# Probability

Computational Mathematics and Statistics Camp
University of Chicago
September 2018

1. If you flip a fair coin 10 times, what is the probability of

   a. getting all tails?

      Multiplication rule for independent processes

      $$P(\text{all tails}) = 0.5^{10} = 0.00098$$

   b. getting all heads?

      Multiplication rule for independent processes

      $$P(\text{all heads}) = 0.5^{10} = 0.00098$$

   c. getting at least one tails?

      Multiplication rule for independent processes

      $$P(\text{all least one tails}) = 1 - P(\text{no tails}) = 1 - (0.5^{10}) \approx 1 - 0.001 = 0.999$$

2. If you roll a pair of fair dice, what is the probability of

   a. getting a sum of 1?

      $$P(\text{sum of } 1) = 0$$

      Since two dice are being rolled, the minimum possible sum is 2. I.e. 1 is not in the **set** of possible outcomes

   b. getting a sum of 5?

      General addition rule

      $$P(\text{sum of } 5) = P(1,4) + P(2,3) + P(3,2) + P(4,1) = \left(\frac{1}{6} \times \frac{1}{6}\right) \times 4 \approx 0.11$$

   c. getting a sum of 12?

      Multiplication rule for independent processes

      $$P(\text{sum of } 12) = P(6,6) = \left(\frac{1}{6} \times \frac{1}{6}\right) = \frac{1}{36} \approx 0.0278$$

3. In parts(a) and(b), identify whether the events are disjoint, independent, or neither (events cannot be both disjoint and independent).

   a. You and a randomly selected student from your class both earn A's in this course.

      If the class is not graded on a curve, they are independent. If graded on a curve, then neither independent nor disjoint – unless the instructor will give only one A, which is a situation we will ignore in parts (b) and (c).

b. You and your class study partner both earn A's in this course.

   They are probably not independent – if you study together, your study habits would be related, which suggests your course performances are also related.

c. If two events can occur at the same time, must they be dependent?

   No. See the answer to part (a) when the course is not graded on a curve. More generally: if two things are unrelated (independent), then one occurring does not preclude the other from occurring.

4. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.

   a. What is the probability that a student chosen at random doesn't miss any days of school due to sickness this year?
   $$P(\text{no misses}) = 1 - (0.25 + 0.15 + 0.28) = 0.32$$

   b. What is the probability that a student chosen at random misses no more than one day?
   $$P(\text{at most 1 miss}) = P(\text{no misses}) + P(\text{1 miss}) = 0.32 + 0.25 = 0.57$$

   c. What is the probability that a student chosen at random misses at least one day?
   $$P(\text{at least 1 miss}) = P(\text{1 miss}) + P(\text{2 misses}) + P(\text{3+ misses}) = 1 - P(\text{no misses}) = 1 - 0.32 = 0.68$$

   d. If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

      Assume that whether or not one kid misses school is independent of the other (not necessarily a safe assumption).
   $$P(\text{neither miss any}) = P(\text{no miss}) \times P(\text{no miss}) = 0.322 = 0.1024$$

   e. If a parent has two kids at a DeKalb County elementary school, what is the probability that both kids will miss some school, i.e. at least one day? Note any assumption you make.

      Assume that whether or not one kid misses school is independent of the other (not necessarily a safe assumption).
   $$P(\text{both miss some}) = P(\text{at least 1 miss}) \times P(\text{at least 1 miss}) = 0.68^2 = 0.4624$$

   f. If you made an assumption in part(d) or(e), do you think it was reasonable? If you didn't make any assumptions, double check your earlier answers.

      These kids are siblings, and if one gets sick it probably raises the chance that the other one will get sick as well. So whether or not one misses school due to sickness is probably not independent of the other.

5. $P(A) = 0.3$, $P(B) = 0.7$

   a. Can you compute P(A and B) if you only know P(A) and P(B)?

      No, we cannot compute $P(A \text{ and } B)$ since we do not know if $A$ and $B$ are independent. We could if A and B were independent.

   b. Assuming that events A and B arise from independent random processes,

      a. what is $P(A \text{ and } B)$?
      $$P(A \text{ and } B) = P(A) \times P(B) = 0.21$$

b. what is $P(A \text{ or } B)$?

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B) = 0.3 + 0.7 - 0.21 = 0.79$$

c. what is $P(A|B)$?

$$P(A|B) = P(A) = 0.3$$

c. If we are given that $P(A \text{ and } B) = 0.1$, are the random variables giving rise to events A and B independent?

No, because $0.1 \neq 0.21$.

d. If we are given that $P(A \text{ and } B) = 0.1$, what is $P(A|B)$?

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.1}{0.7} = 0.143$$

6. Suppose 80% of people like peanut butter, 89% like jelly, and 78% like both. Given that a randomly sampled person likes peanut butter, what's the probability that he also likes jelly?

$$P(J|PB) = \frac{P(PB \text{ and} J)}{P(PB)} = \frac{0.78}{0.80} = 0.975$$

7. A 2010 SurveyUSA poll asked 500 Los Angeles residents, "What is the best hamburger place in Southern California? Five Guys Burgers? In-N-Out Burger? Fat Burger? Tommy's Hamburgers? Umami Burger? Or somewhere else?" The distribution of responses by gender is shown below.

| | | *Gender* | | |
|---|---|---|---|---|
| | | Male | Female | Total |
| *Best hamburger place* | Five Guys Burgers | 5 | 6 | 11 |
| | In-N-Out Burger | 162 | 181 | 343 |
| | Fat Burger | 10 | 12 | 22 |
| | Tommy's Hamburgers | 27 | 27 | 54 |
| | Umami Burger | 5 | 1 | 6 |
| | Other | 26 | 20 | 46 |
| | Not Sure | 13 | 5 | 18 |
| | Total | 248 | 252 | 500 |

a. Are being female and liking Five Guys Burgers mutually exclusive?

No. There are 6 females who like Five Guys Burgers.

b. What is the probability that a randomly chosen male likes In-N-Out the best?

$$P(\text{In-N-Out}|\text{male}) = \frac{162}{248} \approx 0.65$$

c. What is the probability that a randomly chosen female likes In-N-Out the best?

$$P(\text{In-N-Out}|\text{female}) = \frac{181}{252} \approx 0.72$$

d. What is the probability that a man and a woman who are dating both like In-N-Out the best? Note any assumption you make and evaluate whether you think that assumption is reasonable.

Under the assumption of independence of gender and hamburger preference:

$$P(\text{man and woman dating both like In-N-Out burgers the best}) = 0.65 \times 0.72 = 0.468$$

While it is possible there is some mysterious connection between burger choice and finding a partner, independence is probably a reasonable assumption.

e. What is the probability that a randomly chosen person likes Umami best or that person is female?

$$P(\text{Umami or female}) = P(\text{Umami}) + P(\text{female}) - P(\text{Umami and female}) = \frac{6 + 252 - 1}{500} = 0.514$$

8. Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease. The test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease. There is a line from the Fox television show **House** that is often used after a patient tests positive for lupus: "It's never lupus." Do you think there is truth to this statement? Use appropriate probabilities to support your answer.

   Given the probabilities stated in the problem, we can deduce the following:

$$\begin{aligned} P(\text{lupus and positive}) &= P(\text{lupus}) \times P(\text{positive}|\text{lupus}) = 0.02 \times 0.98 = 0.0196 \\ P(\text{lupus and negative}) &= P(\text{lupus}) \times P(\text{negative}|\text{lupus}) = 0.02 \times 0.02 = 0.0004 \\ P(\text{not lupus and positive}) &= P(\text{not lupus}) \times P(\text{positive}|\text{not lupus}) = 0.98 \times 0.26 = 0.2548 \\ P(\text{not lupus and negative}) &= P(\text{not lupus}) \times P(\text{negative}|\text{not lupus}) = 0.98 \times 0.74 = 0.7252 \end{aligned}$$

   The conditional probability $P(\text{lupus}|\text{positive})$ is a function of the joint probability $P(\text{lupus and positive})$ and the overall probability of a positive test result $P(\text{positive})$.

$$\begin{aligned} P(\text{lupus}|\text{positive}) &= \frac{P(\text{lupus and positive})}{P(\text{positive})} \\ &= \frac{0.0196}{0.0196 + 0.2548} \\ &= 0.0714 \end{aligned}$$

   Even when a patient tests positive for lupus, there is only a 7.14% chanve that he actually has lupus.

9. Suppose we pick three people at random. For each of the following questions, ignore the special case where someone might be born on February 29th, and assume that births are evenly distributed throughout the year.

   a. What is the probability that the first two people share a birthday?

$$P(\text{first two people share a birthday}) = 1/365 = 0.0027$$

   b. What is the probability that at least two people share a birthday?

$$\begin{aligned} P(\text{at least one pair of people share a birthday}) &= 1 - P(\text{none of the three people share a birthday}) \\ &= 1 - (\frac{365}{365})(\frac{364}{365})(\frac{363}{365}) = 0.0082 \end{aligned}$$

10. A portfolio's value increases by 18% during a financial boom and by 9% during normal times. It decreases by 12% during a recession. What is the expected return on this portfolio if each scenario is equally likely?

| Event | $X$ | $P(X)$ | $X \times P(X)$ |
|---|---|---|---|
| Boom | 0.18 | $\frac{1}{3}$ | $0.18 \times \frac{1}{3} = 0.06$ |
| Normal | 0.09 | $\frac{1}{3}$ | $0.09 \times \frac{1}{3} = 0.03$ |
| Recession | -0.12 | $\frac{1}{3}$ | $-0.12 \times \frac{1}{3} = -0.04$ |

$$E(X) = 0.06 + 0.03 - 0.04 = 0.05$$

The expected return is a 5% increase in value

11. Sally gets a cup of coffee and a muffin every day for breakfast from one of the many coffee shops in her neighborhood. She picks a coffee shop each morning at random and independently of previous days. The average price of a cup of coffee is \$1.40 with a standard deviation of 30¢(\$0.30), the average price of a muffin is \$2.50 with a standard deviation of 15¢, and the two prices are independent of each other.

    a. What is the mean and standard deviation of the amount she spends on breakfast daily?

    Let $X$ represent the amount Sally spends on coffee (in ¢), and $Y$ represent the amount she spends on muffins (in ¢).

$$E(X) = 140 \quad E(Y) = 250$$
$$SD(X) = 30 \quad SD(Y) = 15$$
$$V(X) = 30^2 = 900 \quad V(Y) = 15^2 = 225$$

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) = 140 + 250 = 390 = 3.90 \\ V(X+Y) &= V(X) + V(Y) = 900 + 225 = 1125 = 1.125 \\ SD(X+Y) &= \sqrt{V(X) + V(Y)} = \sqrt{900 + 225} = \sqrt{1125} = 34 = 0.34 \end{aligned}$$

    The mean is \$3.90 and the standard deviation is \$0.34 (or 34¢)

    b. What is the mean and standard deviation of the amount she spends on breakfast weekly (7 days)?

    Let $W$ represent the amount Sally spends on coffee and breakfast each week. Then,

$$W = (X_1 + Y_1) + \ldots + (X_7 + Y_7)$$

$$\begin{aligned} E(X+Y) &= E((X_1 + Y_1) + \ldots + (X_7 + Y_7)) = 3.90 + \ldots + 3.90 = 7 \times 3.90 = 27.30 \\ V(X+Y) &= V((X_1 + \ldots + X_7)) + V((Y_1 + \ldots + Y_7)) = 1.125 \times 7 = 7.875 \\ SD(X+Y) &= \sqrt{7.875} = 0.89 \end{aligned}$$

    The mean is \$27.30 and the standard deviation is \$0.89 (or 89¢)