Biostatistical Methods I (BIOS 5710)
Breheny

Assignment 4
Due: Wednesday, October 1

1. Suppose we were interested in testing which of two diets was more effective at reducing an individual's cholesterol level. Consider two potential study designs:

   Design A: Fourteen individuals with high cholesterol are placed on either Diet 1 or Diet 2 (randomly); after two weeks, their LDL cholesterol levels were measured. Each individual then switched to the other diet; after two weeks, the LDL levels were recorded again. The difference between the two measurements was then analyzed.

   Design B: Fourteen individuals with high cholesterol were placed on either Diet 1 or Diet 2 (randomly, with 7 individuals assigned to each diet). After two weeks, their LDL cholesterol levels were measured. The difference between the average cholesterol levels in each group was then analyzed.

   Which design will be able to more accurately estimate the difference between the two diets? Why?

2. Describe two situations in which a crossover design would be an unrealistic and/or impractical way to test which of two treatments was better.

3. In a 2006 study published in *The New England Journal of Medicine*, 78 pairs of patients with Parkinson's disease were randomly assigned to receive treatment (which consisted of deep-brain stimulation of a region of the brain affected by the disease) or control (which consisted of taking a prescription drug). The researchers found that in 50 of 78 pairs, the patients who received deep-brain stimulation had improved more than their partner in the control group. The parameter of interest is θ, the probability of doing better on treatment than control.

   (a) Let X denote the number of pairs in which the treatment patient did better than the control patient. What does the null hypothesis hypothesize about the distribution of X. Be specific.

   (b) Test the null hypothesis in (a).

   (c) Construct a 95% confidence interval for θ.

   (d) In the paper, the authors claim that deep-brain stimulation is "more effective than medical management." Based on your answers to (b) and (c), do you agree? How strong is the evidence to support this conclusion?

   (e) Suppose you analyzed this data from a Bayesian perspective, using a uniform prior on θ. What is the posterior distribution of $\theta|x$?

   (f) Plot the posterior density $f(\theta|x)$.

   (g) Calculate a 95% posterior interval for θ.

   (h) What is the posterior probability that $\theta > 0.5$?

   (i) Based on your analysis in (e)-(h), do you agree that deep-brain stimulation is "more effective than medical management." How strong is the evidence to support this conclusion?

(j)-(m) Repeat (e)-(h), only this time assuming a Beta$(11, 11)$ prior.

4. Your friend Andy claims to have ESP (extra-sensory perception). To test this claim, you propose the following experiment: You select one of four cards and Andy tries to identify it using his ESP. Let θ denote the probability that Andy is correct in identifying the card. Suppose that the experiment is repeated ten times and Andy is correct six times and incorrect four times.

   (a) What would the null hypothesis be in this situation?

   (b) Perform a test of the null hypothesis from part (a).

   (c) Suppose that, before carrying out the experiment, you carefully considered your prior beliefs about Andy's ESP ability and came up with the following:

θ	0	.125	.250	.375	.500	.625	.750	.875	1
$p(\theta)$	.001	.001	.950	.008	.008	.008	.008	.008	.008

   Find the posterior probabilities of these values of p. In particular, what is your posterior probability that Andy has no ability?

   (d) Do you reach similar conclusions from your analyses in (b) and (c)? Comment on why or why not.

5. Consider the posterior mode – i.e., the value of θ that maximizes $f(\theta)$ – of a Beta(α, β) distribution.

   (a) Show that the posterior mode is equal to $(\alpha - 1)/(\alpha + \beta - 2)$.

   (b) Comment on the value of the posterior mode in the special case of the uniform prior.

6. Extend the `binom.bayes` function we wrote in class to also calculate the highest posterior density (HPD) interval. For the purposes of this assignment, you need only consider the "vanilla" case where there is a single mode somewhere between 0 and 1 (i.e., your function does not have to work in cases where the mode occurs at one of the boundaries, or when multiple modes are present). Please turn in your function electronically via the Dropbox on ICON. Hint: You may wish to define a `finv` function (or whatever you want to call it) that calculates the inverse pdf of a beta distribution.