

# Simple Random Walk



Two equally matched opponents are competing in a game in which changes in score occur often and in one point increments. (Imagine a basketball game in which every basket counts only one point.) We'll use simulation to investigate the following questions.

- Which is more likely: that one team leads for most of the game, or that the lead tends to change frequently over the course of the game?
- When would you expect the largest lead (or deficit) to occur — near the beginning, the end, or in the middle of the game? (If the largest lead (or deficit) is attained at several points in the game, when you do expect it to first occur?)
- When would you expect the last tie to occur — near the beginning, the end, or in the middle of the game?

*Pause to think about these questions before proceeding.*

Let  $X_n$  be the point difference (team A - team B) after the first  $n$  scores;  $X_n > 0$  if team A is in the lead,  $X_n < 0$  if team B is in the lead, and  $X_n = 0$  if the score is tied. We can model  $X_n$  as a simple symmetric random walk on the integers starting from  $X_0 = 0$ , like in the Harry and Tom example from the handout.

Consider the first  $2n$  steps of the walk. We say  $2n$  because that will be even. We are interested, in particular, in times at which the walk can be in state 0 (that is, the score is tied), which can only happen in an even number of steps.

We are interested in the following random variables, each of which has been scaled to take values between 0 and 1.

- $T/(2n)$ , where  $T$  is the time, between 0 and  $2n$ , at which the walk is last in state 0. For example, if  $n = 50$  and the walk takes 100 steps, and it is last at 0 at step 84, then  $T/(2n) = 0.84$  (the last tie occurs with 16% of the game remaining.)
- $L/(2n)$ , the fraction of time the walk stays above 0, where  $L$  the number of time steps for which the walk is above 0;  $L/(2n)$  is the proportion of the game team A is in the lead.
- $M/(2n)$ , where  $M$  is the first time, between 0 and  $2n$ , at which the maximum value, over time 0 to  $2n$ , of the walk is attained;  $M/(2n)$  is the first time (measured as a proportion of the full game) at which team A's largest lead is attained.

**Write your own code to conduct and run a simulation to approximate the distribution of each of  $T/(2n)$ ,  $L/(2n)$ , and  $M/(2n)$  for  $n = 100$ . Summarize the results with appropriate plots and summary statistics, and describe the distributions. Consider the three questions at the start of this page; what do your simulation results suggest? Write a brief report summarizing your results and conclusions.**

*Optional:* experiment with different values of  $n$  and discuss how the results change with  $n$ .