

CMPT 210: Probability and Computing

Lecture 10

Sharan Vaswani

October 8, 2024

- **Conditional probability:** $\Pr[E|F] = \frac{\Pr[E \cap F]}{\Pr[F]}$.
- **Multiplication Rule:** For events E_1, E_2, \dots, E_n ,
 $\Pr[E_1 \cap E_2 \dots \cap E_n] = \Pr[E_1] \Pr[E_2|E_1] \Pr[E_3|E_1 \cap E_2] \dots \Pr[E_n|E_1 \cap E_2 \cap \dots E_{n-1}]$.
- **Bayes Rule:** For events E and F if $\Pr[E] \neq 0$ and $\Pr[F] \neq 0$, then, $\Pr[F|E] = \frac{\Pr[E|F] \Pr[F]}{\Pr[E]}$.

Conditional Probability for Complement Events

Q: Prove that for events E, F , $\Pr[E^c|F] = 1 - \Pr[E|F]$.

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

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

$$(E \cup E^c) \cap F = (E \cap F) \cup (E^c \cap F) \quad (\text{Distributive Law})$$

$$\implies \Pr[(E \cap F) \cup (E^c \cap F)] = \Pr[F]$$

Since $E \cap F$ and $E^c \cap F$ are mutually exclusive events,

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

$$\implies \Pr[E^c|F] = 1 - \Pr[E|F] \quad (\text{By def. of conditional probability})$$

Law of Total Probability and Bayes rule

Law of Total Probability: For events E and F , $\Pr[E] = \Pr[E|F] \Pr[F] + \Pr[E|F^c] \Pr[F^c]$.

Proof:

$$E = (E \cap F) \cup (E \cap F^c)$$

$$\implies \Pr[E] = \Pr[(E \cap F) \cup (E \cap F^c)] = \Pr[E \cap F] + \Pr[E \cap F^c]$$

(By union-rule for disjoint events)

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

(By definition of conditional probability)

Combining Bayes rule and Law of total probability

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

(By definition of conditional probability)

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

(By law of total probability)

Total Probability - Examples

Q: In answering a question on a multiple-choice test, a student either knows the answer or she guesses. Let p be the probability that she knows the answer and $1 - p$ the probability that she guesses. Assume that a student who guesses at the answer will be correct with probability $\frac{1}{m}$, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that she answered it correctly?

Let C be the event that the student answers the question correctly. Let K be the event that the student knows the answer. We wish to compute $\Pr[K|C]$.

We know that $\Pr[K] = p$ and $\Pr[C|K^c] = 1/m$, $\Pr[C|K] = 1$. Hence,
 $\Pr[C] = \Pr[C|K] \Pr[K] + \Pr[C|K^c] \Pr[K^c] = (1)(p) + \frac{1}{m} (1 - p)$.

$$\Pr[K|C] = \frac{\Pr[C|K] \Pr[K]}{\Pr[C]} = \frac{mp}{1+(m-1)p}.$$

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: 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,

$$\Pr[U_3] = \Pr[U_3|U_2] \Pr[U_2] + \Pr[U_3|B_2] \Pr[B_2] \text{ and}$$

$$\Pr[U_2] = \Pr[U_2|U_1] \Pr[U_1] + \Pr[U_2|B_1] \Pr[B_1].$$

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

Q: 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).

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

Independent Events

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,

$$\Pr[E|F] = \Pr[E] \quad ; \quad \Pr[E \cap F] = \Pr[E] \Pr[F]$$

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 $\{1, 2, \dots, 10\}$. E is the event that the number I picked is a prime. F is the event that the number I picked is odd. 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 = $\Pr[M^c] = 1 - \Pr[M] = 1 - p^2$.

Independent Events - Examples

Q: We have a new machine that has 2 independent components. The machine 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, $\Pr[M'] = \Pr[E_1 \cup E_2]$.

Incorrect: By the union rule for mutually exclusive events, $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] = 2p$.

Mistake: *Independence does not imply mutual exclusivity* and we can not use the union rule. Independence implies that for any two events E and F , $\Pr[E \cap F] = \Pr[E] \Pr[F]$, while mutual exclusivity requires that $\Pr[E \cap F] = 0$.

Correct way:

$$\begin{aligned}\Pr[E_1 \cup E_2] &= \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2] && \text{(By the inclusion-exclusion rule)} \\ &= \Pr[E_1] + \Pr[E_2] - \Pr[E_1] \Pr[E_2] = 2p - p^2 && \text{(Since } E_1 \text{ and } E_2 \text{ are independent.)}\end{aligned}$$

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 hence, 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

As an example, let us focus on A, B being binary 2×2 matrices.

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) Freivald's Algorithm

Q: For $n \times n$ matrices A , B and D , is $D = AB$?

Algorithm:

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 $y = z$ (all entries need to be equal), 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 $O(n^2)$.