Investigation 4: Sports series?
(assigned on Fri Oct 21, due on Fri Oct 28)

You may work with in a group of as many as three students on this assignment, handing in one report with all names, provided that you all contribute to the work. Your report must be word-processed, with computer output integrated into the report.

Suppose that two teams (for fun let’s call them the Domestic Shorthairs and Cache Cows) play a series of games to determine a winner. In a best-of-three series, the games end as soon as one team has won two games. In a best-of-five series, the games end as soon as one team has won three games, and so on. Assume that the Domestic Shorthair’s probability of winning any one game is $p$, where $0.5 < p < 1$. (Notice that this means that the Domestic Shorthairs are the better team.) Also assume that the outcomes are independent from game to game.

We will compare different series configurations/rules on two criteria: the probability that the better team wins the series, and the expected value of the number of games needed to complete the series.

Exact Analyses: Best-of-Three

a) Determine (exactly) the probability that the Domestic Shorthairs win a best-of-three series, as a function of $p$. (Show your work.)

b) Graph this function for values of $0.5 < p < 1$ (include good axis labels), and comment on its behavior.

c) Let the random variable $X_3 =$ number of games played in a best-of-three series. Determine the probability distribution (pmf) of $X_3$, as a function of $p$. (Show your work.)

d) Determine $E(X_3)$, as a function of $p$. Graph this function for values of $0.5 < p < 1$ (include good axis labels), and comment on its behavior.

Exact Analyses: Best-of-Five

e) Repeat (a) and (b) for a best-of-five series.

f) Produce a graph to compare the probability that the better team wins the series between a 1-game, 3-game, and 5-game series. [R hint: You can put multiple graphs on the same axes with commands such as:

```r
> plot(p,winprob1,type="l",col="blue",xlab="gameprob",ylab="seriesprob")
> lines(p,winprob2,col="green")
> lines(p,winprob3,col="red")
```

g) Comment on what the graph from f) reveals.
h) Repeat (c) and (d) for a best-of-five series. Call the random variable $X_5$ in this case.

i) Produce a graph to compare the expected number of games between a 1-game, 3-game, and 5-game series. Comment on what the graph reveals.

**Simulation Analyses**

The code in the file `sportseriesbestofn.R` conducts a simulation analysis of this random process, where the user enters the values of $p$ (probability that better team wins an individual game), $n$ (maximum number of games in the series), and $N$ (number of repetitions).

j) Explain why this program requires a **while** loop rather than a **for** loop.

k) Describe the condition on the **while** loop. (In other words, the loop continues to run while what is true?)

l) Run the code for a best-of-five series with 100,000 repetitions with $p = .6$. Report the approximate probability that the better team wins and the average number of games played. How do these compare to the theoretical values that you calculated earlier?

m) Run the code for a best-of-seven series with 100,000 repetitions with $p = .6$. Report the approximate probability that the better team wins and the average number of games played. Then repeat this for a best-of-9 series, a best-of-15 series, a best-of-25 series, a best-of-49 series, and a best-of-99 series. Produce a table that reports the approximate probabilities for all of these series lengths, and produce another table that reports the average numbers of games played.

n) Describe how the approximate probabilities (that the better team wins the series) change as the maximum number of games increases.

o) Describe how the average numbers of games change as the maximum number of games increases.