CSCI104L Lecture 14: Independence and Bayes' Theorem

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 n processor chips. The manufacturer tested some of batches, but not others. If the manufacturer tested your batch, then all of the chips will work. 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 there are any bad chips in your batch. If you test them all, this will take $O(n)$.

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 a **Monte Carlo** randomized algorithm. A **Monte Carlo** algorithm always returns an answer quickly, but sometimes returns a wrong answer. A **Las Vegas** algorithm always returns the correct answer, but sometimes takes a long time.

Fallacies and Paradoxes

The Prosecutor's Fallacy: You are accused of speeding. The prosecutor notes that you have a radar detector (which detects when police are using their radar to identify speeders). The prosecutor notes that 80% of speeders have radar detectors, so you're probably guilty. Is there a flaw in his reasoning?

Simpson's Paradox: Alice and Bob are studying for the 104 midterm, each doing 25 practice problems split over 2 types (graphs and probability).

- Bob does 5 probability problems and gets 3 right. Alice does 20 probability problems and gets 13 right.
- Bob does 20 graph problems and gets 15 right. Alice does 5 graph problems and gets 4 right.
- Bob got 60% on probability, and Alice got 65%. Bob got 75% on graphs, and Alice got 80%. Alice did better on each.
- Bob got 72% overall and Alice got 68%. Bob did better overall.

Bayes' Theorem

Question 10. There are two boxes. The first contains two gold balls and seven cardinal balls. The second contains four gold balls and three cardinal balls. You choose a box at random, then you choose a ball at random. You draw out a cardinal ball. What is the probability that the ball came from the first box?

Given two events E and F with non-zero probability, then:

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\overline{F})p(\overline{F})}$$

Question 11. There is a rare disease which infects only 1 out of $100,000$ people. You can detect it with a very accurate diagnostic test. If someone has the disease, it correctly identifies it 99% of the time. If someone does not have the disease, it correctly states this 99.5% of the time.

- Suppose the test comes out negative. What is the probability the person does not have the disease?
- Suppose the test comes out positive. What is the probability the person does have the disease?