

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?

(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 $\{A, B, C, D\}$?

(c) Are these events mutually independent?

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.

(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.

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 $1/2$?

(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.

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.

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