CMPT 210: Probability and Computation

Lecture 9

Sharan Vaswani

June 7, 2022

Logistics

- Assignment 1 is checked and the marks are up on Coursys. Average = 149/170 and Median = 157/170.
- Collect your marked Assignment 1 from TASC-1 9203 between 10.30 am - 12 pm.
- Assignment 2 is out: `https://vaswanis.github.io/210-S22/A2.pdf`
  Due Friday 17 June in class.
- For A2, you can use your late-submission and submit on Tuesday 21 June in class.
- To help you prepare for the midterm the solutions will be released on 21 June after class, meaning that no submissions will be allowed after that.
- If you have questions about either assignment or the marking, post it on Piazza: `https://piazza.com/sfu.ca/summer2022/cmpt210/home`

Conditional Probability - Recap

Recall that for events E and F such that $\Pr[F] \neq 0$, $\Pr[E|F] = \frac{\Pr[E \cap F]}{\Pr[F]}$.

For the complement E^c,$\Pr[E^c|F] = 1 - \Pr[E|F]$.

Proof: Since $E \cup E^c = \mathcal{S}$, for an event F such that $\Pr[F] \neq 0$,

$$(E \cup E^c) \cap F = (E \cap F) \cup (E^c \cap F) = \mathcal{S} \cap F = F$$

$$\implies \Pr[E \cap F] + \Pr[E \cap F^c] = \Pr[F] \implies \frac{\Pr[E^c \cap F]}{\Pr[F]} = 1 - \frac{\Pr[E \cap F]}{\Pr[F]}$$

$$\implies \Pr[E^c|F] = 1 - \Pr[E|F]$$

Generalization to multiple events

For events E_1, E_2, E_3, $\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1]\ \Pr[E_2|E_1]\ \Pr[E_3|E_1 \cap E_2]$.

Proof: By the rule of conditional probability,

$$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1]\ \Pr[E_2 \cap E_3|E_1] = \Pr[E_1]\ \Pr[E_2|E_1]\ \Pr[E_3|E_1 \cap E_2]$$

We can order the events that to compute $\Pr[E_1 \cap E_2 \cap E_3]$ more easily. For example,

$$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_2]\ \Pr[E_3|E_2]\ \Pr[E_1|E_2 \cap E_3]$$

Law of Total Probability and Bayes Rule - Recap

For events E and F such that $\Pr[E] \neq 0$ and $\Pr[F] \neq 0$,

$$\Pr[F|E] = \frac{\Pr[E|F]\Pr[F]}{\Pr[E]} \qquad \text{(Bayes Rule)}$$

For events E and F,

$$\Pr[E] = \Pr[E|F]\Pr[F] + \Pr[E|F^c]\Pr[F^c] \qquad \text{(Law of total probability)}$$

Combining the above equations,

$$\Pr[F|E] = \frac{\Pr[E|F]\ \Pr[F]}{\Pr[E|F]\Pr[F] + \Pr[E|F^c]\Pr[F^c]}$$

Generalization to multiple events

For disjoint events E_1, E_2, E_3 such that $E_1 \cup E_2 \cup E_3 = \mathcal{S}$ and $E_1 \cap E_2 \cap E_3 = \{\}$ i.e. events E_1, E_2 and E_3 form a partition, for any event A,

$$A = (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \qquad \text{(Since } E_1 \cup E_2 \cup E_3 = \mathcal{S}\text{)}$$

$$\implies \Pr[A] = \Pr[A \cap E_1] + \Pr[A \cap E_2] + \Pr[A \cap E_2] \qquad \text{(By union-rule for disjoint events)}$$

$$\implies \Pr[A] = \Pr[A|E_1]\Pr[E_1] + \Pr[A|E_2]\Pr[E_2] + \Pr[A|E_3]\Pr[E_3]$$

(By definition of conditional probability)

Similarly, we can obtain the Bayes rule for 3 events,

$$\Pr[E_1|A] = \frac{\Pr[A|E_1]\ \Pr[E_1]}{\Pr[A|E_1]\Pr[E_1] + \Pr[A|E_2]\Pr[E_2] + \Pr[A|E_3]\Pr[E_3]}$$

Questions?

Total Probability - Examples

Q: An insurance company believes that people can be divided into two classes — those that are accident prone and those that are not. Their statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas this probability decreases to 0.2 for a non-accident-prone person. If we assume that 30% of the population is accident prone, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?

Let A = event that a new policy holder will have an accident within a year of purchasing a policy. Let B = event that the new policy holder is accident prone. We know that $\Pr[B] = 0.3$, $\Pr[A|B] = 0.4$, $\Pr[A|B^c] = 0.2$. By the law of total probability,
$\Pr[A] = \Pr[A|B]\Pr[B] + \Pr[A|B^c]\Pr[B^c] = (0.4)(0.3) + (0.2)(0.7) = 0.26$.

Q: Suppose that a new policy holder has an accident within a year of purchasing their policy. What is the probability that they are accident prone?

Compute $\Pr[B|A] = \frac{\Pr[A|B]\ \Pr[B]}{\Pr[A]} = \frac{0.12}{0.26} = 0.4615$.

Total Probability Examples

Q: At a certain stage of a criminal investigation, the inspector in charge is 60% convinced of the guilt of a certain suspect. Suppose now that a new piece of evidence that shows that the criminal has a certain characteristic (such as left-handedness, baldness, brown hair, etc.) is uncovered. If 20% of the general population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect is among this group?

Let G be the event that the suspect is guilty. Let C be the event that the suspect possesses the characteristic found at the crime scene. We wish to compute $\Pr[G|C]$.

We know that $\Pr[G] = 0.6$, $\Pr[C|G] = 1$, $\Pr[C|G^c] = 0.2$.
$\Pr[C] = \Pr[C|G]\Pr[G] + \Pr[C|G^c]\Pr[G^c] = (1)(0.6) + (0.2)(0.4) = 0.68$

$\Pr[G|C] = \frac{\Pr[G]\Pr[C|G]}{\Pr[C]} = \frac{0.6}{0.68} = 0.882.$

Hence, the additional evidence has corroborated the inspector's theory and increased the probability of guilt.

Total Probability - Examples

Alice is taking a probability class and at the end of each week she can be either up-to-date or she may have fallen behind. If she is up-to-date in a given week, the probability that she will be up-to-date (or behind) in the next week is 0.8 (or 0.2, respectively). If she is behind in a given week, the probability that she will be up-to-date (or behind) in the next week is 0.6 (or 0.4, respectively). Alice is (by default) up-to-date when she starts the class. What is the probability that she is up-to-date after three weeks?

Let U_i and B_i be the events that Alice is up-to-date or behind respectively after i weeks. Since Alice starts the class up-to-date, $\Pr[U_1] = 0.8$ and $\Pr[B_1] = 0.2$. We also know that $\Pr[U_2|U_1] = 0.8$, $\Pr[U_3|U_2] = 0.8$ and $\Pr[B_2|U_1] = 0.2$, $\Pr[B_3|U_2] = 0.2$. Similarly, $\Pr[U_2|B_1] = 0.6$, $\Pr[U_3|B_2] = 0.6$ and $\Pr[B_2|B_1] = 0.4$, $\Pr[B_3|B_2] = 0.4$.

We wish to compute $\Pr[U_3]$. By the law of total probability,
$\Pr[U_3] = \Pr[U_3|U_2]\Pr[U_2] + \Pr[U_3|B_2]\Pr[B_2]$ and
$\Pr[U_2] = \Pr[U_2|U_1]\Pr[U_1] + \Pr[U_2|B_1]\Pr[B_1]$.

Hence, $\Pr[U_2] = (0.8)(0.8) + (0.6)(0.2) = 0.76$, and $\Pr[U_3] = (0.8)(0.76) + (0.6)(0.24) = 0.752$.

Simpson's Paradox

In 1973, there was a lawsuit against a university with the claim that a male candidate is more likely to be admitted to the university than a female.

Let us consider a simplified case – there are two departments, EE and CS, and men and women apply to the program of their choice. Let us define the following events: A is the event that the candidate is admitted to the program of their choice, F_E is the event that the candidate is a woman applying to EE, F_C is the event that the candidate is a woman applying to CS. Similarly, we can define M_E and M_C. Assumption: Candidates are either men or women, and that no candidate is allowed to be part of both EE and CS.

Lawsuit claim: Male candidate is more likely to be admitted to the university than a female i.e. $\Pr[A|M_E \cup M_C] > \Pr[A|F_E \cup F_C]$.

University response: In any given department, a male applicant is less likely to be admitted than a female i.e. $\Pr[A|F_E] > \Pr[A|M_E]$ and $\Pr[A|F_C] > \Pr[A|M_C]$.

Simpson's Paradox: Both the above statements can be simultaneously true.

Simpson's Paradox

CS	2 men admitted out of 5 candidates	40%
	50 women admitted out of 100 candidates	50%
EE	70 men admitted out of 100 candidates	70%
	4 women admitted out of 5 candidates	80%
Overall	72 men admitted, 105 candidates	$\approx 69\%$
	54 women admitted, 105 candidates	$\approx 51\%$

In the above example, $\Pr[A|F_E] = 0.8 > 0.7 = \Pr[A|M_E]$ and $\Pr[A|F_C] = 0.5 > 0.4 = \Pr[A|M_C]$. $\Pr[A|F_E \cup F_C] \approx 0.51$. Similarly, $\Pr[A|M_E \cup M_C] \approx 0.69$.

In general, Simpson's Paradox occurs when multiple small groups of data all exhibit a similar trend, but that trend reverses when those groups are aggregated.

Questions?

Back to throwing dice - Independent Events

Suppose we throw two standard dice one after the other. What is the probability that we get two 6's in a row?

E = We get a 6 in the second throw. F = We get a 6 in the first throw. $E \cap F$ = we get two 6's in a row. We are computing $\Pr[E \cap F]$. $\Pr[E] = \Pr[F] = \frac{1}{6}$.

$\Pr[E|F] = \frac{\Pr[E \cap F]}{\Pr[F]} \implies \Pr[E \cap F] = \Pr[E|F] \Pr[F]$.

Since the two dice are *independent*, knowing that we got a 6 in the first throw does not change the probability that we will get a 6 in the second throw. Hence, $\Pr[E|F] = \Pr[E]$ (conditioning does not change the probability of the event).

Hence, $\Pr[E \cap F] = \Pr[E|F] \Pr[F] = \Pr[E] \Pr[F] = \frac{1}{6} \frac{1}{6} = \frac{1}{36}$.

Independent Events

Events E and F are said to be independent, if knowledge that F has occurred does not change the probability that E occurs. Formally,

$$\Pr[E \cap F] = \Pr[E] \Pr[F]$$

Q: I toss two independent, fair coins. What is the probability that I get the HT sequence?

Define E to be the event that I get a heads in the first toss, and F be the event that I get a tails in the second toss. Since the two coins are independent, events E and F are also independent. $\Pr[E \cap F] = \Pr[E] \Pr[F] = \frac{1}{2} \frac{1}{2} = \frac{1}{4}$.

Q: I randomly choose a number from $\{1, 2, \ldots, 10\}$. E is the event that the number I picked is a prime number. F is the event that the event I picked is an odd number. Are E and F independent?

$\Pr[E] = \frac{2}{5}$, $\Pr[F] = \frac{1}{2}$, $\Pr[E \cap F] = \frac{3}{10}$. $\Pr[E \cap F] \neq \Pr[E] \Pr[F]$. Another way: $\Pr[E|F] = \frac{3}{5}$ and $\Pr[E] = \frac{2}{5}$, and hence $\Pr[E|F] \neq \Pr[E]$. Conditioning on F tell us that prime number cannot be 2, so it changes the probability of E.

Independent Events - Example

Q: We have a machine that has 2 independent components. The machine breaks if each of its 2 components break. Suppose each component can break with probability p, what is the probability that the machine does not break?

Let E_1 = Event that the first component breaks, E_2 = Event that the second component breaks. M = Event that the machine breaks = $E_1 \cap E_2$.

$\Pr[M] = \Pr[E_1 \cap E_2]$. Since the two components are independent, E_1 and E_2 are independent, meaning that $Pr[E_1 \cap E_2] = Pr[E_1]\ \Pr[E_2] = p^2$.

Probability that the machine does not break = $\Pr[M^c] = 1 - \Pr[M] = 1 - p^2$.

Q: We have a new machine that breaks if either of its 2 components break. Suppose each component can break with probability p, what is the probability that the machine breaks?

For this machine, let M' be the event that it breaks. In this case, $\Pr[M] = \Pr[E_1 \cup E_2]$. Since E_1 and E_2 are mutually exclusive, by the union rule, $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] = 2p$.

Independent Events

Mutual Independence: A set of events is said to be mutually independent if the probability of each event in the set is the same no matter which of the events has occurred.

For any selection of two or more of the events, the probability that all the selected events occur equals the product of the probabilities of the selected events.

Example: For events E_1, E_2 and E_3 to be mutually independent, all the following equalities should hold:

$$\Pr[E_1 \cap E_2] = \Pr[E_1]\Pr[E_2] \quad \Pr[E_1 \cap E_3] = \Pr[E_1]\Pr[E_3]$$
$$\Pr[E_2 \cap E_3] = \Pr[E_2]\Pr[E_3] \quad \Pr[E_1 \cap E_2 \cap E_3] = \Pr[E_1]\Pr[E_2]\Pr[E_3].$$

Can generalize this concept to n events – E_1, E_2, ..., E_n are mutually independent, if for *every* subset of events, the probability that all the selected events occur equals the product of the probabilities of the selected events. Formally, for every subset $S \subseteq \{1, 2, \ldots, n\}$, $\Pr[\cap_{i \in S} E_i] = \prod_{i \in S} \Pr[E_i]$.

Mutual independence vs Pairwise independence

Q: Suppose that we flip three fair, mutually-independent coins. Define the following events: E_1 is the event that coin 1 matches coin 2, E_2 is the event that coin 2 matches coin 3 and E_3 is the event that coin 3 matches coin 1. Are E_1, E_2 and E_3 mutually independent?

$\mathcal{S} = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$.

$\Pr[E_1] = \Pr[\{HHH, HHT, TTH, TTT\}] \implies \Pr[E_1] = \frac{4}{8} = \frac{1}{2}$. Similarly, $\Pr[E_2] = \Pr[E_3] = \frac{1}{2}$.

$\Pr[E_1 \cap E_2] = \Pr[\{HHH, TTT\}] = \frac{2}{8} = \frac{1}{4}$. Hence, $\Pr[E_1 \cap E_2] = \Pr[E_1] \cdot \Pr[E_2]$. Similarly, $\Pr[E_2 \cap E_3] = \Pr[E_2] \cdot \Pr[E_3]$ and $\Pr[E_1 \cap E_3] = \Pr[E_1] \cdot \Pr[E_3]$.

$\Pr[E_1 \cap E_2 \cap E_3] = \Pr[\{HHH, TTT\}] = \frac{2}{8} = \frac{1}{4} \neq \Pr[E_1]\ \Pr[E_2]\ \Pr[E_3] = \frac{1}{8}$.

Hence, three events (E_1, E_2 and E_3) can be pairwise independent, but not necessarily mutually independent!

Questions?

Matrix Multiplication

Given two $n \times n$ matrices – A and B, if $C = AB$, then,

$$C_{i,j} = \sum_{k=1}^{n} A_{i,k} B_{k,j}$$

Hence, in the worst case, computing $C_{i,j}$ is an $O(n)$ operation. There are n^2 entries to fill in C and in the absence of additional structure, matrix multiplication takes $O(n^3)$ time.

There are non-trivial algorithms for doing matrix multiplication more efficiently:

- (Strassen, 1969) Requires $O(n^{2.81})$ operations.
- (Coppersmith-Winograd, 1987) Requires $O(n^{2.376})$ operations.
- (Alman-Williams, 2020) Requires $O(n^{2.373})$ operations.
- Belief is that it can be done in time $O(n^{2+\epsilon})$ for $\epsilon > 0$.

Verifying Matrix Multiplication

For simplicity, we will focus on matrix multiplication mod 2, i.e. $C_{i,j} = (\sum_{k=1}^{n} A_{i,k} B_{k,j}) \bmod 2$, implying that all entries in C are binary (either 0 or 1).

Example: $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$ then $C = AB = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$

Objective: Verify whether a matrix multiplication operation is correct.

Trivial way: Do the matrix multiplication ourselves, and verify it using $O(n^3)$ (or $O(n^{2.373})$) operations.

Frievald's Algorithm: Randomized algorithm to verify matrix multiplication with high probability in $O(n^2)$ time.

(Basic) Frievald's Algorithm

In order to answer the question: Is $D = AB \pmod 2$?

1. Generate a random n-bit vector x, by making each bit x_i either 0 or 1 *independently* with probability $\frac{1}{2}$. E.g, for $n = 2$, toss a fair coin independently twice with the scheme – H is 0 and T is 1). If we get HT, then set $x = [0\,;\,1]$.
2. Compute $t = Bx$ and $y = At = A(Bx)$ and $z = Dx$.
3. Output "yes" if $y = z$, else output "no".

Computational complexity: Step 1 can be done in $O(n)$ time. Step 2 requires 3 matrix vector multiplications and can be done in $O(n^2)$ time. Step 3 requires comparing two n-dimensional vectors and can be done in $O(n)$ time. Hence, the total computational complexity is $O(n^2)$.

(Basic) Frievald's Algorithm

Let us run the algorithm on an example. Suppose we have generated $x = [1\,;\,0]$

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad ; \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad ; \quad D = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

$$Bx = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad ; \quad y = A(Bx) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad ; \quad z = Dx = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Hence the algorithm will correctly output "no" since $D \neq AB \pmod 2$.

Q: Suppose we have generated $x = [1\,;\,1]$. What is y and z?

In this case, $y = z$ and the algorithm will incorrectly output "yes" even though $D \neq AB \pmod 2$.

(Basic) Frievald's Algorithm

Let us run the algorithm on an example. Suppose we have generated $x = [1\,;\,0]$.

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad ; \quad B = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \quad ; \quad C = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

$$Bx = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad ; \quad y = A(Bx) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad ; \quad z = Cx = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Hence the algorithm will correctly output "yes" since $C = AB \pmod 2$.

Q: Suppose we have generated $x = [1\,;\,1]$. What is y and z?

In this case again, $y = z$ and the algorithm will correctly output "yes".

(Basic) Frievald's Algorithm - Analysis

Case (i): If $D = AB \pmod 2$, does the algorithm always output "yes"? Yes! Since $D = AB \pmod 2$, for any vector x, $Dx = ABx \pmod 2$.

Case (ii) If $D \neq AB \pmod 2$, does the algorithm output "no"?

Claim: For any input matrices A, B, D if $D \neq AB \pmod 2$, then the Frievald's algorithm will output "no" with probability $\geq \frac{1}{2}$.

	Yes	No
$D = AB$	1	0
$D \neq AB$	$\leq \frac{1}{2}$	$\geq \frac{1}{2}$

Table 1: Probabilities for Basic Frievalds Algorithm

(Basic) Frievald's Algorithm - Analysis

If $D \neq AB \pmod 2$, then $(AB - D) \neq \mathbf{0} \pmod 2$. If $E := AB - D$, if $D \neq AB \pmod 2$, then there exists at least one pair (i, j) such that $E_{i,j} = 1 \pmod 2$. To make things more concrete, let us assume that this pair is $(i, j) = (5, 3)$ and that $E_{5,3} = 1$. Consider coordinate 5 of the $Ex = y - z$ vector denoted as $(Ex)_5$. If $(Ex)_5$ is 1, then the algorithm will return "no" since $y - z = 1 \implies y \neq z$.

$$
\begin{aligned}
(Ex)_5 &= [E_{5,1}\, x_1 + E_{5,2}\, x_2 + E_{5,3}\, x_3 + \ldots E_{5,n}\, x_n] \pmod 2 \\
&= \left[\underbrace{[E_{5,1}\, x_1 + E_{5,2}\, x_2 + E_{5,4}\, x_4 + \ldots E_{5,n}\, x_n] \pmod 2}_{\omega} + x_3 \right] \pmod 2
\end{aligned}
$$

- x_3 is independent of ω and is hence equal to 1 (or 0) with probability equal to $\frac{1}{2}$.
- If $\omega = 0$, then, $(Ex)_5 = x_3 \pmod 2 = x_3$. If $\omega = 1$, then, $(Ex)_5 = (1 + x_3) \pmod 2$.

$\implies (Ex)_5 = 1$ with probability equal to $\frac{1}{2}$ *independent of the value of* $\omega \implies Ex$ has a 1 in at least one coordinate with probability $\geq 1/2 \implies$ with probability $\geq \frac{1}{2}$, $y \neq z$ and the algorithm will output "no" $\implies$ we proved our claim.

Questions?