \frac{1}{n} = \frac{\text{\# of outcomes in A}}{\text{\# of outcomes in S}} = \frac{|A|}{|S|}$$

- **Example:** Flip a (fair) coin three times and we record head/tail each time. How large is the sample space S?
  $|S| = 8$.
  If event A is when head appears exactly once during 3 times, what are $|A|$ and $P(A)$? $|A| = 3$ and $P(A) = 3/8$.

Complement rule

- Let A be an event with probability $P(A)$.

- Then, the event $\bar{A}$ is the **complement** of the event A. It contains the set of all outcomes in the sample space that are not in A.

- $P(\bar{A})$ is given by
$$P(\bar{A}) = 1 - P(A)$$

- **Example:** Flip a coin 3 times again. What is the probability that there is at least 1 tail?

Complement rule

- Let A be an event with probability $P(A)$.
- Then, the event $\bar{A}$ is the **complement** of the event A. It contains the set of all outcomes in the sample space that are not in A.
- $P(\bar{A})$ is given by
$$P(\bar{A}) = 1 - P(A)$$
- **Example:** Flip a coin 3 times again. What is the probability that there is at least 1 tail?
  Suppose A is the event such that there is no tail. Then, $\bar{A}$ is the event that there is at least 1 tail.

Complement rule

- Let A be an event with probability $P(A)$.
- Then, the event $\bar{A}$ is the **complement** of the event A. It contains the set of all outcomes in the sample space that are not in A.
- $P(\bar{A})$ is given by
$$P(\bar{A}) = 1 - P(A)$$
- **Example:** Flip a coin 3 times again. What is the probability that there is at least 1 tail?
  Suppose A is the event such that there is no tail. Then, $\bar{A}$ is the event that there is at least 1 tail. We have $|A| = 1$ (i.e. head-head-head) and $P(A) = 1/8$. Therefore, $P(\bar{A}) = 7/8$.

Addition rule

- We say two events are **mutually exclusive** if they have no overlap (i.e. they can't both happen at the same time).

- If A and B are mutually exclusive, then the probability that A or B happens is

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

- **Example:** Flip a fair coin 3 times again. A is the event in which head appears **exactly** once. B is the event in which head appears **exactly** twice. What is the probability that head appears once or twice?

Addition rule

- We say two events are **mutually exclusive** if they have no overlap (i.e. they can't both happen at the same time).

- If A and B are mutually exclusive, then the probability that A or B happens is

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

- **Example:** Flip a fair coin 3 times again. A is the event in which head appears **exactly** once. B is the event in which head appears **exactly** twice. What is the probability that head appears once or twice?
  $|A| = 3$ and $P(A) = 3/8$.

Addition rule

- We say two events are **mutually exclusive** if they have no overlap (i.e. they can't both happen at the same time).

- If A and B are mutually exclusive, then the probability that A or B happens is

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

- **Example:** Flip a fair coin 3 times again. A is the event in which head appears **exactly** once. B is the event in which head appears **exactly** twice. What is the probability that head appears once or twice?
  $|A| = 3$ and $P(A) = 3/8$.
  $|B| = 3$ and $P(B) = 3/8$.

Addition rule

- We say two events are **mutually exclusive** if they have no overlap (i.e. they can't both happen at the same time).

- If A and B are mutually exclusive, then the probability that A or B happens is

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

- **Example:** Flip a fair coin 3 times again. A is the event in which head appears **exactly** once. B is the event in which head appears **exactly** twice. What is the probability that head appears once or twice?
  $|A| = 3$ and $P(A) = 3/8$.
  $|B| = 3$ and $P(B) = 3/8$.
  Therefore, $P(A \cup B) = 3/4$. Is there another way to compute this?

Addition rule

- We say two events are **mutually exclusive** if they have no overlap (i.e. they can't both happen at the same time).

- If A and B are mutually exclusive, then the probability that A or B happens is

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

- **Example:** Flip a fair coin 3 times again. A is the event in which head appears **exactly** once. B is the event in which head appears **exactly** twice. What is the probability that head appears once or twice?
  $|A| = 3$ and $P(A) = 3/8$.
  $|B| = 3$ and $P(B) = 3/8$.
  Therefore, $P(A \cup B) = 3/4$. Is there another way to compute this? Yes, via complement rule.

Summary

- Informally, a probability distribution $p : X \to \mathbb{R}$ over some domain X is a function such that $\sum_{x \in X} p(x) = 1$ and $p(x) \geq 0$ for all $x \in X$.

- $\bar{A}$ is the complement of event A. $P(\bar{A}) = 1 - P(A)$.

- Two events A, B are mutually exclusive if they share no outcomes, i.e. they don't overlap. In this case, the probability that A happens or B happens is $P(A \cup B) = P(A) + P(B)$.

Principle of inclusion-exclusion

Principle of inclusion-exclusion

- If events A and B are not mutually exclusive, then the addition rule becomes more complicated.
- In general, if A and B are any two events, then

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

Discussion Question

Each day when you get home from school, there is a

- 0.3 chance your mom is at home
- 0.4 chance your brother is at home
- 0.25 chance that both your mom and brother are at home

When you get home from school today, what is the chance that neither your mom nor your brother are at home?

A) 0.3
B) 0.45
C) 0.55
D) 0.7
E) 0.75

Discussion Question

Each day when you get home from school, there is a

- 0.3 chance your mom is at home
- 0.4 chance your brother is at home
- 0.25 chance that both your mom and brother are at home

When you get home from school today, what is the chance that neither your mom nor your brother are at home?

A) 0.3
B) 0.45
C) 0.55
D) 0.7
E) 0.75

Answer: C) 0.55

A = mom is at home: $P(A) = 0.3$
B = brother is at home: $P(B) = 0.4$
$A \cap B$ = both mom and brother are at home: $p(A \cap B) = 0.25$
$A \cup B$ = mom or brother is at home:

A = mom is at home: $P(A) = 0.3$
B = brother is at home: $P(B) = 0.4$
$A \cap B$ = both mom and brother are at home: $p(A \cap B) = 0.25$
$A \cup B$ = mom or brother is at home:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.4 - 0.25 = 0.45$$

$\overline{A \cup B}$ = neither mom nor brother is at home:

A = mom is at home: $P(A) = 0.3$
B = brother is at home: $P(B) = 0.4$
$A \cap B$ = both mom and brother are at home: $p(A \cap B) = 0.25$
$A \cup B$ = mom or brother is at home:

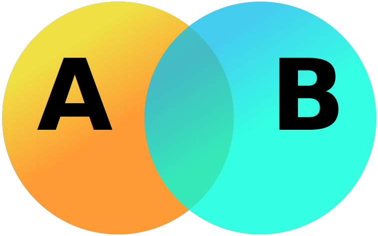
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.4 - 0.25 = 0.45$$

$\overline{A \cup B}$ = neither mom nor brother is at home:

$$P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.45 = 0.55$$

Generalization

Venn diagram:



Sets *A* and *B*:

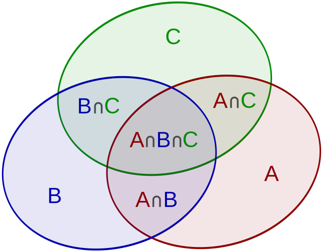
$$|A \cup B| = |A| + |B| - |A \cap B|$$

Events *A* and *B*:

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

Generalization

Venn diagram:



Sets *A*, *B*, and *C*:

$$|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$$

Events *A*, *B* and *C*:

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

Generalization

For n sets $A_1, A_2, .., A_n$:

$$\left|A_1 \cup A_2 \cup \ldots \cup A_n\right| = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \le i_1 < .. < i_k \le n} \left|A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right| \right)$$

For n events $A_1, A_2, .., A_n$:

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = \sum_{k=1}^{n} (-1)^{k+1} \left(\sum_{1 \le i_1 < .. < i_k \le n} P\left(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}\right) \right)$$

Multiplication rules

Multiplication rule and independence

- The probability that events A and B both happen is

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

- $P(B|A)$ is read “the probability that B happens, given that A happened.” It is a **conditional probability**.

- If $P(B|A) = P(B)$, events A and B are **independent**.
  - Intuitively, A and B are independent if knowing that A happened gives you no additional information about event B, and vice versa.

  - For two independent events,

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

Example: rolling a die

Let’s consider rolling a fair 6-sided dice. The results of each dice roll are independent from one another.

- Suppose we roll the dice twice. What is the probability that the faces are 1 and then 2?

Example: rolling a die

Let’s consider rolling a fair 6-sided dice. The results of each dice roll are independent from one another.

- Suppose we roll the dice twice. What is the probability that the faces are 1 and then 2?
  Let X be a random variable denoting the face we get when rolling a the dice. Two times we roll the dice are independent. We have the result:

$$P(X = 1) \cdot P(X = 2) = 1/36$$

Example: rolling a die

- Suppose we roll the dice 3 times. What is the probability that the face 1 never appears in any of the rolls?

Example: rolling a die

- Suppose we roll the dice 3 times. What is the probability that the face 1 never appears in any of the rolls?

$$P(X \neq 1) \cdot P(X \neq 1) \cdot P(X \neq 1) = P(X \neq 1)^3 = \left(\frac{5}{6}\right)^3$$

- Suppose we roll the dice 3 times. What is the probability that the face 1 appears at least once?

Example: rolling a die

- Suppose we roll the dice 3 times. What is the probability that the face 1 never appears in any of the rolls?

$$P(X \neq 1) \cdot P(X \neq 1) \cdot P(X \neq 1) = P(X \neq 1)^3 = \left(\frac{5}{6}\right)^3$$

- Suppose we roll the dice 3 times. What is the probability that the face 1 appears at least once?

$$1 - P(X \neq 1)^3 = 1 - \left(\frac{5}{6}\right)^3$$

Example: rolling a die

- Suppose we roll the dice n times. What is the probability that only the faces 2, 4, and 5 appear?

Example: rolling a die

- Suppose we roll the dice n times. What is the probability that only the faces 2, 4, and 5 appear?

$$\left(\frac{1}{2}\right)^n$$

- Suppose we roll the dice twice. What is the probability that the two rolls have different faces?

Example: rolling a die

- Suppose we roll the dice n times. What is the probability that only the faces 2, 4, and 5 appear?

$$\left(\frac{1}{2}\right)^n$$

- Suppose we roll the dice twice. What is the probability that the two rolls have different faces?

$$\frac{5}{6}$$

Conditional probability

Conditional probability

- The probability of an event may **change** if we have additional information about outcomes.

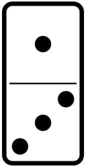
- Starting with the multiplication rule, $P(A \cap B) = P(A)P(B|A)$, we have that

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

  assuming that $P(A) > 0$.

Example: dominoes (source: 538)

In a set of dominoes, each tile has two sides with a number of dots on each side: zero, one, two, three, four, five, or six. There are 28 total tiles, with each number of dots appearing alongside each other number (including itself) on a single tile.



Example: dominoes

Question 1: What is the probability of drawing a “double” from a set of dominoes — that is, a tile with the same number on both sides?

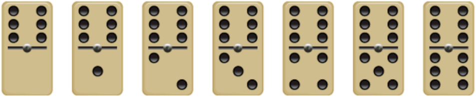
Example: dominoes

Question 1: What is the probability of drawing a “double” from a set of dominoes — that is, a tile with the same number on both sides?

$$\frac{7}{28} = \frac{1}{4}$$

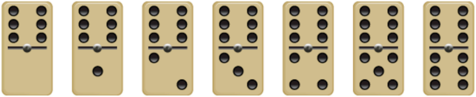
Example: dominoes

Question 2: Now your friend picks a random tile from the set and tells you that at least one of the sides is a 6. What is the probability that your friend's tile is a double, with 6 on both sides?



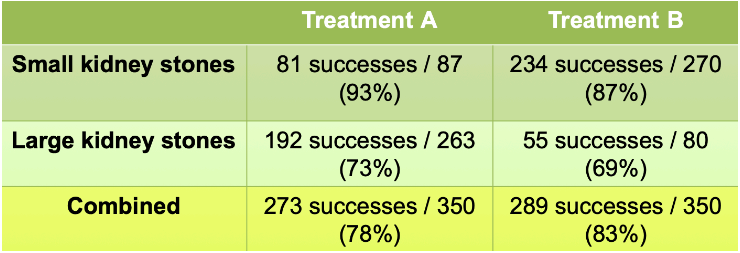
Example: dominoes

Question 2: Now your friend picks a random tile from the set and tells you that at least one of the sides is a 6. What is the probability that your friend's tile is a double, with 6 on both sides?



$$\frac{1}{7}$$

Simpson’s Paradox (source: nih.gov)



	Treatment A	Treatment B
Small kidney stones	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones	192 successes / 263 (73%)	55 successes / 80 (69%)
Combined	273 successes / 350 (78%)	289 successes / 350 (83%)

Discussion Question

Which treatment is better?

A) Treatment A for all cases.
B) Treatment B for all cases.
C) Treatment A for small stones and B for large stones.
D) Treatment A for large stones and B for small stones.

Simpson's Paradox (source: nih.gov)

Let A be a random variable taking value True if treatment A is effective, or False otherwise. Let X be a random variable taking values, small or large, denoting the size of the kidney stone.

By the **Law of Total Probability**, We have:

$$P(A = \text{True}) = P(A = \text{True}|X = \text{small}) \cdot P(X = \text{small}) +$$

$$P(A = \text{True}|X = \text{large}) \cdot P(X = \text{large})$$

Theat is equal to:

$$P(A = \text{True}) = \frac{81}{87} \cdot \frac{87}{350} + \frac{192}{263} \cdot \frac{263}{350} = \frac{273}{350} = 78\%$$

Simpson’s Paradox (source: nih.gov)

Let B be a random variable taking value True if treatment B is effective, or False otherwise. Let Y be a random variable taking values, small or large, denoting the size of the kidney stone. We use Y not X because for each experiment for each treatment, 350 different people.

By the **Law of Total Probability**, We have:

$$P(B = \text{True}) = P(B = \text{True}|Y = \text{small}) \cdot P(Y = \text{small}) +$$

$$P(B = \text{True}|Y = \text{large}) \cdot P(Y = \text{large})$$

Theat is equal to:

$$P(B = \text{True}) = \frac{234}{270} \cdot \frac{270}{350} + \frac{55}{80} \cdot \frac{80}{350} = \frac{289}{350} = 83\%$$

Simpson's Paradox (source: nih.gov)

It is called a **paradox** because:

$$P(B = \text{True} | Y = \text{small}) < P(A = \text{True} | X = \text{small})$$

$$P(B = \text{True} | Y = \text{large}) < P(A = \text{True} | X = \text{large})$$

But

$$P(B = \text{True}) > P(A = \text{True}).$$

The problem lies in the fact that distributions of X and Y are approximations (based on sampling) of the actual distribution of patients with small or large kidney stones.

How can we fix this?

We need to make a better approximation of the distribution of patients with small or large stones.

Simpson's Paradox (source: nih.gov)

There are totally 700 = 350 + 350 patients in which:

- 87 + 270 = 357 have small stones: 357/700 = 51%, denoted by $P(\text{small})$

- 263 + 80 = 343 have large stones: 343/700 = 49%, denoted by $P(\text{large})$

Simpson's Paradox (source: nih.gov)

There are totally 700 = 350 + 350 patients in which:

- 87 + 270 = 357 have small stones: 357/700 = 51%, denoted by $P(\text{small})$
- 263 + 80 = 343 have large stones: 343/700 = 49%, denoted by $P(\text{large})$

By the **Law of Total Probability**, we have the actual effectiveness of A is:

$$P(A = \text{True}) \approx P(A = \text{True}|\text{small})\cdot P(\text{small}) + P(A = \text{True}|\text{large})\cdot P(\text{large})$$

That equals to:

$$P(A = \text{True}) \approx 93\% \cdot 51\% + 73\% \cdot 49\% = 83.2\%$$

Simpson's Paradox (source: nih.gov)

By the **Law of Total Probability**, we have the actual effectiveness of B is:

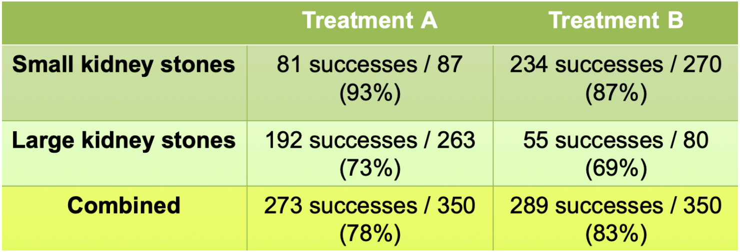
$P(B = \text{True}) \approx P(B = \text{True}|\text{small}) \cdot P(\text{small}) + P(B = \text{True}|\text{large}) \cdot P(\text{large})$

That equals to:

$$P(B = \text{True}) \approx 87\% \cdot 51\% + 69\% \cdot 49\% = 81.24\%$$

Now, we can conclude that treat A is better in general.

Simpson’s Paradox (source: nih.gov)



	Treatment A	Treatment B
Small kidney stones	81 successes / 87 (93%)	234 successes / 270 (87%)
Large kidney stones	192 successes / 263 (73%)	55 successes / 80 (69%)
Combined	273 successes / 350 (78%)	289 successes / 350 (83%)

Simpson’s Paradox occurs when an association between two variables exists when the data is divided into subgroups, but reverses or disappears when the groups are combined.

- See more in DSC 80.

Summary, next time

Summary

- $\bar{A}$ is the complement of event A. $P(\bar{A}) = 1 - P(A)$.

- Two events A, B are mutually exclusive if they share no outcomes, i.e. they don't overlap. In this case, the probability that A happens or B happens is $P(A \cup B) = P(A) + P(B)$.

- More generally, for any two events, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

- The probability that events A and B both happen is $P(A \cap B) = P(A)P(B|A)$.
  - $P(B|A)$ is the probability that B happens given that you know A happened.
  - Through re-arranging, we see that $P(B|A) = \frac{P(A \cap B)}{P(B)}$.