Assignment #5 — Due Wednesday, October 7, 2009, by 5:00 P.M.

Turn in homework in lecture, discussion, or your TA's mailbox (just inside the main entrance of MSC, 1300 University Avenue) Please circle the discussion section you expect to attend to pick up this assignment.

311: Monday 1:20–2:10 **312:** Monday 12:05–12:55 **313:** Tuesday 8:25–9:15

The following problems all involve discrete probability distributions. You may need to think a bit to identify the random variables, their distributions, and associated parameters in each problem.

1. Single strands of DNA sequences can be modeled as long words from a four-letter alphabet, A, C, G, and T. A simple (and unrealistic) model is that all such sequences of a given length are equally likely. This is equivalent to a model in which each DNA base in a sequence is uniformly distributed over the four possible letters. Assume this simple model for these problems:
   (a) A DNA sequence contains 120 bases. Calculate the probability that exactly 25 of these bases are As.
   (b) Some DNA sequences can be divided into *codons* which are three consecutive bases, so there are 64 possible codons ranging from AAA to TTT. In reading DNA in a particular phase, codons do not overlap, and so 120 bases could be read as 40 codons. What is the probability that two or more codons out of 40 have the pattern ACT?

2. Five people play a game where they take turns rolling a pair of regular, fair, 6-sided dice in order: A, B, C, D, E, A, B, C, D, E, ... The game continues until someone rolls double ones (so it is possible that A can win immediately before anyone else rolls the dice).
   (a) Find the probability that everyone rolls the dice at least twice.
   (b) Find the probability that each person wins the game.

3. Suppose that X_1, X_2, and X_3 are all Poisson(θ) random variables for some unknown θ.
   (a) Find the value of θ that makes the probability of $P(X_1 = 2) \times P(X_2 = 3) \times P(X_3 = 5)$ as large as possible.
   (b) Evaluate this probability.

4. One ball is drawn at random from a bucket with 3 red balls and 7 white balls. A second ball is drawn at random from a bucket with 4 red balls and 6 white balls. Let X_i be an indicator random variable that the ith ball is red ($i = 1, 2$) and let Y be the total number of red balls drawn.
   (a) Find the probability distribution of Y.
   (b) Show that Y does not have a binomial distribution.

5. Suppose that $X_i \sim \text{Bernoulli}(\theta_i)$ for $i = 1, 2$ and that $P((X_1 = x_1) \cap (X_2 = x_2)) = P(X_1 = x_1) P(X_2 = x_2)$ for all x_1 and x_2. Let $Y = X_1 + X_2$.
   (a) Find the probability distribution of Y in terms of θ_1 and θ_2.
   (b) Show that if $\theta_1 \neq \theta_2$, then there does not exist a number θ such that $Y \sim \text{Binomial}(2, \theta)$. *(Hint: Assume that such a θ does exist. This will imply three equations relating θ with θ_1 and θ_2. Show that the only valid solution requires $(\theta_1 - \theta_2)^2 = 0$ which contradicts the assumption that $\theta_1 \neq \theta_2$.)*

Work to do, but not turn in.

- Read Chapter 2, remainder.