CMPT 210: Probability and Computation

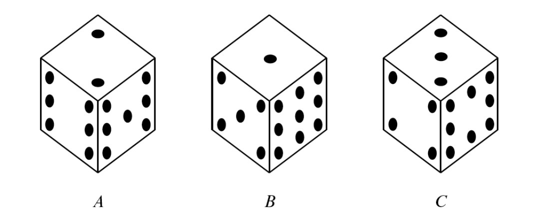
Lecture 8

Sharan Vaswani

June 3, 2022

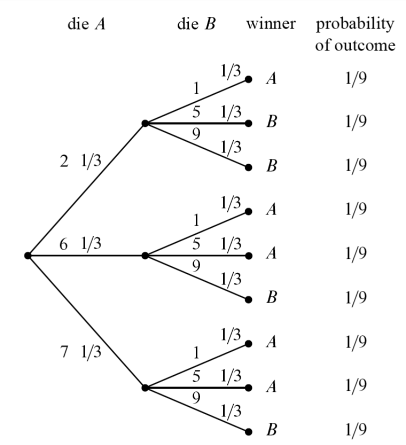
Conditional Probability - Examples

Let us play a game with three strange dice shown in the figure. Each player selects one die and rolls it once. The player with the lower value pays the other player \$100. We can pick a die first, after which the other player can pick one of the other two.



Q: Suppose we choose die B because it has a 9, and the other player selects die A. What the probability is that we will win?

Conditional Probability - Examples



Identify Outcomes: Each leaf is an outcome and $\mathcal{S} = \{(2,1),(2,5),(2,9),(6,1),(6,5),(6,9),(7,1),(7,5),(7,9)\}$.

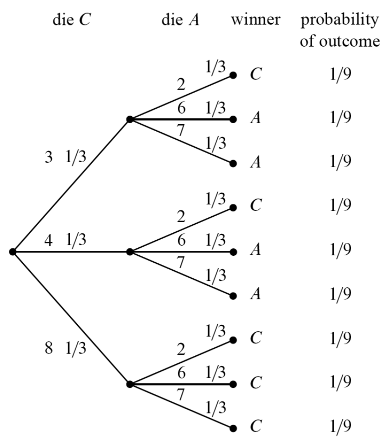
Identify Event: $E = \{(2,5),(2,9),(6,9),(7,9)\}$.

Compute probabilities: $\Pr[\text{Dice 1 is 6}] = \frac{1}{3}$.
$\Pr[(6,5)] = \Pr[\text{Dice 2 is 5} \; \cap \; \text{Dice 1 is 6}] =$
$\Pr[\text{Dice 2 is 5} \mid \text{Dice 1 is 6}] \Pr[\text{Dice 1 is 6}] = \frac{1}{3}\frac{1}{3} = \frac{1}{9}$.

$\Pr[E] = \Pr[(2,5)] + \Pr[(2,9)] + \Pr[(6,9)] + \Pr[(7,9)] = \frac{4}{9}$.
Meaning that there is less than 50% chance of winning.

Conditional Probability - Examples

Q: We get another chance – this time we know that die A is good (since we lost to it previously), we choose die A and the other player chooses die C. What is our probability of winning?



Now, $E = \{(3,6),(3,7),(4,6),(4,7)\}$ and hence $\Pr[E] = \frac{4}{9}$. Meaning that there is less than 50% chance of winning.

Conditional Probability - Examples

We get yet another chance, and this time we choose die C, because we reason that die A is better than B, and C is better than A.

By similar reasoning, we can construct a tree diagram to show that unfortunately, the probability that we win is again $\frac{4}{9}$.

So we conclude that,

- A beats B with probability $\frac{5}{9}$ (first game).
- C beats A with probability $\frac{5}{9}$ (second game).
- B beats C with probability $\frac{5}{9}$ (third game).

Since A will beat B more often than not, and B will beat C more often than not, it seems like A ought to beat C more often than not, that is, the "beats more often" relation ought to be transitive. But this intuitive idea is false: whatever die we pick, we the second player can pick one of the others and be likely to win. So picking first is actually a disadvantage!

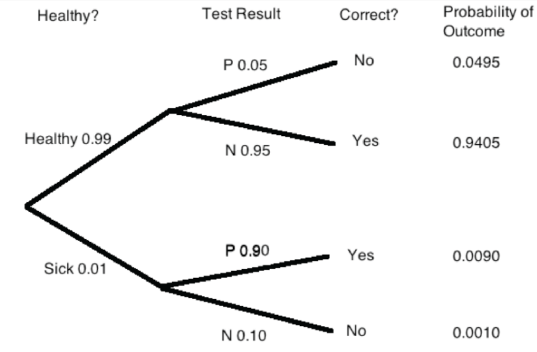
Conditional Probability - Examples

A test for detecting cancer has the following accuracy – (i) If a person has cancer, there is a 10% chance that the test will say that the person does not have it. This is called a “false negative” and (ii) If a person does not have cancer, there is a 5% chance that the test will say that the person does have it. This is called a “false positive”. For patients that have no family history of cancer, the incidence of cancer is 1%. Person X does not have any family history of cancer, but is detected to have cancer. What is the probability that the Person X does have cancer?

Conditional Probability - Examples

$\mathcal{S} = \{(Healthy, Positive), (Healthy, Negative), (Sick, Positive), (Sick, Negative)\}$.

A is the event that Person X has cancer. B is the event that the test is positive.



$$\Pr[A|B] = \frac{\Pr[A \cap B]}{\Pr[B]} = \frac{\{(S,P)\}}{\{(S,P),(H,P)\}} = \frac{0.0090}{0.0090+0.0495} \approx 15.4\%.$$

Questions?

Bayes Rule

For events E and F if $\Pr[E] \neq 0$ and $\Pr[F] \neq 0$, then,

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

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

$$\implies \Pr[E|F]\Pr[F] = \Pr[F|E]\Pr[E]$$

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

Allows us to compute $\Pr[F|E]$ using $\Pr[E|F]$. Later in the course, we will see an application of the Bayes rule to machine learning.

Law of Total Probability and Bayes rule

For events E and F,

$$\begin{aligned} & E = (E \cap F) \cup (E \cap F^c) \\ \implies & \Pr[E] = \Pr[(E \cap F) \cup (E \cap F^c)] = \Pr[E \cap F] + \Pr[E \cap F^c] \\ & \hfill \text{(By union-rule for disjoint events)} \\ & \Pr[E] = \Pr[E|F]\Pr[F] + \Pr[E|F^c]\Pr[F^c] \qquad \text{(By definition of conditional probability)} \end{aligned}$$

$$\Pr[F|E] = \frac{\Pr[F \cap E]}{\Pr[E]} = \frac{\Pr[E|F]\,\Pr[F]}{\Pr[E]} \qquad \text{(By definition of conditional probability)}$$

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

Law of Total Probability and Bayes rule

For disjoint events E_1, E_2, E_3 such that $\Pr[E_1 \cup E_2 \cup E_3] = 1$ and $\Pr[E_1 \cap E_2 \cap E_3] = 0$ i.e. events E_1, E_2 and E_3 form a partition, for any event A,

$$A = (A \cap E_1) \cup (A \cap E_2) \cup (A \cap E_3) \qquad \text{(Since } \Pr[E_1 \cup E_2 \cup E_3] = 1\text{)}$$

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

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

(By definition of conditional probability)

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

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

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 his policy. What is the probability that he is 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: 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: 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

Q: We flip a fair coin. If heads comes up, then we roll one die and take the result. If tails comes up, then we roll two dice and take the sum of the two results. What is the probability that this process yields a 2?

Q: What is the probability that this process yields a (i) 4, (ii) 6, (iii) 8?

Q: What is the probability that the first dice (in the two dice when we get a tails) is 4 given that the process yields a 6?

.

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?