

Solutions to In-Class Problems Week 13, Mon.

Problem 1. Suppose that you flip three fair, mutually independent coins. Define the following events:

- Let A be the event that *the first* coin is heads.
- Let B be the event that *the second* coin is heads.
- Let C be the event that *the third* coin is heads.
- Let D be the event that *an even number of* coins are heads.

(a) Are these events pairwise independent?

Solution. The sample space consists of eight, equiprobable outcomes:

$$\{H, T\}^3 = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Each sequence of three symbols specifies the outcomes of the three coin tosses.

The first three events (A , B , and C) are pairwise independent, since they are mutually independent. All that remains is to check that each of these is independent of D . By symmetry, we need only check just one of the three, say A :

$$\Pr\{A\} = \frac{1}{2}$$

$$\Pr\{D\} = \Pr\{\{HHT, HTH, THH, TTT\}\} = \frac{1}{2}$$

$$\Pr\{A \cap D\} = \Pr\{\{HHT, HTH\}\} = \frac{1}{4}$$

Therefore, $\Pr\{A \cap D\} = \Pr\{A\} \cdot \Pr\{D\}$, and so these events are independent. We conclude that all four events are pairwise independent. ■

(b) Are these events three-way independent? That is, does

$$\Pr \{X \cap Y \cap Z\} = \Pr \{X\} \cdot \Pr \{Y\} \cdot \Pr \{Z\}$$

always hold when X , Y , and Z are distinct events drawn from the set $\{A, B, C, D\}$?

Solution. Because the coin tosses are mutually independent, we know:

$$\Pr \{A \cap B \cap C\} = \Pr \{A\} \cdot \Pr \{B\} \cdot \Pr \{C\}.$$

What remains is to check that equality holds for the other subsets of three events: $\{A, B, D\}$, $\{A, C, D\}$, and $\{B, C, D\}$. By symmetry, again, we need only check one, say the first one.

$$\Pr \{A \cap B \cap D\} = \Pr \{\{HHT\}\} = \frac{1}{8}.$$

Since this is equal to $\Pr \{A\} \cdot \Pr \{B\} \cdot \Pr \{D\}$, these three events are independent.

We conclude that all four events are three-way independent. ■

(c) Are these events mutually independent?

Solution. No, because:

$$\Pr \{A \cap B \cap C \cap D\} \neq \Pr \{A\} \cdot \Pr \{B\} \cdot \Pr \{C\} \cdot \Pr \{D\}$$

The probability on the left is zero, but the product on the right is equal to $1/16$. ■

Problem 2. There is a set of 1000 people.

- The favorite color of 20% of these people is blue.
- The favorite color of 30% is green.
- The favorite color of 50% is red.

(a) Suppose we select a sequence of two distinct people (p_1, p_2) by selecting p_1 uniformly at random, and then selecting p_2 uniformly at random among the remaining 999 people. Let the random variable C_i be the favorite color of person p_i . Are C_1 and C_2 independent? Justify your answer.

Solution. No. For example,

$$\Pr \{C_1 = \text{blue}\} \Pr \{C_2 = \text{blue}\} = \frac{200}{1000}.$$

However,

$$\Pr \{C_2 = \text{blue} \mid C_1 = \text{blue}\} = \frac{199}{999}.$$

since 199 of the remaining 999 people like blue after one person who likes blue is selected.

■

(b) Suppose we select a sequence of two people (q_1, q_2) uniformly at random. (Now it may be that $q_1 = q_2$.) Let the random variable D_i be the favorite color of q_i . Now are D_1 and D_2 independent? Justify your answer.

Solution. Yes. Let $c(p)$ be the color that person p likes. The random variables q_1 and q_2 are independent. Functions of independent random variables are independent, so $C_1 ::= c(q_1)$ and $C_2 ::= c(q_2)$ are independent. ■

Problem 3. Suppose that we create a national database of DNA profiles. Let's make some simplifying assumptions (for the purposes of this problem, don't worry about whether these assumptions are reasonable):

- Each person can be classified into one of 20 billion different "DNA types". (For example, you might be type #13,646,572,661 and the person next to you might be type #2,785,466,098.)
- Each DNA type is equally probable.
- The DNA types of Americans are mutually independent.

(a) A congressman argues that there are only about 250 million Americans, so even if a profile for every American were stored in the database, the probability of even one coincidental match would be very small. How many profiles must the database actually contain in order for the probability of at least one coincidental match be about 1/2?

Solution. By the birthday principle, the probability of a match is around half when the number of entries is:

$$\sqrt{2 \ln 2 \cdot 20,000,000,000} = 166,511.$$

■

(b) Person x is arrested for a crime that was committed by person y . At trial, jurors must determine whether $x = y$. The crime lab says x and y have the same DNA type. The prosecutor argues that the probability that x and y are different people is only 1 in 20 billion. Let $T(x)$ denote the DNA type of person x . Using T , write the prosecutor's assertion in mathematical notation, and explain her error.

Solution. The prosecutor is asserting that:

$$\Pr \{x \text{ and } y \text{ are different people} \mid T(x) = T(y)\} = 2 \cdot 10^{-10}.$$

This assertion is false using the probability model that is implicitly being assumed. Namely, we associate each person with a DNA-type-valued random variable with uniform distribution over the $2 \cdot 10^{10}$ types, and we assume all these variables are mutually independent. We also consider x and y to be distinct people-valued independent random variables with uniform distribution over the $2.5 \cdot 10^8$ people in the US.

With this model, the actual probability is

$$\Pr \{x \neq y \mid T(x) = T(y)\} = \frac{1}{126}, \quad (1)$$

as we show below.

The prosecutor's argument *sounds* confusingly similar to a correct assertion:

$$\Pr \{T(x) = T(y) \mid x \neq y\} = 2 \cdot 10^{-10}.$$

The prosecutor can validly argue that *either* an amazing 1-in-20 billion coincidence involving DNA has occurred *or else* x and y are the same person. On this basis, a jury might conclude that x is guilty.

On the other hand, the defense attorney could argue from (1) that, given the DNA match, the probability that x is not guilty is much greater $2 \cdot 10^{-10}$, but is in fact $1/126$. He could then argue to the jury that $125/126$ does not meet the legal standard of being guilty "beyond reasonable doubt," and so the jury should not convict x .

To prove (1), note that

$$\Pr \{x \neq y \mid T(x) = T(y)\} = 1 - \Pr \{x = y \mid T(x) = T(y)\}.$$

Then we have,

$$\begin{aligned}
 & \Pr \{x = y \mid T(x) = T(y)\} \\
 & ::= \frac{\Pr \{x = y \text{ and } T(x) = T(y)\}}{\Pr \{T(x) = T(y)\}} \\
 & = \frac{\Pr \{x = y\}}{\Pr \{T(x) = T(y) \mid x = y\} \Pr \{x = y\} + \Pr \{T(x) = T(y) \mid x \neq y\} \Pr \{x \neq y\}} \\
 & = \frac{2.5 \cdot 10^{-8}}{1 \cdot 2.5 \cdot 10^{-8} + 2 \cdot 10^{-10}(1 - 2.5 \cdot 10^{-8})} \\
 & \approx \frac{2.5 \cdot 10^{-8}}{2.5 \cdot 10^{-8} + (2.5 \cdot 10^{-8}/125) \cdot 1} \\
 & = \frac{1}{1 + 1/125} = \frac{125}{126}.
 \end{aligned}$$

■

Problem 4. Suppose there are 100 people in a room. Assume that their birthdays are independent and uniformly distributed. As stated in lecture:

If there are N days in a year and m people in a room, then the probability that no two people in the room have the same birthday is very close to:

$$e^{-m^2/(2N)}$$

It follows that with probability $> 99\%$ there will be two people with the same birthday.

Now suppose you learn the birthdays of all the people in the room except one—call her “Jane”—and find all 99 dates to be different.

(a) What’s wrong with the following argument:

With probability greater than 99%, some pair of people in the room have the same birthday. Since the 99 people we asked all had different birthdays, it follows that with probability greater than 99% Jane has the same birthday as some other person in the room.

Solution. Here’s the problem with the argument: Let A be the event that some two people in the room have the same birthday. Let B be the event that the 99 people we asked all have different birthdays.

It is true that $\Pr\{A\} > 0.99$ (that is indeed the probability spoken about in the lecture notes). However, that is the *a priori* probability, that is, assuming all the birthdays are uniform and independent, with no other constraints. The argument above makes the erroneous assumption that event A has probability of at least 99% even once we know that event B holds. But once we know that event B holds, the birthdays are no longer independent. Thus $\Pr\{A \mid B\}$ is not necessarily equal to $\Pr\{A\}$ (in our case, they are actually quite different, as will be computed in part (b)). ■

(b) What is the actual probability that Jane has the same birthday as some other person in the room?

Solution. Let S be the set of birthdays of the 99 people in the room other than Jane. By assumption, $|S| = 99$. Let b be the date of Jane's birthday. Since b is uniformly distributed over a set of size 365, and b is independent of all the birthdays in S , we have

$$\Pr\{A \mid B\} = \Pr\{b \in S\} = \frac{|S|}{365} = \frac{99}{365} \approx 27.1\%,$$

where B is the event that the 99 people we asked all have different birthdays. ■