Stat 450

R Assignment 2

Fall 2016

R code of solutions to be submitted via D2L dropbox by 5pm on Friday, September 23. You do not need to submit a .Rmd file; just the R code. See the HW 1 key for an example of a suitable submission.

In the Major League Baseball World Series, two teams play a best-of-seven series. The first team to win four of the seven games is the champion. Suppose Team A and Team B play a best-of-seven championship. Assume that the probability Team A wins any game against Team B is p, and assume the games are independent of each other. Of interest is the distribution of Y, defined to be the *number of games played in the series*. Note that $p(y) > 0$ for $y \in \{4, 5, 6, 7\}$ only.

Write a function that includes a `while()` loop that can be used to generate random realizations of Y, i.e., many simulations of a World Series between Team A and Team B assuming probability p that Team A wins any one game. Getting you started:

```
one.world.series <- function(p)
{
  number.A.wins <- 0
  number.B.wins <- 0
  series.over <- ifelse(max(number.A.wins,number.B.wins)==4, 1, 0)
  total.games <- number.A.wins + number.B.wins
  while(series.over == 0 ) {
    #Simulate a single game, then update number.A.wins, number.B.wins,
    #total.wins, series.over, and ngames, as appropriate.
  }
  return(ngames)
}
```

1. (5pts) Obtain 1000 simulated world series for $p \in \{1/4, 1/3, 1/2, 2/3, 3/4\}$.
2. (4pts) What is $E(Y)$ for each of the p?
3. (5pts) Now write a function that takes as argument `p` and returns the *theoretical* (i.e., **not simulated**) pmf as a length-4 vector. Verify that you have correctly defined the pmf by verifying that $\sum_{y=4}^{7} p(y) = 1$ for various values of p. (Hint: for each $y \in \{4, 5, 6, 7\}$, $p(y)$ can be calculated with `dnbinom()`.)
4. (6pts) Write another function that takes as argument `p` and returns $E(Y)$. You may find it helpful to call your function from Question 3 in this function. Use the output from your newest function to plot $E(Y)$ as a function of p when p is a length-100 vector in $\{0, 1\}$. For what value of p (approximately) does $E(Y)$ appear to be maximized, and why does this make sense?