CS109 January 24, 2024

Section 2: Conditional Probabililty and Bayes

Chris Piech, Mehran Sahami, Jerry Cain, Lisa Yan, and numerous CS109 CA's.

Overview of Section Materials

The warm-up questions provided will help students practice concepts introduced in lectures. The section problems are meant to apply these concepts in more complex scenarios similar to what you will see in problem sets and exams. In fact, many of them **are** old exam questions.

Before you leave lab, make sure you click here so that you're marked as having attended this week's section. The CA leading your discussion section can enter the password needed once you've submitted.

Warm-ups

1. Definitions: Cite Bayes' Theorem. Can you explain why $P(A|B)$ is different than $P(B|A)$?
2. True or False. Note that true means true for ALL cases.
   (a) In general, $P(AB|C) = P(B|C)P(A|BC)$
   (b) If A and B are independent, so are A and B^C.

1. Bayes' Theorem: $P(E|F) = \frac{P(F|E)P(E)}{P(F)}$
2. (a) True

$P(AB\|C)$	Left side
$\frac{P(ABC)}{P(C)}$	Def'n Cond'n Prob
$\frac{P(A\|BC)P(BC)}{P(C)}$	Chain Rule
$\frac{P(A\|BC)P(B\|C)P(C)}{P(C)}$	Chain Rule
$P(A\|BC)P(B\|C)$	Cancellation
$P(B\|C)P(A\|BC)$	■

(b) True

Start from Law of Total Probability (this is a good candidate for a starting point because it relates A, B and B^C). We will employ the assumption that $A \perp B$ (i.e. A is independent of B) somewhere, and then try to see if we can arrive at the equation $P(AB^C) = P(A)P(B^C)$ (this is what it means for A and B^C to be independent mathematically).

$$
\begin{aligned}
P(A) &= P(AB) + P(AB^C) && \text{LOTP} \\
P(A) &= P(A)P(B) + P(AB^C) && A \perp B \\
P(A) - P(A)P(B) &= P(AB^C) && \text{Subtract } P(A)P(B) \\
P(A)(1 - P(B)) &= P(AB^C) && \text{Factor } P(A) \\
P(A)P(B^C) &= P(AB^C) && 1 - P(B) := P(B^C) \\
P(AB^C) &= P(A)P(B^C) && \blacksquare
\end{aligned}
$$

1 Malware Detection

Jerry's niece, Lauren, started 8^{th} grade this past September! But because there are so many snowstorms each school year that require students be schooled remotely, her middle school has provided her with an old laptop so she can Zoom in for class. To better protect its own equipment from malware and viruses, the school installed two browser extensions to scrutinize all web downloads. Each download is either safe (event S) or unsafe (event S^C), and all downloads are examined by both extensions. The first extension marks the download as either safe (event A) or unsafe (event A^C), and the second extension marks the download as safe (event B) or unsafe (event B^C).

Assume that 94% of Lauren's web downloads are safe, and that:

- the first browser extension accurately marks unsafe downloads as unsafe with probability 0.93, but improperly marks safe downloads as unsafe with probability 0.04.
- the second browser extension accurately marks unsafe downloads as unsafe with probability 0.85, but improperly marks safe downloads as unsafe with probability 0.02.

Assume that given a download is safe, the two browser extensions independently mark that download as safe. Similarly, the two browser extensions independently mark unsafe downloads as unsafe.

a. Assuming a downloaded document is safe, what is the probability that at least one of the two extensions marks the download as unsafe?

Let's jot down everything we've been given:

$$P(S) = 0.94, \text{ so } P(S^C) = 1 - P(S) = 0.06$$

$$P(A^C|S^C) = 0.93, \text{ so } P(A|S^C) = 1 - P(A^C|S^C) = 0.07$$
$$P(A^C|S) = 0.04, \text{ so } P(A|S) = 1 - P(A^C|S) = 0.96$$

$$P(B^C|S^C) = 0.85, \text{ so } P(B|S^C) = 1 - P(B^C|S^C) = 0.15$$
$$P(B^C|S) = 0.02, \text{ so } P(B|S) = 1 - P(B^C|S) = 0.98$$

$$P(AB|S) = P(A|S)P(B|S)$$
$$P(A^C B^C|S^C) = P(A^C|S^C)P(B^C|S^C)$$

The probability that one or both extensions mark a safe download as unsafe is more easily computed as 1 minus the probability that **both** mark a safe download as safe.

$$\begin{aligned} P((AB)^C|S) &= 1 - P(AB|S) \\ &= 1 - P(A|S)P(B|S) \\ &= 1 - 0.96 \cdot 0.98 \\ &= 0.0592 \end{aligned}$$

The second line above follows from the first because A and B are conditionally independent given S.

b. Assuming a downloaded document is safe, what is the probability that exactly one of the two extensions marks the download as unsafe?

We're still in the world of safe downloads, so:

$$\begin{aligned} P(A^C B|S) + P(AB^C|S) &= P(A^C|S)P(B|S) + P(A|S)P(B^C|S) \\ &= 0.04 \cdot 0.98 + 0.96 \cdot 0.02 \\ &= 0.0584 \end{aligned}$$

Note that if A and B are conditionally independent given S, then so are A and B^C, as are A^C and B.

c. Given that both extensions mark the download as safe, what is the probability that the download is unsafe?

We are interested in $P(S^C|AB)$, which according to Bayes' Theorem is $\frac{P(AB|S^C)P(S^C)}{P(AB)}$. Some of these probabilities are given, but others have to be computed from scratch.

$$
\begin{aligned}
P(AB|S^C) &= P(A|S^C) \cdot P(B|S^C) \\
&= 0.07 \cdot 0.15 \\
&= 0.0105 \\
P(S^C) &= 0.06 \\
P(AB) &= P(AB|S)P(S) + P(AB|S^C)P(S^C) \\
&= P(A|S)P(B|S)P(S) + 0.0105 \cdot 0.06 \\
&= 0.96 \cdot 0.98 \cdot 0.94 + 0.0105 \cdot 0.06 \\
&= 0.884982
\end{aligned}
$$

Now we can compute $P(S^C|AB)$ as $\frac{0.0105 \cdot 0.06}{0.884982}$, or 0.000712. That's less than $\frac{1}{10}^{th}$ of a percent, which means the probability an unsafe download goes undetected is super small.

d. Are the unconditioned events where the two extensions mark a download as safe independent? Why or why not?

Nope. Intuitively, you would expect the second extension is more likely to flag a download as safe when the first extension does, and vice versa. Mathematically, we examine $P(AB)$, which we've already computed to be 0.88498, and compare that to $P(A)P(B)$. We haven't computed $P(A)$ or $P(B)$ yet, so let's do that now:

$$
\begin{aligned}
P(A) &= P(A|S)P(S) + P(A|S^C)P(S^C) \\
&= 0.96 \cdot 0.94 + 0.07 \cdot 0.06 \\
&= 0.9066 \\
P(B) &= P(B|S)P(S) + P(B|S^C)P(S^C) \\
&= 0.98 \cdot 0.94 + 0.15 \cdot 0.06 \\
&= 0.9302 \\
P(A)P(B) &= 0.9066 \cdot 0.9302 = 0.84332 \neq 0.884982 = P(AB)
\end{aligned}
$$

That's a solid mathematical defense that A and B are **not independent**.

2 Taking Expectation: Breaking Vegas

Preamble: When a random variable fits neatly into a family we've seen before (e.g. Binomial), we get its expectation for free. When it does not, we have to use the definition of expectation.

Problem: If you bet on "Red" in Roulette, there is $p = 18/38$ that you with win \$Y and a $(1-p)$ probability that you lose \$Y. Consider this algorithm for a series of bets:

Let Y = \$1. First you bet Y. If you win, then stop. If you lose, then set Y to be 2Y and repeat.

What are your expected winnings when you stop? It will help to recall that the sum of a geometric series $a^0 + a^1 + a^2 + \cdots = \frac{1}{1-a}$ if $0 < a < 1$. Vegas breaks you: Why doesn't everyone do this?

Let X be the number of dollars that your earn.

The possible values of x are from the outcomes of: winning on your first bet, winning on your second bet, and so on.

$$\begin{aligned} E[X] &= \frac{18}{38} + \frac{20}{38}\frac{18}{38}(2-1) + \left(\frac{20}{38}\right)^2 \frac{18}{38}(4-2-1) + \ldots \\ &= \sum_{i=0}^{\infty} \left(\frac{20}{38}\right)^i \left(\frac{18}{38}\right)\left(2^i - \sum_{j=0}^{i-1} 2^j\right) \\ &= \left(\frac{18}{38}\right) \sum_{i=0}^{\infty} \left(\frac{20}{38}\right)^i \\ &= \left(\frac{18}{38}\right) \frac{1}{1-\frac{20}{38}} = 1 \end{aligned}$$

Real games have maximum bet amounts. You have finite money and casinos can kick you out. But, if you had no betting limits and infinite money, then go for it! (and tell me which planet you are living on).

3 Conditional Probabilities: Missing Not at Random

Preamble: We have three big tools for manipulating conditional probabilities:

- Definition of conditional probability: $P(EF) = P(E|F)P(F)$
- Law of Total Probability: $P(E) = P(EF) + P(EF^C) = P(E|F)P(F) + P(E|F^C)P(F^C)$
- Bayes Rule: $P(E|F) = \frac{P(F|E)P(E)}{P(F)} = \frac{P(F|E)P(E)}{P(F|E)P(E)+P(F|E^C)P(E^C)}$

This is a good time to commit these three to memory and start thinking about when each of them is useful.

Problem: You collect data on whether or not people intend to vote for Ayesha, a candidate in an upcoming election. You send an electronic poll to 100 randomly chosen people. You assume all 100 responses are independent and identically distributed.

User Response	**Count**
Responded that they will vote for Ayesha	40
Responded that they will **not** vote for Ayesha	45
Did not respond	15

Let A be the event that a person responds saying they'll vote for Ayesha. Let M be the event that a user did not respond to the poll. We are interested in estimating $P(A)$, though computing that estimate is difficult, given that 15 users didn't actually respond.

a. What is the probability that a user said they will vote for Ayesha and that they responded to the poll $P(A \text{ and } M^C)$?

b. Which formula from class would you use to calculate $P(A)$? Your formula should rely on the context that voters for Ayesha are in one of two (mutually exclusive) groups: those that missed the poll, and those that did not.

c. Calculate the $P(A)$. You estimate that the probability that a voter is missing, given that they were going to vote for Ayesha is $P(M|A) = \frac{1}{5}$.

a. $P(A \text{ and } M^C) = \frac{40}{100}$. The M^C part is redundant.

b. The law of total probability. It breaks down $P(A)$ into two parts, the part which intersects with M and the part that intersections with M^C.

$$P(A) = P(A \text{ and } M) + P(A \text{ and } M^C)$$

c.

$$\begin{aligned}
P(A) &= P(A \text{ and } M^C) + P(A \text{ and } M) && \text{Law of total probability} \\
&= \frac{40}{100} + P(A \text{ and } M) && \text{From part a} \\
&= \frac{40}{100} + P(M|A)P(A) && \text{Chain rule} \\
P(A) - P(M|A)P(A) &= \frac{40}{100} && \text{The rest is algebra} \\
P(A) \cdot [1 - P(M|A)] &= \frac{40}{100} \\
P(A) \cdot \frac{4}{5} &= \frac{40}{100} \\
P(A) &= \frac{40}{100} \cdot \frac{5}{4} \\
P(A) &= \frac{1}{2}
\end{aligned}$$

4 Sending Bits to Space

Preamble: When sending binary data to satellites (or really over any noisy channel), the bits can be flipped with high probability. In 1947, Richard Hamming developed a system to more reliably send data. By using Error Correcting Hamming Codes, you can send a stream of 4 bits along with 3 redundant bits. If zero or one of the seven bits are corrupted, using error correcting codes, a receiver can identify the original 4 bits.

Problem: Lets consider the case of sending a signal to a satellite where each bit is independently flipped with probability $p = 0.1$.

a. If you send 4 bits, what is the probability that the correct message was received (i.e. none of the bits are flipped).

b. If you send 4 bits, with 3 Hamming error correcting bits, what is the probability that an interpretable message (i.e. a message with zero or one errors) was received?

c. Instead of using Hamming codes, you decide to send 100 copies of each of the four bits. If for every single bit, more than 50 of the copies are not flipped, the signal will be correctable. What is the probability that a correctable message was received?

Hamming codes are super interesting. It's worth looking up if you haven't seen them before!

a. Let Y be the number of 4 bits corrupted. Then $P(Y = k)$ is given as:

$$P(Y = 0) = \binom{4}{0}(0.1)^0(0.9)^4 = 0.656$$

b. Let Z be the number of 7 bits corrupted. A correctable message is received if Z equals 0 or 1:

$$\begin{aligned} P(\text{correctable}) &= P(Z = 0) + P(Z = 1) \\ &= \binom{7}{0}(0.1)^0(0.9)^7 + \binom{7}{1}(0.1)^1(0.9)^6 = 0.850 \end{aligned}$$

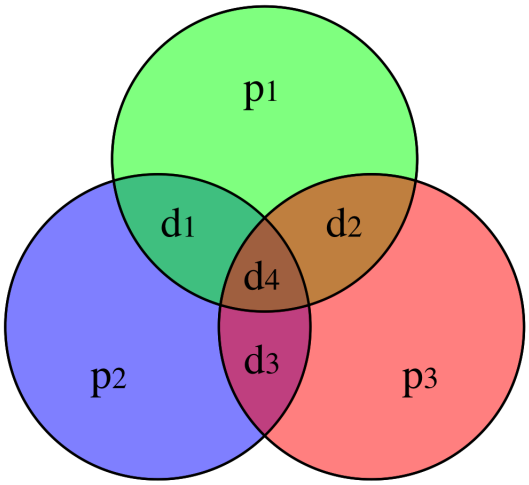
That is a 30% improvement!

c. Let X_i be the number of copies of bit i which are not corrupted. We can represent each as a random variable as we did in parts a and b.

$$\begin{aligned} P(\text{correctable}) &= \prod_{i=1}^{4} P(X_i > 50) \\ &= \prod_{i=1}^{4} \sum_{j=51}^{100} P(X_i = j) \\ &= \prod_{i=1}^{4} \sum_{j=51}^{100} \binom{100}{j}(0.9)^j(0.1)^{100-j} \\ &= \left(\sum_{j=51}^{100} \binom{100}{j}(0.9)^j(0.1)^{100-j} \right)^4 > 0.999 \end{aligned}$$

But now you need to send 400 bits, instead of the 7 required by hamming codes :-).

Extra: Explanation of the "`Hamming(7,4)`" technique



If we are trying to transmit 4 bits, we can send an additional 3 ”parity” bits that we can use to correct our original message if a bit gets flipped due to an error in transmission. Consider the diagram. The data bits are d_1 through d_4. The ”parity” bits are p_1 through p_3. A parity bit is set to whatever value would make it's large circle have an even number of bits. For example, the green circle consists of p_1, d_1, d_2, and d_4. If $d_1 = 1$, $d_2 = 1$, and $d_4 = 1$, then p_1 would be set to 1 in order to ensure there are an even number of bits in that circle (in this case, 4 bits).

Convince yourself that a single error which appeared in any bit could be identified and corrected! For example, if d_2 is flipped, it would throw off the parity for the green and red circles. Therefore, flipping d_2 back is the only way to correct the parity. As another example, if p_2 is flipped, then only the blue circle would have a parity issue, and flipping p_2 back is the unique solution to fixing the parity.