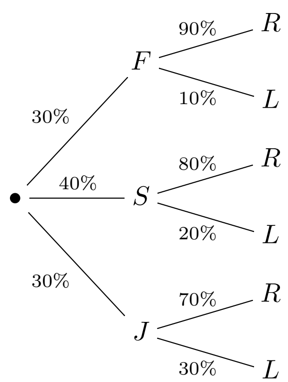
More on Conditional Probability

Example 1. In a class, 30% of the students are freshmen, 40% are sophomores, and the other 30% are juniors. Suppose that 90% of the freshmen, 80% of the sophomores, and 70% of the juniors in this class are right-handed. (Assume that each student is either left-handed or right-handed, and no one is ambidextrous.) Pick a student in this class at random. (a) Find the probability that the student is left-handed. (b) Suppose that the chosen student is left-handed, find the probability that he/she is a sophomore.

Solution. We start by drawing a tree diagram.



$$P(F \cap R) = P(F)P(R|F) = (0.3)(0.9) = 27\%$$
$$P(F \cap L) = P(F)P(L|F) = (0.3)(0.1) = 3\%$$
$$P(S \cap R) = P(S)P(R|S) = (0.4)(0.8) = 32\%$$
$$P(S \cap L) = P(S)P(L|S) = (0.4)(0.2) = 8\%$$
$$P(J \cap R) = P(J)P(R|J) = (0.3)(0.7) = 21\%$$
$$P(J \cap L) = P(J)P(L|J) = (0.3)(0.3) = 9\%$$

(a) We want to find the fraction of the students who are left-handed. We can see above, where we divide the class into 6 groups according to the year and the hand, that three groups are left-handed. So we simply add the fractions of these three groups to get the answer: $3\% + 8\% + 9\% = 20\%$. In terms of probability,

$$P(L) = P(F \cap L) + P(S \cap L) + P(J \cap L) = 3\% + 8\% + 9\% = 20\%.$$

(b) Now we know that the chosen student comes from these 20% of the class, and we want to find the fraction of sophomores *among these* 20%. Recall that 20% are made up from 3%, 8%, and 9%, who are freshmen, sophomores, and juniors, respectively. So the fraction of the students in these 20% who are sophomores is simply $\frac{8\%}{20\%} = \frac{8}{20} = 40\%$. In terms of probability,

$$P(S|L) = \frac{P(S \cap L)}{P(L)} = \frac{8\%}{20\%} = \frac{8}{20} = 40\%. \quad \blacksquare$$

Random Variables - Basics

A random variable is a quantity that depends on the outcome; it assigns a real number to each outcome in the sample space. To describe a (discrete) random variable, we can list all its possible values and the probability that each value occurs.

Example 2. Two boys each choose a number randomly from 1 to 3. There are $3 \times 3 = 9$ possible outcomes. Let's consider the maximum of the two numbers that they choose. Let's call it X. We see that X is a random variable because X varies according to the outcome:

outcome	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
X	1	2	3	2	2	3	3	3	3

X takes on three possible values: 1, 2 or 3. As the nine outcomes listed above are equally likely, we have that $P(X = 1) = \frac{1}{9}$, $P(X = 2) = \frac{3}{9}$ and $P(X = 3) = \frac{5}{9}$. (There are 1, 3 and 5 outcomes such that X = 1, 2 and 3, respectively.)

Expected Value. The expected value is the average of a random variable. Continuing from Example 2, we have that

$$\begin{aligned} EX &= 1 \cdot P(X = 1) + 2 \cdot P(X = 2) + 3 \cdot P(X = 3) \\ &= 1 \cdot \frac{1}{9} + 2 \cdot \frac{3}{9} + 3 \cdot \frac{5}{9} = \frac{22}{9}. \end{aligned}$$

We can find the expected value of any function of X. For example, using again Example 2,

$$\begin{aligned} E(X^2) &= 1^2 P(X = 1) + 2^2 P(X = 2) + 3^2 P(X = 3) \\ &= 1 \cdot \frac{1}{9} + 2^2 \cdot \frac{3}{9} + 3^2 \cdot \frac{5}{9} = \frac{58}{9}. \end{aligned}$$

Variance. The variance roughly measures how much the values of a random variable fluctuate. To compute, we can use the formula $\boxed{\operatorname{var} X = E(X^2) - (EX)^2}$. From Example 2, we have that $\operatorname{var} X = E(X^2) - (EX)^2 = \frac{58}{9} - \left(\frac{22}{9}\right)^2 = \frac{38}{81}$.

Properties

- $E(aX + b) = a(EX) + b$.
- $\operatorname{var}(aX + b) = a^2 \operatorname{var} X$.
- $E(X + Y) = EX + EY$. (linearity of expectation)

Notes

- The expected value and the variance do not depend on the outcome.
- The expected value and the variance might not exist for some random variables.
- We do not have linearity of variance in general. However, we will see that if X and Y are independent random variables, then $\operatorname{var}(X+Y) = \operatorname{var} X + \operatorname{var} Y$.

Exercises

1. (a) 6% of the population have a certain type of disease. Suppose that 85% of the screening tests on the people with the disease show positive results, and 10% of the tests on the people without the disease show positive results (aka false positives). Given a person with a positive test result, what is the probability that this person actually has the disease?

   (b) From part (a), suppose that we are not given the prevalence of the disease (i.e. the number 6% is not given) but instead we know that the probability that the screening test on a random person shows a positive result is 25%. What is the actual prevalence of the disease?

   (c) From part (a), suppose that we are not given the prevalence of the disease, but instead we know that the probability that a person with a positive test result actually has the disease is 80%. What is the actual prevalence of the disease?

2. In Example 2, two boys each choose a number randomly from 1 to 3. We have defined X to be the maximum of the two chosen numbers. Here, let's define Y to be the minimum of the two numbers, and Z to be the difference of the two numbers.

outcome	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)	(3,1)	(3,2)	(3,3)
X	1	2	3	2	2	3	3	3	3
Y	1	1							
Z	0	1							

   (a) Complete the table above.

   (b) Compute EY, EZ, $\operatorname{var} Y$, and $\operatorname{var} Z$.

   (c) Explain why $X = Y + Z$, then verify that $EX = EY + EZ$.

   (d) Is $\operatorname{var} X$ equal to $\operatorname{var} Y + \operatorname{var} Z$?

3. You have a biased coin that shows heads 2/3 of the time. Toss this coin three times. Let X be the number of times it shows heads. Complete the table below. (Note that the listed outcomes are not equally likely.) Then compute EX.

outcome	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
probability	$\left(\frac{2}{3}\right)^3$							
X	3							

4. Roll a fair, regular six-sided die (with numbers 1, 2, ... , 6) four times. For each roll that you make except the first roll, if you get the same number as the previous number, you gain \$6; otherwise you lose \$1. For example, if you get (3,3,5,3), you gain \$6 on the second roll, but lose \$1 on the third and the fourth rolls, making the net gain of \$4. If you have to pay to play this game, how much should one game cost so that the game is fair or even in your favor? (Hint: Consider the amount of money you gain on *each* roll.)