Discrete Structures for Computer Science

William Garrison
bill@cs.pitt.edu
6311 Sennott Square

Lecture #25: Bayes' Theorem

University of Pittsburgh

Based on materials developed by Dr. Adam Lee

Today's Topics

- Bayes' Theorem
  - What do we compute conditional probabilities with incomplete information?

Conditional Probability

Definition: Let E and F be events with p(F) > 0. The conditional probability of E given F, denoted p(E | F), is defined as:

$$p(E \mid F) = \frac{p(E \cap F)}{p(F)}$$

Intuition:

- Think of the event F as reducing the sample space that can be considered
- The numerator looks at the likelihood of the outcomes in E that overlap those in F
- The denominator accounts for the reduction in sample size indicated by our prior knowledge that F has occurred

Bayes' Theorem

Bayes' Theorem allows us to relate the conditional and marginal probabilities of two random events.

?

In English: Bayes' Theorem will help us assess the probability that an event occurred given only partial evidence.

Doesn't our formula for conditional probability do this already?

We can't always use this formula directly…

A Motivating Example

Suppose that a certain opium test correctly identifies a person who uses opiates as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. If a company suspects that 0.5% of its employees are opium users, what is the probability that an employee that tests positive for this drug is actually a user?

Question: Can we use our simple conditional probability formula?

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

X is a user *X tested positive*

The 1,000 foot view…

In situations like those on the last slide, Bayes' theorem can help!

Essentially, Bayes' theorem will allow us to calculate P(E|F) assuming that we know (or can derive):

- $P(E)$ *Probability that X is a user*
- $P(F|E)$ *Test success rate*
- $P(F|\bar{E})$

Test false positive rate

Probability that X is an opium user given a positive test

Returning to our earlier example:

- Let E = "Person X is an opium user"
- Let F = "Person X tested positive for opium"

It sounds like Bayes' Theorem could help in this case…

New Notation

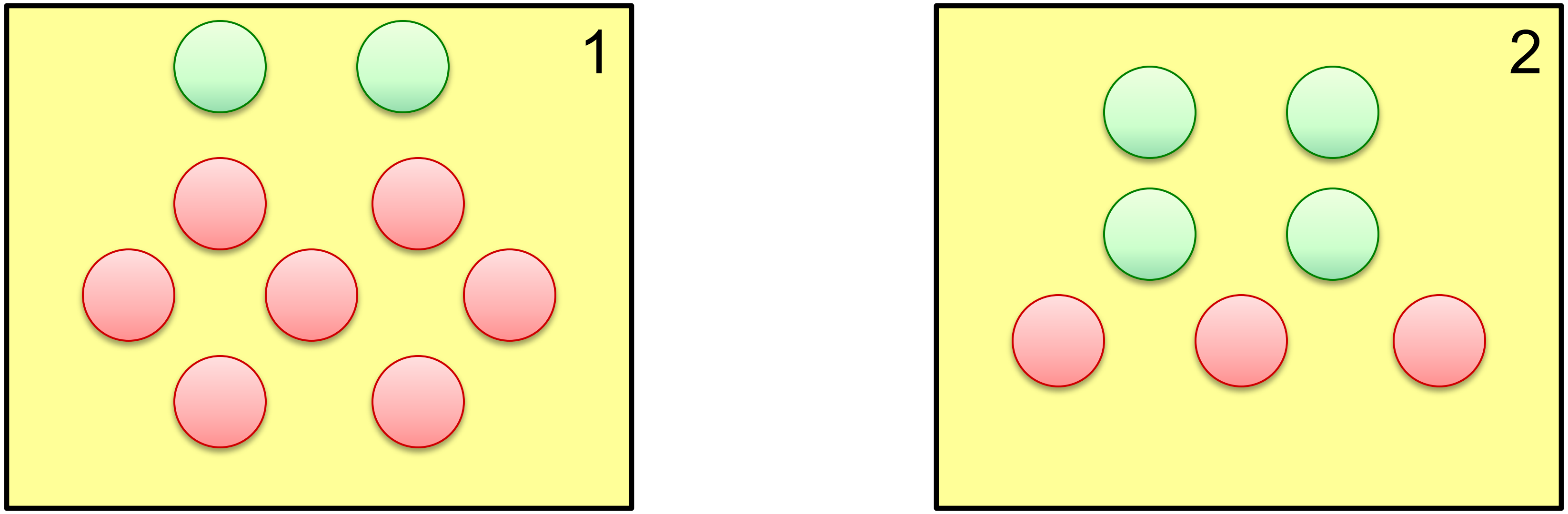
To simplify expressions, we will use the notation E^C to denote the complementary event of E

That is:

$$\overline{E} = E^C$$

A Simple Example

We have two boxes. The first contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects a ball by first choosing a box at random. He then selects one of the balls from that box at random. If Bob has selected a red ball, what is the probability that he took it from the first box?



Picking the problem apart…

First, let's define a few events relevant to this problem:

- Let E = Bob has chosen a red ball
- By definition E^C = Bob has chosen a green ball
- Let F = Bob chose his ball from this first box
- Therefore, F^C = Bob chose his ball from the second box

We want to find the probability that Bob chose from the first box, given that he picked a red ball. That is, we want $p(F|E)$.

Goal: Given that $p(F|E) = p(F \cap E)/p(E)$, use what we know to derive $p(F \cap E)$ and $p(E)$.

What do we know?

We have two boxes. The first contains two green balls and seven red balls. The second contains four green balls and three red balls. Bob selects a ball by first choosing a box at random. He then selects one of the balls from that box at random. If Bob has selected a red ball, what is the probability that he took it from the first box?

Statement: *Bob selects a ball by first choosing a box at random*

- Bob is equally likely to choose the first box, or the second box
- $p(F) = p(F^C) = 1/2$

Statement: *The first contains two green balls and seven red balls*

- The first box has nine balls, seven of which are red
- $p(E|F) = 7/9$

Statement: *The second contains four green balls and three red balls*

- The second box contains seven balls, three of which are red
- $p(E|F^C) = 3/7$

Now, for a little algebra…

Recall:
- $p(F) = p(F^C) = 1/2$
- $p(E|F) = 7/9$
- $p(E|F^C) = 3/7$

The end goal: Compute $p(F|E) = p(F \cap E)/p(E)$

Note that $p(E|F) = p(E \cap F)/p(F)$

- If we multiply by $p(F)$, we get $p(E \cap F) = p(E|F)\, p(F)$
- Further, we know that $p(E|F) = 7/9$ and $p(F) = 1/2$
- So $p(E \cap F) = 7/9 \times 1/2 = 7/18$

Similarly, $p(E \cap F^C) = p(E|F^C)\, p(F^C) = 3/7 \times 1/2 = 3/14$

Observation: $E = (E \cap F) \cup (E \cap F^C)$

- This means that $p(E) = p(E \cap F) + p(E \cap F^C)$
- $= 7/18 + 3/14$
- $= 49/126 + 27/126$
- $= 76/126$
- $= 38/63$

Denouement

The end goal: Compute $p(F|E) = p(F \cap E)/p(E)$

Recall:
- $p(F) = p(F^C) = 1/2$
- $p(E|F) = 7/9$
- $p(E|F^C) = 3/7$
- $p(E \cap F) = 7/18$
- $p(E) = 38/63$

So, $p(F|E) = (7/18) \,/\, (38/63) \approx 0.645$

How did we get here?

1. Extract what we could from the problem definition itself
2. Rearrange terms to derive $p(F \cap E)$ and $p(E)$
3. Use our trusty definition of conditional probability to do the rest!

The reasoning that we used in the last problem essentially derives Bayes' Theorem for us!

Bayes' Theorem: Suppose that E and F are events from some sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then:

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{P(E \mid F)p(F) + p(E \mid F^C)p(F^C)}$$

Proof:

- The definition of conditional probability says that
  - $p(F|E) = p(F \cap E)/p(E)$
  - $p(E|F) = p(E \cap F)/p(F)$
- This means that
  - $p(E \cap F) = p(F|E)p(E)$
  - $p(E \cap F) = p(E|F)p(F)$
- So $p(F|E)p(E) = p(E|F)p(F)$
- Therefore, $p(F|E) = p(E|F)p(F)/p(E)$

Proof (continued)

Note: To finish, we must prove $p(E) = p(E \mid F)p(F) + p(E \mid F^C)p(F^C)$

- Observe that $E = E \cap S$
- $\quad = E \cap (F \cup F^C)$
- $\quad = (E \cap F) \cup (E \cap F^C)$
- Note also that $(E \cap F)$ and $(E \cap F^C)$ are disjoint (i.e., no x can be in both F and F^C)
- This means that $p(E) = p(E \cap F) + p(E \cap F^C)$
- We already have shown that $p(E \cap F) = p(E|F)p(F)$
- Further, since $p(E \mid F^C) = p(E \cap F^C)/p(F^C)$, we have that $p(E \cap F^C) = p(E|F^C)p(F^C)$
- So $p(E) = p(E \cap F) + p(E \cap F^C) = p(E|F)p(F) + p(E|F^C)p(F^C)$

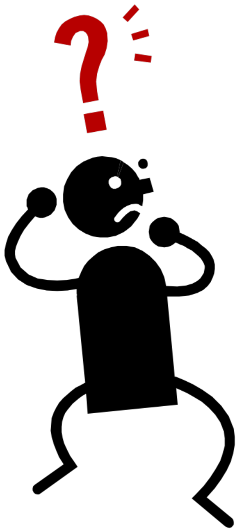
Putting everything together, we get:

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{P(E \mid F)p(F) + p(E \mid F^C)p(F^C)}$$

❑

And why is this useful?

In a nutshell, Bayes' Theorem is useful if you want to find p(F|E), but you don't know p(E ∩ F) or p(E).



Here's a general solution tactic

Step 1: Identify the independent events that are being investigated. For example:

- F = Bob chooses the first box, F^C = Bob chooses the second box
- E = Bob chooses a red ball, E^C = Bob chooses a green ball

Step 2: Record the probabilities identified in the problem statement. For example:

- $p(F) = p(F^C) = 1/2$
- $p(E|F) = 7/9$
- $p(E|F^C) = 3/7$

Step 3: Plug into Bayes' formula and solve

Example: Pants and Skirts

Suppose there is a co-ed school having 60% boys and 40% girls as students. The girl students wear pants or skirts in equal numbers; the boys all wear pants. An observer sees a (random) student from a distance; all they can see is that this student is wearing pants. What is the probability this student is a girl?

Step 1: Set up events

- E = X is wearing pants
- E^C = X is wearing a skirt
- F = X is a girl
- F^C = X is a boy

Step 2: Extract probabilities from problem definition

- $p(F) = 0.4$
- $p(F^C) = 0.6$
- $p(E|F) = p(E^C|F) = 0.5$
- $p(E|F^C) = 1$

Pants and Skirts (continued)

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{P(E \mid F)p(F) + p(E \mid F^C)p(F^C)}$$

Recall:

- $p(F) = 0.4$
- $p(F^C) = 0.6$
- $p(E|F) = p(E^C|F) = 0.5$
- $p(E|F^C) = 1$

Step 3: Plug in to Bayes' Theorem

- $p(F|E) = (0.5 \times 0.4)/(0.5 \times 0.4 + 1 \times 0.6)$
- $= 0.2/0.8$
- $= 1/4$

Conclusion: There is a 25% chance that the person seen was a girl, given that they were wearing pants.

Drug screening, revisited

Suppose that a certain opium test correctly identifies a person who uses opiates as testing positive 99% of the time, and will correctly identify a non-user as testing negative 99% of the time. If a company suspects that 0.5% of its employees are opium users, what is the probability that an employee that tests positive for this drug is actually a user?

Step 1: Set up events

- F = X is an opium user
- F^C = X is not an opium user
- E = X tests positive for opiates
- E^C = X tests negative for opiates

Step 2: Extract probabilities from problem definition

- $p(F) = 0.005$
- $p(F^C) = 0.995$
- $p(E|F) = 0.99$
- $p(E|F^C) = 0.01$

Drug screening (continued)

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{P(E \mid F)p(F) + p(E \mid F^C)p(F^C)}$$

Recall:
- $p(F) = 0.005$
- $p(F^C) = 0.995$
- $p(E|F) = 0.99$
- $p(E|F^C) = 0.01$

Step 3: Plug in to Bayes' Theorem

- $p(F|E) = (0.99 \times 0.005)/(0.99 \times 0.005 + 0.01 \times 0.995)$
- $= 0.3322$

Conclusion: If an employee tests positive for opiate use, there is only a 33% chance that they are actually an opium user!

In-class exercises

Suppose that 1 person in 100,000 has a particular rare disease. A diagnostic test is correct 99% of the time when given to someone with the disease, and is correct 99.5% of the time when given to someone without the disease.

Problem 1: Calculate the probability that someone who tests positive for the disease actually has it.

Problem 2: Calculate the probability that someone who tests negative for the disease does not have the disease.

Application: Spam filtering

Definition: Spam is unsolicited bulk email

I didn't ask for it, I probably don't want it

Sent to lots of people…

In recent years, spam has become increasingly problematic. For example, in 2015, spam accounted for ~50% of all email messages sent.

To combat this problem, people have developed spam filters based on Bayes' theorem!

How does a Bayesian spam filter work?

Essentially, these filters determine the probability that a message is spam, given that it contains certain keywords.

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{P(E \mid F)p(F) + p(E \mid F^C)p(F^C)}$$

Message is spam

Message contains questionable keyword

In the above equation:

- $p(E|F)$ = Probability that our keyword occurs in spam messages
- $p(E|F^C)$ = Probability that our keyword occurs in legitimate messages
- $p(F)$ = Probability that an arbitrary message is spam
- $p(F^C)$ = Probability that an arbitrary message is legitimate

Question: How do we derive these parameters?

We can learn these parameters by examining historical email traces

Imagine that we have a corpus of email messages…

We can ask a few intelligent questions to learn the parameters of our Bayesian filter:

- How many of these messages do we consider spam?
- In the spam messages, how often does our keyword appear?
- In the good messages, how often does our keyword appear?

Aside: This is what happens when you click the “mark as spam” button in your email client!

Given this information, we can apply Bayes’ theorem!

Filtering spam using a single keyword

Suppose that the keyword “Rolex” occurs in 250 of 2000 known spam messages, and in 5 of 1000 known good messages. Estimate the probability that an incoming message containing the word “Rolex” is spam, assuming that it is equally likely that an incoming message is spam or not spam. If our threshold for classifying a message as spam is 0.9, will we reject this message?

Step 1: Define events

- F = message is spam
- F^C = message is good
- E = message contains the keyword “Rolex”
- E^C = message does not contain the keyword “Rolex”

Step 2: Gather probabilities from the problem statement

- $p(F) = p(F^C) = 0.5$
- $p(E|F) = 250/2000 = 0.125$
- $p(E|F^C) = 5/1000 = 0.005$

Spam Rolexes (continued)

$$p(F \mid E) = \frac{p(E \mid F)p(F)}{p(E \mid F)p(F) + p(E \mid F^C)p(F^C)}$$

Recall:
- $p(F) = p(F^C) = 0.5$
- $p(E|F) = 0.125$
- $p(E|F^C) = 0.005$

Step 3: Plug in to Bayes' Theorem
- $p(F|E) = (0.125 \times 0.5)/(0.125 \times 0.5 + 0.005 \times 0.5)$
- $= 0.125/(0.125 + 0.005)$
- ≈ 0.962

Conclusion: Since the probability that our message is spam given that it contains the string "Rolex" is approximately 0.962 > 0.9, we will discard the message.

Problems with this simple filter

How would you choose a single keyword/phrase to use?

- "All natural"
- "Nigeria"
- "Click here"
- …

Users get upset if false positives occur, i.e., if legitimate messages are incorrectly classified as spam

- When was the last time you checked your spam folder?

How can we fix this?

- Choose keywords so p(spam | keyword) is very high or very low
- Filter based on multiple keywords

Specifically, we want to develop a Bayesian filter that tells us $p(F \mid E_1 \cap E_2)$

First, some assumptions

1. Events E_1 and E_2 are independent
2. The events $E_1|F$ and $E_2|F$ are independent
3. $p(F) = p(F^C) = 0.5$

Now, let's derive formula for this $p(F \mid E_1 \cap E_2)$

By Bayes' theorem

$$p(F \mid E_1 \cap E_2) = \frac{p(E_1 \cap E_2 \mid F)p(F)}{p(E_1 \cap E_2 \mid F)p(F) + p(E_1 \cap E_2 \mid F^C)p(F^C)}$$

Assumption 3

$$= \frac{p(E_1 \cap E_2 \mid F)}{p(E_1 \cap E_2 \mid F) + p(E_1 \cap E_2 \mid F^C)}$$

Assumptions 1 and 2

$$= \frac{p(E_1 \mid F)p(E_2 \mid F)}{p(E_1 \mid F)p(E_2 \mid F) + p(E_1 \mid F^C)p(E_2 \mid F^C)}$$

Spam filtering on two keywords

Suppose that we train a Bayesian spam filter on a set of 2000 spam messages and 1000 messages that are not spam. The word "stock" appears in 400 spam messages and 60 good messages, and the word "undervalued" appears in 200 spam messages and 25 good messages. Estimate the probability that a message containing the words "stock" and "undervalued" is spam. Will we reject this message if our spam threshold is set at 0.9?

Step 1: Set up events

- F = message is spam, F^C = message is good
- E_1 = message contains the word "stock"
- E_2 = message contains the word "undervalued"

Step 2: Identify probabilities

- $P(E_1|F) = 400/2000 = 0.2$
- $p(E_1|F^C) = 60/1000 = 0.06$
- $p(E_2|F) = 200/2000 = 0.1$
- $p(E_2|F^C) = 25/1000 = 0.025$

Two keywords (continued)

$$p(F \mid E_1 \cap E_2) = \frac{p(E_1 \mid F)p(E_2 \mid F)}{p(E_1 \mid F)p(E_2 \mid F) + p(E_1 \mid F^C)p(E_2 \mid F^C)}$$

Recall:
- $p(E_1|F) = 0.2$
- $p(E_1|F^C) = 0.06$
- $p(E_2|F) = 0.1$
- $p(E_2|F^C) = 0.025$

Step 3: Plug in to Bayes' Theorem

- $p(F|E_1 \cap E_2) = (0.2 \times 0.1)/(0.2 \times 0.1 + 0.06 \times 0.025)$
- $= 0.02/(0.02 + 0.0015)$
- ≈ 0.9302

Conclusion: Since the probability that our message is spam given that it contains the strings "stock" and "undervalued" is $\approx 0.9302 > 0.9$, we will reject this message.

In-class exercises

Problem 3: A business records incoming emails for 1 week and collects 1,000 spam messages and 400 non-spam messages. The word "opportunity" appears in 175 spam messages and 20 non-spam messages. Assuming this week's emails were typical, should an incoming message be labeled as spam if it contains the word "opportunity" and the threshold for rejecting is 0.9?

Problem 4: Suppose that a Bayesian spam filter is trained on a set of 10,000 spam messages and 5,000 messages that are not spam. The word "enhancement" appears in 1,500 spam messages and 20 non-spam messages, while the word "herbal" appears in 800 spam messages and 200 non-spam messages. Estimate the probability that a received message containing both the words "enhancement" and "herbal" is spam. You may assume that 50% of emails are spam.

Final Thoughts

- Conditional probability is very useful

- Bayes' theorem
  - Helps us assess conditional probabilities
  - Has a range of important applications

- Next time:
  - Expected values and variance (Section 7.4)