Biology 3550: Physical Principles in Biology
Fall Semester - 2016

Problem Set 2

Due by 11:59 PM on Monday, 26 September.

Please submit your work electronically via Canvas, following the guidelines below.

This problem set includes some relatively large calculations and graphing that you will probably not want to do by hand! There are a variety of computer programs that can be used for this sort of thing, and you are welcome to use whatever tools you like. Although I am not particularly fond of using spreadsheets, such as Microsoft Excel, for scientific computing and graphing, Excel is probably what most students are most familiar with, and it is fine to use it for this assignment.

Some other options are programs specifically designed for scientific graphing, including Kaleidagraph and SciDAVis, both of which are available in the Biology Department student computer lab and are briefly described on the Biol 3550 web page, at `http://courses.biology.utah.edu/goldenberg/biol3550/internet.shtml`. SciDAVis has the significant advantage of being an open source project that is free, but the documentation is quite limited.

Whatever method you use for the calculations and graphing, your numerical results, graphs and other illustrations should be incorporated into the pdf files you submit, and your calculations should be submitted as a separate file. Your graphs should be clearly labeled and easy to read. If you use Excel, please include all of your calculations in a single "workbook" file, with the calculations for different problems on different pages. If you use another program, please check in with the instructor so that we can figure out an appropriate way for you to submit your calculations.

If there is interest, I would be happy to organize a tutorial session on calculations and graphing, probably using SciDAVis.

1. Sports fans and gamblers are frequently obsessed with winning streaks. In sports and some forms of gambling (*e.g.* poker) many factors, including skill and determination, contribute to the likelihood of a winning streak, but in pure games of chance, it is just a matter of probabilities. Even in sports, probability plays a role, and most people probably underestimate the likelihood of a fairly long winning streak. Calculating these probabilities is actually rather involved and requires careful definition of the assumptions and enumeration of the possible outcomes. But, a relatively simple treatment can give some insight.

   (a) Consider a season of n games in which each game leads to a win for one team and a loss for the other (no ties!). Assume also that each team has an intrinsic probability of winning each of its games and this probability, p_{w}, is constant over the season. A winning streak of k games can occur in the following four ways:

      i. At the beginning of the season, ending with a loss in game $k+1$.

ii. At the end of the season, starting with game $n-k$ (following a loss in game $n-k-1$) and ending with a win in game n.

iii. A grand slam, running from the first to last game.

iv. Beginning after the first game and ending before the last game; starting with game j (following a loss in game $j-1$) and ending with a loss in game $j+k-1$.

As defined above, the four kinds of winning streaks have different probabilities, because they include 0, 1 or 2 losses, depending on the position in the season.

Write expressions for the probabilities for each of the four kinds of winning streaks (including the losses that bracket them) for a given sequence of k (type iii), $k+1$ (type i and ii) or $k+2$ (type iv) games.

(b) For a season of n games, we can calculate the expected number of winning streaks of length k. This expectation depends not only on the probability of a streak k long, but also on the number of opportunities for this number of continuous wins during the season. At the extremes, there is only one opportunity for a winning streak of n, but there are $n-1$ opportunities to start a winning streak of 2. The expected number of winning streaks of length k is the sum of the probabilities of all possible winning streaks of that length:

$$E(ws_k) = \sum_{j=1}^{n-k} p_{k,j}$$

where the index j represents the game number of the first win in the streak, and $p_{k,j}$ is the probability of a winning streak k games long and starting at game j.

For $k < n-1$, write an equation to calculate $E(ws_k)$ in terms of k, n and p_{w}, the probability of winning a single game.

(c) In the NBA, each team plays 82 games in a season. Use the equation derived in part (b) to calculate the expected number of winning streaks of 2 to 80 games, for a team that has a 50% chance of winning each game. Make a plot of $E(ws_k)$ versus k for this team.

(d) The longest winning streak in NBA history was 33, by the Los Angeles Lakers in the 1971–72 season. That season, the Lakers won 69 of their 82 games, or 84%. Assuming that this win ratio represents the intrinsic probability of winning each game, use your equation to to calculate the expected number of winning streaks of 2 to 80 games, and make a plot of $E(ws_k)$ versus k.

How likely do you think it is for this team to have a winning streak of 33 games in one season?

What do you conclude from all of this about the 71-72 Lakers? In particular, do the overall win-loss record and the winning streak seem to be consistent with one another? Why or why not? If there seems to be an inconsistency, how might it be explained?

(e) What do you think is the weakest assumption in these calculations? Explain your reasoning.

2. In class, we have focused attention on the six-row plinko, but the situation looks a bit different when there is an odd number of rows. So, this problem considers a seven-row plinko.

   (a) To make sure that you have visualized the situation clearly, make a drawing of the plinko, like the ones we used in class, with the buckets numbered 0 to 7 from the left.

       What is the total number of paths through the plinko?

   (b) For each bucket,a calculate the number of paths to the bucket. Make a bar graph showing the number of paths versus bucket number. Each bar should represent one bucket, labeled 0 to 7.

   (c) Assuming that there is nothing funny about the plinko (that is that a ball is equally likely to turn right as left at each peg), calculate the probability of a single ball falling into each of the buckets, and make a bar graph representing the distribution.

       In what general way does this distribution of probabilities differ from the one that we calculated for the six-row plinko.

   (d) Suppose that all of the pegs in the seven-row plinko has been tampered with so that the probability of turning right at each peg is 0.7, instead of 0.5. Recalculate the probabilities of a single ball falling into each of the buckets and make a bar graph representing the distribution.

3. Now, someone has done something very devious to the seven-row plinko. This sneak has modified only the bottom row of pegs, so that a ball falling on any peg in this row has a probability of 0.3 of turning right and a probability of 0.7 of turning left. All of the other pegs are normal. Follow the steps below to calculate the probabilities for this modified plinko.

   (a) For all but buckets 0 and 7, the paths leading to a given bucket can be divided into to classes: The path ends with either a right turn from the peg above and to the left of the bucket or a left turn from the peg above and to the right[1]. Buckets 0 and 7 each have a single path, as usual.

       Using the general symbols, p_{l} and p_{r} for the probabilities of a left or right turn at row 7, respectively, write expressions for the probabilities for:

       - The single path to bucket 0.
       - A single path of the type that leads to bucket k, for $1 \leq k \leq 6$, and ends with a left turn at row 7.
       - A single path of the type that leads to bucket k, for $1 \leq k \leq 6$, and ends with a right turn at row 7.
       - The single path to bucket 7.

   (b) For bucket 1, calculate the number of paths that end with a left turn and the number that end with a right turn. Write an expression for the total probability of a ball falling in bucket 1.

[1]The buckets should not be confused with baskets: None of the buckets or paths should be considered "deplorable."

(c) For bucket 2, calculate the number of paths that end with a left turn and the number that end with a right turn. Write an expression for the total probability of a ball falling in bucket 2.

(d) Continue for buckets 3 to 6. For each, calculate the number of paths that end with a left turn and the number that end with a right turn, and write an expression for the total probability of a ball falling in the bucket. You should see patterns that will enable you to do this without explicitly counting the paths.

(e) Using the expressions derived in the previous steps, check that the sum of probabilities for all 8 buckets is 1. If it isn't, something has gone wrong.

(f) For the specific case where $p_l = 0.7$, calculate all of the probabilities and make a bar graph of probability versus bucket number.

If you can do this problem, pat yourself on the back and consider yourself a plinko master!

4. An ant is taking a two-dimensional random walk on a flat surface. We will distinguish steps that the ant takes with its feet from the random walk (RW) steps, of length δ, which are made up of lots of footsteps. In a RW step, the ant choose a direction randomly and walks distance δ in that direction. She then choose another direction randomly and walks distance δ in that direction. She repeats this for a total of n RW steps. The surface conveniently has x- and y-coordinates drawn on it, so that we can monitor the ant's progress along each direction, as well as the distance from the starting point.

   The ant walks at a rate of 5 cm/min and normally walks for 30 s in each of its random walk steps.

   (a) If the ant walks for 30 min, without stops to rest and always in segments of 30 s, calculate the following:
      - The total distance walked (as measured by a pedometer on one of its six legs).
      - The number of RW steps, n in a single walk.
      - The distance covered, δ, in a single RW steps.

   (b) At the end of each n RW steps, the ant takes a rest for a few minutes and then starts off on another random walk of n steps. Assuming that the ant completes a large number, N, of random walks, all the same, calculate the following:
      - The average (RMS) distance that the ant moves from the starting point of each walk.
      - The average (RMS) distance that the ant moves from the starting point of each walk in the x-direction.
      - The expected total RMS distance from the original starting point after N random walks of n steps each. (This question assumes that the ant would be up to repeating all of the N random walks a large number of times, and the distance would be measured from the start of each set of N walks to the end of each set.)

   (c) It's not clear why the ant is doing all of this walking, but she seems to have become dissatisfied, and has changed her routine. Now, she walks in straight segments for only 15 s (at the same speed), but maintains the pattern of walking for 30 min between rests, thus breaking her travels up into distinct random walks. For this new routine, calculate the following.

- The number of RW steps, n, in a single walk.
- The distance covered, δ, in a single RW step.
- The average (RMS) distance that the ant moves from the starting point of each walk.
- The average (RMS) distance that the ant moves from the starting point of each walk in the x-direction.
- The expected total RMS distance from the original starting point after N random walks of n steps each.

(d) Describe how the pattern of random walks observed in the two scenarios would differer qualitatively. Suggest a (serious) reason that an ant might choose longer or shorter steps in a random walk.