Homework 7

Due date: Homework is due on Wednesday, November 26 at 10 a.m. Homework should be turned into the box for your section (outside of Room 253 Sloan).

1. Moore and McCabe, Chapter 8, Problem 8.48.

2. It has been suggested that dying people may be able to postpone their death until after an important occasion, such as a wedding or a birthday. Philips and King (1988) studied the patterns of death surrounding Passover, an important Jewish holiday, in California during the years 1966-1984. They compared the number of deaths during the week before Passover to the number of deaths during the week after Passover for 1919 people who had Jewish surnames. Of these, 922 occurred in the week before and 997 in the week after Passover. The significance of this discrepancy can be assessed by statistical calculations. We can think of the counts as follows: if there is no holiday effect, then a death has probability 1/2 of occurring in the week before the holiday and probability 1/2 of occurring after.

   (a) Propose a test and present your conclusions.

   (b) Repeat your analysis for a group of male Chinese and Japanese ancestry, of whom 418 died in the week before Passover and 434 died in the week after.

   (c) What is the relevance of this latter analysis to the former?

3. Suppose that the $X \sim$ binomial(n, p). Show that the maximum likelihood estimate of p is X/n. If $n = 10$ and $X = 5$, plot the log-likelihood function.

4. In an effort to determine the size of an animal population, 100 animals are captured and tagged. Some time later, another 50 animals are captured and it is found that 20 of them are tagged. How would you estimate the population size? Justify your answer. What assumptions about the capture/recapture process do you make? (*Hint:* you will assume there are N animals of which 100 are tagged. You then calculate the likelihood of recapturing 20 of them.)

5. For two factors – starchy or sugary and green or white base leaf – the following counts for the progeny of self-fertilized plants were observed:

Type	Count
Starchy and green	1997
Starchy and white	906
Sugary and green	904
Sugary and white	32
	n = 3839

   According to genetic theory, the cell probabilities are $(2 + \theta)/4, (1 - \theta)/4, (1 - \theta)/4$, and $\theta/4$ where $0 < \theta < 1$ is a parameter related to the genetic linkage of factors.

   (a) Find the maximum likelihood estimate for θ.

   (b) Construct an approximate 95% confidence interval for θ.

(c) Let X_1 denote the count in the first cell (starchy and green) and X_4 the count in the last cell (sugary and white). Show that $4X_1/n - 2$ and $4X_4/n$ are both unbiased estimates of θ. Write down the variances of these unbiased estimates and compare them with the asymptotic variance of the maximum likelihood estimate.