Wednesday, September 20 **Fall 2023**

Section 2: Conditional Probability

Srihari Ganesh

Based section template and practice problems by Rachel Li and Ginnie Ma '23
Practice problem selection also aided by Liz Li and Karina Wang

1 Summary

Notation 1. See that we use commas and intersections interchangeably (i.e., $P(A, B, C) = P(A \cap B \cap C)$).

Remark 2. A rough workflow for solving probability problems:

1. Define events for every aspect of the problem (e.g., "A = the event that it rains tomorrow, B = the event that it rained today")
2. Write out the probabilities that you are given in the problem using notation (e.g., "$P(A|B) = 1/2$, $P(B) = 1/4$).
3. Write the probability that you want to calculate using notation (e.g., we want to calculate the unconditional probability that it rains tomorrow, $P(A)$).
4. Figure out how the tools we have learned allow you to utilize the probabiliies that you do know (step 2) to calculate the probabilities that you don't know (step 3).

1.1 Definition of Probability

Definition 3 (Axioms of Probability)**.** With sample space S,

1. $P(S) = 1, P(\emptyset) = 0$.
2. For $A_1, A_2, \dots,$ that partition B (this can be finite or infinite),

$$P(B) = \sum_{j=1}^{\infty} P(A_j)$$

Result 4 (Probability of a complement)**.** For event A,

$$P(A) = 1 - P(A^c)$$

Result 5 (Probability of a union)**.** For events A and B,

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

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

Result 6 (Principle of Inclusion-Exclusion)**.** For events $A_1, \dots, A_n$,

$$P(\bigcup_{i=1}^{n} A_i) = \sum_i P(A_i) - \sum_{i<j} P(A_i \cap A_j)$$

$$+ \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \cdots + (-1)^{n+1} P(\bigcap_{i=1}^{n} A_i).$$

1.2 Conditional Probability

Definition 7 (Conditional probability)**.** The probability of event A given that we know B occured is

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

...with extra conditioning:

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

1.3 Conditional Probability Tools

Remark 8 (First-step analysis)**.** If you ever need to solve a problem involving a sequence of things (like a game with many turns, or a random walk, or so on) and are stuck, try first-step analysis: conditioning what happens after the first step. You'll often be able to get a recursive equation that is easier to solve.

Result 9 (Probability of an intersection)**.**

$$P(A_1, A_2, \dots A_n) = P(A_1)P(A_2|A_1)\cdots P(A_n|A_1, \dots, A_{n-1})$$

$$= P(A_n)P(A_{n-1}|A_n)\cdots P(A_1|A_2, \dots, A_n),$$

$$= [\text{chaining in any order that is convenient for you}].$$

...with extra conditioning:

$$P(A_1, A_2, \dots A_n|C) = P(A_1|C)P(A_2|A_1, C)\cdots P(A_n|A_1, \dots, A_{n-1}, C)$$

Result 10 (Law of Total Probability (LOTP))**.** for events $A_1, A_2, \dots, A_n$ that partition S, we can find $P(B)$ by

$$P(B) = P(B, A_1) + P(B, A_2) + \cdots + P(B, A_n)$$

$$= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + \cdots + P(B|A_n)P(A_n).$$

We pick $A_1, A_2, \dots, A_n$ to "condition on what we wish we knew." These are situations where you don't know $P(B)$, but you know $P(B|A_1)$, $(B|A_2)$, etc.

...with extra conditioning:

$$P(B|C) = P(B, A_1|C) + P(B, A_2|C) + \cdots + P(B, A_n|C)$$

$$= P(B|A_1, C)P(A_1|C) + P(B|A_2, C)P(A_2|C) + \cdots + P(B|A_n, C)P(A_n|C).$$

Result 11 (Bayes' Rule)**.** for events A, B, if we want to calculate $P(B|A)$ but can only know how to calculate $P(A|B)$,

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

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

where we commonly expand the denominator using the Law of Total Probability (LOTP).

...with extra conditioning:

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

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

1.4 Independence

Definition 12 (Independence)**.** A, B are defined to be independent if

$$P(A, B) = P(A)P(B).$$

Note that if A, B are independent, then A^c, B^c are independent, as are A, B^c, and so on; in generality, for functions f, g, the events $f(A), g(B)$ are independent.

A set of events $A_1, A_2, \ldots, A_n$ is independent if any subset of the events $A_{j_1}, \ldots, A_{j_k}$ follows the equation.

$$P(A_{j_1}, \ldots, A_{j_k}) = P(A_{j_1}) \cdots P(A_{j_k}).$$

Basically, for any combination of independent events, we should be able to factor out the probabilities.

Definition 13 (Conditional independence)**.** A, B are conditionally independent given C if

$$P(A, B|C) = P(A|C)P(B|C).$$

VERY IMPORTANT:

- Disjointness and independence are not the same: in fact, if A, B are disjoint, then they are **VERY DEPENDENT**, because you know if A happens, then B definitely didn't happen!
- Independence and conditional independence are not the same/do not imply each other. There is no guarantee that independent events are conditionally independent, or vice versa.

2 Practice Problems

Warning for this week: these are some time-consuming problems! I do not expect you to be able to solve all of these in the allotted section time!

1. **[Probability results, maybe some LOTP]** For each of the following, fill in the blank with a $\geq, \leq, =$, or ?. You can reason mathematically or with a picture (e.g., Venn diagram)

$$\begin{aligned} P(B^c) &\;\_\_\_\; P(A) \\ P(A \cup B) &\;\_\_\_\; 1 - P(A^c \cap B^c) \\ P(A) &\;\_\_\_\; P(A \cap B^c). \end{aligned}$$

Solution

- $P(B^c)$? $P(A)$. We can break each side down into disjoint pieces (similar to LOTP) and get that

$$\begin{aligned} P(B^c) &= P(A^c \cap B^c) + P(A \cap B^c) \\ P(A) &= P(A \cap B) + P(A \cap B^c). \end{aligned}$$

  With $P(A^c \cap B^c)$ and $P(A \cap B)$ being disjoint, we can't say anything about the relationship between their probabilities.
- $P(A \cup B) = 1 - P(A^c \cap B^c)$. This follows by complementary counting and DeMorgan's law to calculate the complement:

$$\begin{aligned} P(A \cup B) &= 1 - P((A \cup B)^c) \\ &= 1 - P(A^c \cap B^c). \end{aligned}$$

- $P(A) \geq P(A \cap B^c)$. We can break down A into disjoint pieces to get

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

  where the fact that probabilities are nonzero gives $P(A \cap B)$ to yield the inequality.

2. **[LOTP]** You have a project due with two intermediate milestones (i.e., checkpoints that help you tell whether you're on track to complete on time). Let A_1 be the event that you complete your first milestone on time, A_2 be the event that you complete your second milestone on time, and A_3 be the event that you complete your project on time. For $j = 1, 2$,

$$P(A_{j+1}|A_j) = 0.8$$
$$P(A_{j+1}|A_j^c) = 0.3.$$

Also assume that if we know your status on the second milestone (whether you completed it on time or not), the first milestone is no longer relevant to whether you complete the project on time.

(a) State the previous paragraph in terms of independence or conditional independence.

> **Solution**
> Given the status of the second milestone, the status of the first milestone gives us no information about the timely completion of the project. This means that given A_2, A_1 and A_3 are conditionally independent; similarly, given A_2^c, A_1 and A_3 are conditionally independent.

(b) Find the probability that you complete the project on time, given that you complete the first milestone on time.

> **Solution**
> We want to calculate $P(A_3|A_1)$. We don't know how to calculate this, but we do know things like $P(A_3|A_2), P(A_2|A_1)$, and so on. Concisely, we wish we knew if we hit the second milestone on time (so whether or not A_2 occurs). Let's use LOTP (with extra conditioning!) to expand that:
>
> $$P(A_3|A_1) = P(A_3|A_1, A_2)P(A_2|A_1) + P(A_3|A_1, A_2^c)P(A_2^c|A_1).$$
>
> Due to the conditional independence between A_3, A_1 given A_2 or A_2^c, conditioning on both A_1, A_2 is equivalent to conditioning on just A_2 (and similarly for A_2^c). So we can reduce this to
>
> $$\begin{aligned} P(A_3|A_1) &= P(A_3|A_2)P(A_2|A_1) + P(A_3|A_2^c)P(A_2^c|A_1) \\ &= (0.8)(0.8) + (0.3)(1 - 0.8) \\ &= 0.64 + 0.06 \\ &= \boxed{0.7} \end{aligned}$$

3. **[Bayes' Rule]** It's that time of year for Datamatch! A recent survey states that if a participant likes their match, there is a $\frac{3}{4}$ chance they will match back, and if they don't like their match, there is a $\frac{1}{2}$ chance they will match back (for the fooood). Let's be honest, you're a stunner so you assign a prior that your match likes you of $\frac{2}{3}$. What's the probability that your match likes you given that they matched back?

> **Solution**
> Let M be the event that your match, matched back; define L be the event that your match

likes you. We are given that $P(L) = \frac{2}{3}, P(M|L) = \frac{3}{4}, P(M|L^c) = \frac{1}{2}$. This calls for Bayes rule, since we know M given L but want to find L given M.

$$\begin{aligned} P(L|M) &= \frac{P(M|L)P(L)}{P(M|L)P(L) + P(M|L^c)P(L^c)} \\ &= \frac{\frac{3}{4}\frac{2}{3}}{\frac{3}{4}\frac{2}{3} + \frac{1}{2}\frac{1}{3}} \\ &= \frac{\frac{1}{2}}{\frac{1}{2} + \frac{1}{6}} \\ &= \boxed{\frac{3}{4}} \end{aligned}$$

4. **[LOTP/first-step-analysis][If you have extra time!]** Two players (A, B) take turns tossing a fair coin, with A going first. The sequence of heads and tails is recorded, with H representing heads and T representing tails. If a head is followed by a tail, the player who flipped the tail wins. What is the probability A wins?

Solution
Let W be the event that A wins. We want to calculate $P(W)$.
This problem seems difficult to reason about without any information. Let's use LOTP, conditioning on whether the first toss is heads, to see what happens. Define H_i to be the event that the i^{th} toss is heads.

$$P(W) = P(W|H_1)P(H_1) + P(W|H_1^c)P(H_1^c)$$

We know $P(H_1) = P(H_1^c) = 1/2$ because the coin is fair. Now let's consider the cases we conditioned on:

- If H_1 occurs, then we still want to consider another step:

  $$P(W|H_1) = P(W|H_1, H_2)P(H_2|H_1) + P(W|H_1, H_2^c)P(H_2^c|H_1).$$

  The coin tosses are fair, so they should be independent, so $P(H_2|H_1) = P(H_2) = 1/2$ and similarly for $P(H_2^c|H_1) = 1/2$. Now we consider that B wins if H_1, H_2^c occur (since this is an HT pattern, but B tosses the T), so $P(W|H_1, H_2^c) = 0$. For the other case, H_1, H_2, we actually can see that $P(W|H_1, H_2) = P(W^c|H_1)$. This comes from the fact that only the previous toss matters in the game — any older tosses are irrelevant. We get a new relation that $P(W|H_1) = (1 - P(H_1))(1/2)$, which can be solved to get $P(W|H_1) = 1/3$.
  Another (potentially more elegant) way to think about the problem is that, once H_1 occurs, whoever tosses T next will win the game (and player B gets to go first). Try using similar first-step analysis to see the probability that A tosses the first T.
- If H_1^c occurs, then player B is in basically the same situation as player A - previous tails cannot lead to winning combinations. Thus,

  $$P(W|H_1^c) = P(W^c).$$

We plug these back in to get

$$\begin{aligned} P(W) &= P(W|H_1)P(H_1) + P(W|H_1^c)P(H_1^c) \\ &= \frac{1}{3}\frac{1}{2} + P(W^c)\frac{1}{2} \\ P(W) &= \frac{1}{6} + \frac{1}{2}(1 - P(W)). \end{aligned}$$

This can be solved to get $\boxed{P(W) = \frac{4}{9}}$.