STAT 325 – Handout 10
Binomial Random Variables (3.1)

For the next couple of days we will examine five special cases of discrete random variables that arise often enough to be widely studied and to have their own names.

Example 10-1: Exam guessing
Suppose that you take a multiple choice quiz for which you guess randomly among the options on every question (well, not really you, of course, I mean some clueless student instead). Suppose there are 5 questions, with 3 options to choose (guess) from on each question. Let the random variable X be the number of questions that you answer correctly.

a) What are the possible values of X?

b) Determine the probability that you answer all 5 questions correctly. Also indicate what rule you use to do this calculation.

c) Determine the probability that you answer 0 questions correctly.

d) Determine the probability that you answer exactly 2 questions correctly. [*Hint*: First find the probability of one particular sequence with exactly 2 correct. Then multiply by the number of such sequences.]

e) Determine a general expression for the probability that you answer k questions correctly.

f) Now generalize to n questions with probability p of answering any one question correctly. Determine a general expression for the probability of answering k questions correctly in this case.

- A random variable X is said to have a **binomial** distribution with parameters *n* and *p* if:
  - X can be thought of as the number of successes in *n* trials
  - Each trial has two possible outcomes (S, F)
  - The trials are independent
  - The probability of success is the same (*p*) on every trial
- Notation: X ~ Bin(*n*, *p*)
- The pmf for a binomial random variable X is: $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$ for $x = 0, 1, \ldots, n$.
- It can be shown that E(X) = *np* and Var(X) = *np*(1-*p*).
- Calculations:
  - Minitab: `Graph> Probability distribution plot> Binomial`
  - R: pmf: `dbinom(x, n, p)`; cdf: `pbinom(x, n, p)`, inverse cdf: `qbinom(prob, n, p)`, simulation: `rbinom(numreps, n, p)`

g) Suppose that you must answer at least half of the questions correctly in order to pass the quiz. Determine the probability of passing, first directly from the pmf and then using Minitab and then using R with the cdf.

h) Would you expect this probability of passing to be larger, the same, or smaller, if the quiz consists of 13 questions rather than 5? Explain your reasoning.

i) Determine the probability of passing the quiz if it consists of 13 questions rather than 5. Is your intuition from h) supported?

j) How would you expect these probabilities to change if each question has 5 options rather than 3? Explain your reasoning.

k) Make this change (from 3 options per question to 5), and recalculate the probabilities. How do these compare to the previous ones?

Example 10-2: Naughty or nice?

We all recognize the difference between naughty and nice, right? What about children less than a year old- do they recognize the difference and show a preference for nice over naughty? In a study reported in the November 2007 issue of *Nature*, researchers investigated whether infants take into account an individual's actions towards others in evaluating that individual as appealing or aversive, perhaps laying for the foundation for social interaction (Hamlin, Wynn, and Bloom, 2007). In one component of the study, 10-month-old infants were shown a "climber" character (a piece of wood with "google" eyes glued onto it) that could not make it up a hill in two tries. Then they were alternately shown two scenarios for the climber's next try, one where the climber was pushed to the top of the hill by another character ("helper") and one where the climber was pushed back down the hill by another character ("hinderer"). The infant was alternately shown these two scenarios several times. Then the child was presented with both pieces of wood (the helper and the hinderer) and asked to pick one to play with. The researchers found that the 14 of the 16 infants chose the helper over the hinderer. (You can see the videos at: http://www.yale.edu/infantlab/socialevaluation/Helper-Hinderer.html.)

a) If the infants really had no preference between the toys and so were just choosing a toy at random, what is the probability that 14 or more of the 16 infants would have chosen the helper toy? Explain how your answer is based on the binomial distribution.

b) Is this probability small enough to cast strong doubt on the "no preference" assumption, which would suggest that infants really do sow a preference for the helper toy? Explain.

c) How many of the 16 infants would have to choose the helper toy, in order for the probability to be no more than .05 of obtaining that many or more choosing the helper toy under the "no preference" assumption?

Example 10-3: Random babies (cont.)
Consider yet again the "random babies" process with 4 mothers/babies. Suppose that you simulate this process 10 times. Let the random variable X be the number of times that you obtain 0 matches.

a) What probability distribution does X have? (Specify its parameter values as well as its name.)

b) Determine the expected value, variance, and standard deviation of X.

c) Determine the most likely value of X and its probability.

d) Determine the probability that you obtain 0 matches more than half the time in these 10 repetitions.

e) Determine the probability that you obtain 0 matches no more than twice.

f) Change the number of repetitions from 10 to 10,000 and consider the random variable Y = number of repetitions that produce 4 matches. Specify the probability distribution of Y, and also determine its expected value, variance, and standard deviation.

g) Determine the probability that Y falls within one standard deviation of its expected value.

h) Determine the probability that Y falls within two standard deviations of its expected value.