Homework 5

Due on Gradescope 10/25/2016 at 4:00PM (Before Lecture)

The first two problems should look familiar!

1. Deriving Parameters, to appear in TMD: The goal of this problem is for you to use the definitions and notation from class in familiar contexts. For each random variable, derive the expectation $E(X)$, the variance $\mathrm{Var}(X)$, and the standard deviation $SD(X)$ from their definitions (i.e., work out the sum).
   (a) X is the number of spots that show on one roll of a fair six-sided die.
   (b) X is an "indicator" random variable; it has the value 1 with probability p, and the value 0 with probability 1 p. This random variable is a Boolean, that is, it can only be 0 or 1. Just as 0s and 1s are powerful in computing, so also indicators are powerful in probability theory. Youll see how next week.
   (c) X is the number of heads in one toss of a fair coin.
   (d) X is the number of heads in two tosses of a fair coin.
   (e) X is the number of red cards among two cards picked at random without replacement from a standard deck (52 cards of which 26 are red).

2. Stock Price: (**Note:** In this problem, you may use these facts presented in class: If X is a binomial random variable with size n and probability p, $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.)

   Suppose that TechCo is one of San Francisco's hottest publicly traded tech startups, and its stock price moves in the following way: every day, it either increases by \$1 with probability p or decreases by \$1 with probability $1-p$, and the change on each day is independent. Let Z be the change in the price of TechCo's stock over two weeks; that is, Z is the price of TechCo's stock on Oct 19 minus the price of TechCo's stock today, Oct 5.
   (a) What is $E(Z)$?
   (b) What is $SD(Z)$?
   (c) Suppose that p is 0.51, so the stock price has a very slight upward drift. How many days would it take for $E(Z)$ to be more than 2 standard deviations ($SD(X)$) away from 0?

3. Polling Margin of Error: (**Note:** In this problem, you may use these facts presented in class: If X is a binomial random variable with size n and probability p, $E(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.)

   When pollsters conduct surveys of political opinions, they draw a sample of size n without replacement from a population of size N, where N is much larger than n. Nonetheless, they often model the number of people supporting a particular candidate (for the purposes of this quesion, let's say Clinton), X, as a binomial random variable with size n and probability p. In this case, the pollster does not know the true value p, but we can write down expectations and variances of important quantities in terms of p.

(a) Technically, what are the requirements for a random variable X to be binomially distributed? Which of these requirements are violated in this case? Intuitively, why is the binomial model approximately correct here? (Hint: Sizes of N and n matter).

(b) Define Y as the *proportion* of respondents who say they will vote for Clinton, $\frac{X}{n}$. What is $E(Y)$?

(c) Using the same definition of Y as above, what is $\text{Var}(Y)$?

(d) For what value of p is $\text{Var}(Y)$ maximized? Use calculus or an algebraic argument.

(e) Simple polls will report Y as their estimate of the unknown proportion of Clinton supporters p. They will also report a "margin of error", which is 2 times the worst case standard deviation of Y (that is, the largest SD among all potential parameters p). For a poll with n respondents, what is the margin of error?