CSCI104L Lecture 13: Independence

Two events E and F are **independent** if and only if $p(E|F) = p(E)$. Equivalently, $p(E \cap F) = p(E)p(F)$.

Question 1. A family has two children, the probability of each being a boy or girl is uniformly generated. E is the event that the family has two boys. F is the event that the family has at least one boy. Are E and F independent?

Question 2. A family has three children, again uniformly generated. E is the event that the family has both a boy and a girl. F is the event that the family has at most 1 boy. Are E and F independent?

Question 3. There are 4 people in line, and each person is either a pirate or a ninja generated uniformly at random. E is the event that the first person in line is a ninja. F is the event that there are an even number of ninja. Are E and F independent?

A **Bernoulli trial** is an experiment with two outcomes that do not necessarily have equal probability (for example, a biased coin). We consider one outcome the **success** and the other outcome the **failure**.

The probability of exactly k successes in n independent Bernoulli trials with probability of success p and failure $q = 1 - p$ is $C(n, k)p^k q^{n-k}$. This should look familiar from the Binomial Theorem.

Question 4. A coin is biased so that probability of heads is $\frac{2}{3}$. Given 4 mutually independent flips, what is the probability that we get exactly 2 heads?

Question 5. There are 10 people in a line. Each person has a 90% chance of being a pirate, and a 10% chance of being a ninja. Everyone is exactly one or the other; no one is both. What is the probability that there are exactly 8 pirates?

Question 6. The Birthday Problem

- What are the odds that two people in this room share the same birthday?
- Given n people, what is the probability that no two have the same birthday?

  For this, assume that any given day of the 366 possible birthdays is equally likely (not actually true).

A **Random Variable** is a function from the set of outcomes to the set of real numbers. That is, it assigns a real number to each possible outcome.
A coin is flipped 3 times. Let $X(t)$ be the random variable which outputs the number of heads. Then: X(HHH) = 3, X(HHT)=2, X(HTH)=2, ..., X(TTT) = 0
A Random Variable is a **function**. It is not random (though the input to the function *is* random), and it is not a variable. The name is something of a misnomer: the concept is dirt simple, but it sounds like something very different than it actually is.
The **distribution** of a random variable is the probability distribution over all possible outputs.

Question 7. What is the distribution of the above random variable?

Question 8. Roll 2 fair dice. Let $X(t)$ be the random variable which outputs their sum. What is the distribution of X?

Question 9. In 1940, two mathematicians (W. Feller and J.L. Doob) were trying to decide whether both would use the term "random variable" or "chance variable" to describe this concept in the books they were writing. Given random variables are based on probability, how do you suppose they decided?

Application: Randomized Algorithms

You order a batch of processor chips of size n. Due to feasibility constraints, the manufacturer has tested some of the batches of processor chips, but not all of them. If the manufacturer tested the batch, then all of the chips will work (because faulty ones were replaced). If the manufacturer did not test the batch, then the probability that an arbitrary chip is bad is 0.1.

You want to determine if your batch of chips have been tested (that is, they're all good). You could obviously test every chip, but this will take $O(n)$ time.

You could instead rely on probability. You could test k of the chips (chosen randomly). If all tested chips work, you claim the batch has been tested.

- If the manufacturer tested the chips, what is the probability you correctly identify this?
- If the manufacturer did not test the chips, what is the probability you get a *false positive*?

This is an example of a randomized algorithm. Specifically, it is a **Monte Carlo** algorithm. A **Monte Carlo** algorithm will always return an answer to a question quickly, but has some small probability that it returns the incorrect answer. A **Las Vegas** algorithm will always return the correct answer, but has some small probability that it will take a long time.