Assignment 2: More interesting probability calculations

Due Tuesday, January 19 by 4:00 pm at 253 Sloan
(January 18 is Martin Luther King Day, an Institute holiday)

Instructions: For each exercise please rate its difficulty (on a scale of your choosing—just explain it), and record how much time you spent on it.

When asked for a probability, give both a formula and an explanation for why you used that formula, and also give a numerical value when available. When asked for a numerical probability, evaluate the formula numerically.

Recall the role of optional exercises. Grades will calculated without taking the optional exercises into account, but the maximum grade will be an A. If you want an A+, you will have to earn an A and also accumulate sufficiently many optional points. No collaboration is allowed on optional exercises.

For some of these problems you may find some scientific computing software, such as Mathematica, useful. Feel free to use such a program.

Exercise 1 (The World Series) (50 pts) The “World Series” is a tournament between the champion of the USA’s National League and American League to decide the U.S. Major League Baseball champion. At present, it is won by the first team to win four games out of a possible seven. Since baseball games do not end in ties, at most seven games are ever played.

\footnote{This is a lie. In the dim past (before night baseball) there were three World Series that had a tie game, when the game was called (shortened) on account of darkness. There were also three Series that used a best-of-nine format.}
It is often said that “baseball is a game of inches.” This means that small changes in the physical outcomes of a given play can lead to loss or victory. It also means that the outcome of a game between two teams is effectively random. Let us say that if the probability \( p \) that Team A beats Team B is strictly greater than \( 1/2 \), then Team A is better than Team B. Note that it is possible (with probability \( 1 - p \)) for the better team to lose a game. Frederick Mosteller estimated (based on data from 44 Series from the first half of the 20th century) that the probability that the better team wins any given World Series game is 0.65 and that the outcomes of the games are stochastically independent. I redid his calculation for all 108 Series through 2012 and came up with 0.59. (You will have a chance to figure this out later in the course with data through the 2015 Series.)

Let \( p \) be the probability that Team A wins any given game. Assume that it is the same for every game, and that the outcomes of the game are stochastically independent. The probability that Team B wins is thus \( 1 - p \).

We can describe the general rule to determine the winner in two ways. Either the winner is first team to win \( m \) games, or as the team that wins the most out of \( 2m - 1 \) games. In practice, the series is over as soon as one team wins \( m \) games.

1. What is the probability that Team A wins the series in exactly \( m \) games? (Give the formula, and explain it.)

2. What is the probability that Team B wins the series in exactly \( m \) games?

3. What is the probability that the series is over in exactly \( m \) games?

4. What is the probability that Team A wins in exactly \( m + 1 \) games?

5. What is the probability that the series lasts exactly \( 2m - 1 \) games?

6. What is the probability that Team A wins the series by being the first team to win \( m \) games?

7. Suppose the rule was that the teams had to play all \( 2m - 1 \) games. What is the probability that Team A wins the series? What interesting algebraic fact does this prove?

8. What is the probability that the Team A wins a best-of-7 series \( (m = 4) \) if \( p = 0.65 \)? (Remember to give both a formula and a numeric answer.)

Exercise 2 (Sampling schemes) (50 pts)

A child is selected at random from a group of children. What is the probability that it is the first born in its family? For simplicity, we will only consider families that have
children. (Who is the firstborn in a childless family?) (Note that even with multiple births, e.g., twins, one will be born before the others.)

The answer depends on the sampling scheme. Consider the two following schemes:

1. There is an urn for each family, which contains all the children in that family. A family is selected at random, and then a child is selected randomly from the family urn.

2. All the children are put into one urn, and a child is selected at random.

One of these schemes always yields a greater probability of finding a firstborn child than the other, with equality only if all families are the same size. Which one? Prove it.

Exercise 3 (The missing women) (50 pts) For the sake of simplicity let us assume that the probability of being born a boy $P(B)$ is the same as the probability of being born a girl $P(G)$, namely $1/2$. Let us assume that the sex of different children are stochastically independent, and that there are no multiple births or adoptions.

In this case we would expect the population to be about half male and half female. But is it? According to Nobel Prize winning economist Amartya Sen, due to differential mortality, in Europe and North America there are about 105 females for every 100 males. But in other countries the ratio is considerably lower. The number of females per 100 males is 94 in China, 93 in India, and 92 in Pakistan, and 84 in Saudi Arabia (which has a large migrant male workforce). These latter countries are sometimes described as having “missing women.” Sen conjectures that this discrepancy is due to the neglect of female children, causing them to have higher mortality rates.

An alternative explanation might be that in some countries, parents prefer to have boys, so they may continue to have children until they have a boy or maybe two boys. Let’s see if this resolves Sen’s problem.

Let $B$ denote the number of boys in a family, and let $G$ denote the number of girls. The total number of children is then $N = B + G$. (These are random variables.) For each of the following parental decision rules, compute these expectations:

$$E G, E B, E N, E \left( \frac{G}{N} \right), E \left( \frac{B}{N} \right), E \left( \frac{G - B}{N} \right), \frac{E G}{E B}, E \left( \frac{G}{B} \right).$$

1. Parents have exactly one child.

2. Parents stop having children once they have a boy or two girls, whichever comes first.

3. Parents always have two children.
4. Parents have children until they have a boy. Take the idealization that the family has no limit on the number of children. (My own great-great-grandfather had 22 children that survived infancy. Not all of his three wives survived childbirth.)

(Note: I had to look up some of the infinite series involved here, but you may not need to. But if you do, you may want to consult the notes at http://www.math.caltech.edu/%7E2015-16/2term/ma003/Notes/FunSeries.pdf.)

- There is a problem with the way Sen chose to characterize the problem in terms of ratios. What is it? What is a better way to capture the intuition that the population should be about half men and half women?

- The analysis above ignored parents. Does this matter for Sen’s point? We also assumed that all the families observed were complete, but at any given point in time some of the families may not have stopped having children. Under the decision rules above, does this ameliorate the missing women problem?

Exercise 4 (10 pts) How much time did you spend on the previous exercises?

Exercise 5 (Optional Exercise 30 pts)
Alex tosses a fair coin $n$ independent times and Blair tosses a fair coin $m$ independent times. Find an elegant or clever argument to compute the probability that they have equal numbers of Tails. (I will be the judge of whether the argument is elegant, but it had better not involve any lengthy sums.)

References

