Problem 1.

Out of 350 applicants for a job, 160 are female and 54 are female and have a graduate degree.

(a) What is the probability that a randomly chosen applicant has a graduate degree, given that they are female?

(b) If 104 of the applicants have graduate degrees, what is the probability that a randomly chosen applicant is female, given that the applicant has a graduate degree?

Solution

Recall the Conditional Probability formula, for two events A and B, where $\mathbb{P}(B) > 0$, the probability of A given B is:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Let F be the set of females, and G be the set of applicants that have a graduate degree.

(a) Using the formula, we can say that $\mathbb{P}(F) = \frac{160}{350}$, and $\mathbb{P}(F \cap G) = \frac{54}{350}$. Use the Conditional Probability formula to solve $\mathbb{P}(G \mid F)$:

$$\mathbb{P}(G \mid F) = \frac{\mathbb{P}(G \cap F)}{\mathbb{P}(F)} = \frac{\frac{54}{350}}{\frac{160}{350}} = \frac{54}{160} \approx 0.3375 = 33.75\%$$

So the probability that a randomly chosen applicant has a graduate degree given that they are female is about 33.75%.

(b) Since we have the number of total graduate degree applicants, we can calculate $\mathbb{P}(G)$ which is $\frac{104}{350}$. Use the Conditional Probability formula to solve $\mathbb{P}(F \mid G)$:

$$\mathbb{P}(F \mid G) = \frac{\mathbb{P}(F \cap G)}{\mathbb{P}(G)} = \frac{\frac{54}{350}}{\frac{104}{350}} = \frac{54}{104} \approx 0.5192 = 51.92\%$$

So the probability that a randomly chosen applicant is female given that they have a graduate degree is about 51.92%.

Problem 2.

Find the probability of drawing a spade, then a heart, then a diamond, and then a heart from a shuffled deck of cards (without replacement)

Solution

In order to find the probability, we can realize that since there are 13 cards of each suit originally in the deck (52 total), and the size of the deck decreases after each draw, you can write the conditional terms and solve as:

$$\mathbb{P}(\text{S then H then D then H}) = \frac{13}{52} \cdot \frac{13}{51} \cdot \frac{13}{50} \cdot \frac{12}{49} = \frac{26364}{6497400} = \frac{169}{41650} \approx 0.0041 = 0.41\%$$

So the probability that you draw a spade, a heart, a diamond, a heart, is about 0.41%.

Problem 3.

Find the probability of drawing a Jack or a red card and then a Jack from a shuffled deck of cards (without replacement).

Solution

We are dealing with two draws. The first one is inclusive, the first card being any red card (26) **or** any Jack (4), calculating the probability would be:

$$\mathbb{P}(J \cup R) = \mathbb{P}(J) + \mathbb{P}(R) - \mathbb{P}(J \cap R) = 4 + 26 - 2 = 28.$$

We can split the first two draws into cases:

- First is a Jack $\frac{4}{52}$, then a Jack again $\frac{3}{51}$.

- First is red non-Jack $\frac{24}{52}$, then a Jack $\frac{4}{51}$.

Add them to get our answer:

$$\frac{4}{52} \cdot \frac{3}{51} + \frac{24}{52} \cdot \frac{4}{51} = \frac{3}{663} + \frac{24}{663} = \frac{27}{663} = \frac{9}{221} \approx 0.0407 = 4.07\%.$$

So the probability of drawing a Jack or a red card and then a Jack from a shuffled deck of cards is about 4.07%.

Problem 4.

Three dice are rolled. If each lands on a different number, find the probability that one is 3.

Solution

This is another problem that we can use the Conditional Probability formula:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Let A be the set that at least one die is 3, and B be the set that all three dice show different numbers. The probability B would be:

$$\mathbb{P}(B) = \frac{6}{6} \cdot \frac{5}{6} \cdot \frac{4}{6} = \frac{5}{9},$$

since the second die must differ from the first, and the third must differ from both. For $A \cap B$, we can choose which die rolls the 3 (3 ways), then the other two must be distinct and not 3:

$$\mathbb{P}(A \cap B) = 3 \cdot \frac{1}{6} \cdot \frac{5}{6} \cdot \frac{4}{6} = \frac{5}{18}.$$

Now we can solve $\mathbb{P}(A \mid B)$:

$$\mathbb{P}(A \mid B) = \frac{\frac{5}{18}}{\frac{5}{9}} = \frac{5}{18} \cdot \frac{9}{5} = \frac{9}{18} = \frac{1}{2}.$$

So the probability that one dice is 3 given three die land on a different number is $\frac{1}{2}$.

Problem 5.

Suppose there is a lottery where 3 balls are randomly selected from a group of balls numbered from 1 to 50 (without replacement).

(a) What is the probability that we win the lottery if we have one ticket with 3 numbers?

(b) Suppose the first of 3 numbers announced is on our ticket. Find the probability that we will win given this information.

Solution

(a) Out of the 50 balls, there will be 3 balls that will match our 3 numbers on our ticket. Therefore, the probability that we win will be:

$$\mathbb{P}(\text{win}) = \frac{1}{\binom{50}{3}} = \frac{1}{\frac{50!}{47!3!}} = \frac{1}{19600} \approx 5.10 \times 10^{-5} = 0.00510\%$$

0.00510% are not good odds.

(b) Let A be the set that all 3 balls drawn match our ticket (win), and B be the set of the first announced number being on our ticket. We can use the Conditional Probability formula to solve this question:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Solving the probability of B is easy:

$$\mathbb{P}(B) = \frac{3}{50},$$

since the first draw has a $\frac{3}{50}$ chance of matching one of our 3 numbers. Solving the probability of $A \cap B$:

$$\mathbb{P}(A \cap B) = \frac{3}{50} \cdot \frac{2}{49} \cdot \frac{1}{48},$$

since if the first drawing is one of ours, then both remaining draws must be the other two numbers. Therefore $\mathbb{P}(A \mid B)$ would be:

$$\mathbb{P}(A \mid B) = \frac{\frac{3}{50} \cdot \frac{2}{49} \cdot \frac{1}{48}}{\frac{3}{50}} = \frac{2}{49} \cdot \frac{1}{48} = \frac{1}{1176} \approx 8.50 \times 10^{-4}$$

These are also not good odds.

Problem 6.

If the letters of the word "uncopyrightable" are put into a random order, what is the probability that letters will be in their correct locations given that the rearrangement begins with "unc"?

Solution

Once again, we can solve this using the Conditional Probability formula:

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Let A be the set of the entire word being in correct order (where the permutation equals the original), and B be the set of the permutation which begins with "unc". Notice that since the original word **does** start with "unc,"

$$A \subseteq B \implies \mathbb{P}(A \cap B) = P(A).$$

Thus,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A)}{\mathbb{P}(B)}.$$

We can now compute the pieces:

- $\mathbb{P}(A) = \frac{1}{15!}$ (1 out of 15! permutations match exactly)

- $\mathbb{P}(B) = \frac{1}{15} \cdot \frac{1}{14} \cdot \frac{1}{13}$ (first letter must be u, then n, then c)

Solve $\mathbb{P}(A \mid B)$:

$$\mathbb{P}(A \mid B) = \frac{\frac{1}{15!}}{\frac{1}{15 \cdot 14 \cdot 13}} = \frac{15 \cdot 14 \cdot 13}{15!} = \frac{1}{12!}$$

Since $12! = 479001600$,

$$\mathbb{P}(A \mid B) \approx 2.09 \times 10^{-9}.$$

Problem 7.

Three machines (A, B, C) manufacture screws. They manufacture 25%, 35%, and 40% of the screws, respectively. The output screws are defective at 2%, 3%, and 5%, respectively. If you choose a random screw produced at the factory and it is defective, what is the probability it came from each machine A, B, and C?

Solution

Let A be the set of screws from machine A, B be the set of screws from machine B, C be the set of screws from machine C, and D be the set of defective screws.

The first thing we need to calculate is the probability that a screw is defective. This can be done by computing the three products of the machine's manufacture percentage and their respective defective rate:

- A: $0.25 \times 0.02 = 0.005$

- B: $0.35 \times 0.03 = 0.0105$

- C: $0.40 \times 0.05 = 0.02$

and computing their sum (Law of Total Probability):

$$\mathbb{P}(D) = 0.005 + 0.0105 + 0.02 = 0.0355$$

We can now for the first time utilize Bayes' theorem to find the probability that a defective screw came from machine A, B, or C:

$$\mathbb{P}(A \mid D) = \frac{\mathbb{P}(D \mid A)\mathbb{P}(A)}{\mathbb{P}(D)} = \frac{0.005}{0.0355} = \frac{10}{71} \approx 0.1408$$

$$\mathbb{P}(B \mid D) = \frac{\mathbb{P}(D \mid B)\mathbb{P}(B)}{\mathbb{P}(D)} = \frac{0.0105}{0.0355} = \frac{21}{71} \approx 0.2958$$

$$\mathbb{P}(C \mid D) = \frac{\mathbb{P}(D \mid C)\mathbb{P}(C)}{\mathbb{P}(D)} = \frac{0.020}{0.0355} = \frac{40}{71} \approx 0.5634$$

As a quick check, we can see that the fractions and decimals both add to 1:

$$\frac{10}{71} + \frac{21}{71} + \frac{40}{71} = \frac{71}{71} = 1$$

$$0.1408 + 0.2958 + 0.5634 = 1$$

Problem 8.

Suppose we have an e-mail spam filter. If a message is spam, it has a 95% chance of blocking it. If a message is not spam, it still has a 2% chance to block it. Assume 20% of e-mails received are actually spam. If the filter blocks a message, what is the probability that it was actually spam?

Solution

Let S be the set of spam messages, and B be the set of blocked messages. From the question, we are given:

- $\mathbb{P}(S) = 0.20$, $\mathbb{P}(S^C) = 0.80$

- $\mathbb{P}(B \mid S) = 0.95$, $\mathbb{P}(B \mid S^C) = 0.02$

First we can get $\mathbb{P}(B)$ by using the Law of Total Probability:

$$\mathbb{P}(B) = 0.95 \cdot 0.20 + 0.02 \cdot 0.80 = 0.19 + 0.016 = 0.206$$

Then use Bayes' Theorem:

$$\mathbb{P}(S \mid B) = \frac{\mathbb{P}(B \mid S)\mathbb{P}(S)}{\mathbb{P}(B)} = \frac{0.95 \cdot 0.20}{0.206} = \frac{95}{103} \approx 0.9223 = 92.2\%$$

So, given a message is blocked, there is a 92.2% chance it was actually spam.

Problem 9.

There are four blood types (A, B, AB, and O) and two Rh factors (+ or -). Each person has a blood type and an Rh factor. The percentages of blood types and Rh factors have a strong relationship to ethnicity in the US (abbreviations represent Caucasian, Latin American, African American, and Asian American), as we can see in the tables below.

	+	-
O	0.38	0.07
A	0.34	0.06
B	0.09	0.02
AB	0.03	0.01

Figure 2.3: Prevalence of blood types in the US

	Ca.	Af.	La.	As.
O+	0.37	0.47	0.53	0.39
O-	0.08	0.04	0.04	0.01
A+	0.33	0.24	0.29	0.27
A-	0.07	0.02	0.02	0.005
B+	0.09	0.18	0.09	0.25
B-	0.02	0.01	0.01	0.004
AB+	0.03	0.04	0.02	0.07
AB-	0.01	0.003	0.002	0.001

Figure 2.4: Prevalence of blood types by ethnicity

Ethnicity	US Pop
Ca.	0.63
La.	0.176
Af.	0.133
As.	0.056

Figure 2.5: US population by ethnicity.

(a) What is the probability an American has blood type A? Rh factor positive? O+ blood?

(b) What is the probability an American is Caucasian and has type O+ blood?

(c) If a randomly selected American has type O+ blood, what is the probability this person is Caucasian?

(d) How does $\mathbb{P}(\text{Ethnicity} \mid \text{Blood Type})$ relate to $\mathbb{P}(\text{Blood Type} \mid \text{Ethnicity})$

Solution

(a) For $\mathbb{P}(\text{American has blood type A})$, we can look at Figure 2.3 in the second line and add the Rh factors (+ and -) up:

$$\mathbb{P}(\text{American has blood type A}) = 0.34 + 0.06 = 0.40$$

For $\mathbb{P}(\text{Rh factor } +)$, Add up the column of (+) on Figure 2.3:

$$\mathbb{P}(\text{Rh factor } +) = 0.38 + 0.34 + 0.09 + 0.03 = 0.84$$

For $\mathbb{P}(O+)$, Simply look at the first column of the first row on Figure 2.3:

$$\mathbb{P}(O+) = 0.38$$

(b) To find $\mathbb{P}(\text{Caucasian} \cap \text{O+})$, we use the multiplication rule:

$$\mathbb{P}(\text{Caucasian} \cap \text{O+}) = \mathbb{P}(\text{Caucasian})\mathbb{P}(\text{O+} \mid \text{Caucasian}) = 0.63 \cdot 0.37 = 0.2331$$

To find $\mathbb{P}(\text{O+} \mid \text{Caucasian})$, we looked at Figure 2.4 and found the prevalence of blood type O+ within Caucasians. We found $\mathbb{P}(\text{Caucasian})$ by looking at Figure 2.5 and see the U.S. population of Caucasians.

(c) We use the Conditional Probability formula to find the probability an American is Caucasian given they have type O+ blood (using the answer we got from part (b)):

$$\mathbb{P}(\text{Caucasian} \mid \text{O+}) = \frac{\mathbb{P}(\text{Caucasian} \cap \text{O+})}{\mathbb{P}(\text{O+})} = \frac{0.2331}{0.38} \approx 0.6134$$

$\mathbb{P}(\text{O+})$ was found in part (a), or by just looking at Figure 2.3.

(d) $\mathbb{P}(\text{Ethnicity} \mid \text{Blood Type})$ and $\mathbb{P}(\text{Blood Type} \mid \text{Ethnicity})$ are not equal in general, but are related to each other by Bayes' Theorem:

$$\mathbb{P}(\text{Ethnicity} \mid \text{Blood Type}) = \frac{\mathbb{P}(\text{Blood Type} \mid \text{Ethnicity})\mathbb{P}(\text{Ethnicity})}{\mathbb{P}(\text{Blood Type})}$$

$$\mathbb{P}(\text{Blood Type} \mid \text{Ethnicity}) = \frac{\mathbb{P}(\text{Ethnicity} \mid \text{Blood Type})\mathbb{P}(\text{Blood Type})}{\mathbb{P}(\text{Ethnicity})}$$

Problem 10.

Consider a blood test for a disease. If the patient has the disease, test is positive with probability 0.97. If the patient does not have the disease, the test is negative with probability 0.92. 1 in 10,000 people have the disease. If a person tests positive, what is the probability he or she actually has the disease?

Solution

Let D be the set of patients having disease, and P be the set of patients testing positive. We are given:

- $\mathbb{P}(P \mid D) = 0.97$

- $\mathbb{P}(P^C \mid D^C) = 0.92 \implies \mathbb{P}(P \mid D^C) = 0.08$ (false positive)

- $\mathbb{P}(D) = \frac{1}{10000} = 0.0001 \implies \mathbb{P}(D^C) = 0.9999$

First, we need to calculate the total probability of a test being positive using the Law of Total Probability:

$$\mathbb{P}(P) = \mathbb{P}(P \mid D)\mathbb{P}(D) + \mathbb{P}(P \mid D^C)\mathbb{P}(D^C) = 0.97(0.0001) + 0.08(0.9999) = 0.080089.$$

Now we can use Bayes' Theorem to solve for the probability that a person has the disease given the test is positive $\mathbb{P}(D \mid P)$:

$$\mathbb{P}(D \mid P) = \frac{\mathbb{P}(P \mid D)\mathbb{P}(D)}{\mathbb{P}(P)} = \frac{0.97 \cdot 0.0001}{0.080089} \approx 0.0012 = 0.12\%$$

So if someone tests positive, the chance they actually have the disease is about 0.12%.