

Notes for Recitation 19

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 :

$$\begin{aligned}\Pr(A) &= \frac{1}{2} \\ \Pr(D) &= \Pr(\{HHT, HTH, THH, TTT\}) = \frac{1}{2} \\ \Pr(A \cap D) &= \Pr(\{HHT, HTH\}) = \frac{1}{4}\end{aligned}$$

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 $\frac{1}{16}$.

Fact from 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 about:

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

Problem 2. Suppose that we create a national database of DNA profiles. Let's make some simplistic assumptions:

- 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.) Let $T(x)$ denote the type of person x .
- 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. 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 at best false and arguably not even a well-formed mathematical statement. Either x and y are the same person or different people, regardless of the DNA types of all the people in the world. Thus, either x and y are the same person in every outcome or they are different people. Consequently, the probability above is either 0 or 1, but we don't know which.

The prosecutor's argument *sounds* confusingly similar to a correct assertion: if x and y are different people, then:

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

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 almost surely guilty, but there is nothing in our probability model to justify that conclusion directly.

Problem 3. Suppose there are 100 people in a room. Assume that their birthdays are independent and uniformly distributed. As stated in lecture notes, with probability $> 99\%$ there will be two that have the same birthday.

Now suppose you find out 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, i.e., 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 $\Pr(A | B)$ 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

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

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

Problem 4. There were n Immortal Warriors born into our world, but in the end *there can be only one*. The Immortals' original plan was to stalk the world for centuries, dueling one another with ancient swords in dramatic landscapes until only one survivor remained. However, after a thought-provoking discussion of probabilistic independence, they opt to give the following protocol a try:

1. The Immortals forge a coin that comes up heads with probability p .
2. Each Immortal flips the coin once.
3. If *exactly one* Immortal flips heads, then he or she is declared The One. Otherwise, the protocol is declared a failure, and they all go back to hacking each other up with swords.

(a) One of the Immortals (the Kurgan from the Russian steppe) argues that as n grows large, the probability that this protocol succeeds must tend to zero. Another (McLeod from the Scottish highlands) argues that this need not be the case, provided p is chosen *very carefully*. What does your intuition tell you?

Solution. Your intuition tells you that a short nap would be nice right now. As would a couple cookies to dunk in a cold glass of milk.

(b) What is the probability that the experiment succeeds as a function of p and n ?

Solution. The sample space consists of all possible results of n coin flips, which we can represent by the set $\{H, T\}^n$. Let E be the event that the experiment successfully selects The One. Then E consists of the n outcomes which contain a single head. In general, the probability of an outcome with h heads and $n - h$ tails is:

$$p^h(1 - p)^{n-h}$$

Summing the probabilities of the n outcomes in E gives the probability that the procedure succeeds:

$$\Pr(E) = np(1 - p)^{n-1}$$

(c) How should p , the bias of the coin, be chosen in order to maximize the probability that the experiment succeeds? (You're going to have to compute a derivative!)

Solution. We compute the derivative of the success probability:

$$\frac{d}{dp} np(1 - p)^{n-1} = n(1 - p)^{n-1} - np(n - 1)(1 - p)^{n-2}$$

Now we set the right side equal to zero to find the best probability p :

$$\begin{aligned}n(1-p)^{n-1} &= np(n-1)(1-p)^{n-2} \\(1-p) &= p(n-1) \\p &= 1/n\end{aligned}$$

This answer makes sense, since we want the coin to come up heads exactly 1 time in n .

(d) What is the probability of success if p is chosen in this way? What quantity does this approach when n , the number of Immortal Warriors, grows large?

Solution. Setting $p = 1/n$ in the formula for the probability that the experiment succeeds gives:

$$\Pr(E) = \left(1 - \frac{1}{n}\right)^{n-1}$$

In the limit, this tends to $1/e$. McLeod is right.