Homework 3
Bayesian Data Analysis

Instructions

- Do Problems 1 and 4.
- Choose between Problems 2 and 3 (PhD in Stats should really try to do problem 3 though 😜)

Problem 1

Exercise 5.6 in the textbook

Problem 2 - A Meta-Analysis of Parapsychology Data

The term extrasensory perception (ESP) refers to the ability to acquire information without using the five known senses. Psychokinesis (PK) refers to the ability to change the physical world without apparent physical force. A general term used to describe both PK and ESP abilities is *psi*, from the letter that begins the Greek word for *"psyche"*. Some of the work investigating ESP uses the "ganzfeld" technique. Ganzfeld is German for "whole field". You can read more about the procedure here: `https://psi-encyclopedia.spr.ac.uk/articles/ganzfeld`.

Here, we consider a meta-analysis of 56 ganzfeld studies, that met criteria for methodological rigor and adherence to standard ganzfeld procedures. We thank Jessica Utts and Wes Johnson (professors Emeriti at UCI) for the data.

```
ESP.data=read.csv("./GanzStudiesUsed-56.csv", header=T)
```

```
head(ESP.data)
```

```
##    n hits
## 1 32   14
## 2  7    6
## 3 30   13
## 4 30    7
## 5 20    2
## 6 10    9
```

The data report on the number of "hits" (number of correct targets chosen by the "receiver" in the Ganzfeld experiment) out of n trials in each study. We observe that if psychic abilities are real, the assumption of a constant probability of "hits" among studies may not be valid. For instance, it is well-known in psychology that there can be an experimenter effect, in which some experimenters are able to put participants more at ease and thus obtain more favorable results than others. Another reason for differences may be that the participant pool differs. For instance for two of the studies in Table 1 (#24 and #34) the participants were creative artists, dancers and musicians, and those studies obtained very successful results. In our Bayesian analysis we want to allow for the possibility of different true probabilities of success for each study.

1. Propose a model (in formulas!) to conduct a meta-analysis study of the parapsychology data above

2. Write the previous model in a Jags/Stan script.

3. **Choice of Priors** We want to consider three different priors for this analysis:

   - **Open-minded prior**, which uses 0.25 as best estimate for the mode probability of a hit, and for which we are 95% sure that the probability is below 0.30.
   - **Psi believer's prior**, which uses 0.33 as the best estimate for the mode probability of a hit, for which we are 95% sure that the probability is less than 0.36
   - **Psi skeptics prior**, which uses 0.25 as the best estimate for the mode probability and for which we are 95% sure that theprobability is less than 0.255

   Using Betabuster and the model you have proposed above, propose prior distributions for the probability of a hit corresponding to the three priors (scenarios) above.

4. Obtain the posterior mean and 95% posterior credible intervals for the probability of a hit when using each of the previous priors.

5. Plot the prior and posterior distributions for each of the three scenarios above.

6. Comment on the results. What can you say about the posterior beliefs of the open-minded, skeptic and the believer investigators? How have they changed with the data?

Problem 3

Phase IIB trials are multi-arm studies to compare the efficacy of new drugs, with the goal of screening out those that are ineffective. The multiple arms could be different treatments (possibly including a control arm, such as "standard of care"), different doses or schedules of the same agent, or any combination of such comparisons. Typycally, Phase IIB trials are randomized, that is patients are assigned to any of the arms following a randomization scheme. Here, we do not consider the different randomization schemes. Instead, we are interested to assess how a drug compares with another, after the patients have already been

assigned to each arm, and the outcome is recorded.

We consider a multi-institution trial of two drug compounds: gemcitabine + docetaxel (G+D) versus gemcitabine alone (G, standard of care) for patients with advanced/metastatic unresectable soft tissue sarcoma. For more details, see Maki et al, A SARC multicenter phase III study of gemcitabine vs. gemcitabine and docetaxel in patients with metastatic soft tissue sarcomas. *J Clinical Oncology*. 2006;24(18S):9514; and Thall & Wathen (2007), Practical Bayesian Adaptive Randomization in Clinical Trials, *Eur J. Cancer*, for the description of the precise Bayesian methodology used in the trial (which we only slightly simplify here).

The primary end point of the study was tumor response R, defined as complete or partial response within 24 weeks, or stable disease lasting at least 24 weeks. The investigators were interested in the probabilities of overall treatment success, θ_R in each of the two arms of the study. One hundred nineteen patients had assessable outcomes at the end of the study. The adaptive randomization assigned 73 patients to the G+D arm and 49 patients to the G arm (again, we do not bother here about the randomization scheme, but only look at the sample size collected in each cohort). The primary end point was reached by 13 patients receiving G alone and 23 patients receiving G+D.

1. Propose a model to analyze these data. You want to make sure that the treatment effects are defined specifically for each arm, but there may still be some sharing due to the presence of G in both arms. For the prior specification, you may want to consider the following information: based on the current standard of care, the mode response rate should be around around 25%, and less than 0.7 with high probability (say, 95%) for both treatments. In addition, we want to base our judgement on the trial data only, so the information provided by the available prior data (prior sample size) could be considered as a positive random variable roughly centered around 1 and large variance (say, 3).

2. Write the previous model in Rjags (or rstan).

3. Find the posterior distribution of the response rates in each of the two arms, say θ^{G+D} and θ^{G}. Summarize the posterior distribution by using the posterior means, and 90% posterior credible intervals.

4. Based on the data collected in the two arms, would you recommend the trial for a larger Phase III study? What would you base your decision upon?

5. It is often the case that the decision to recommend a trial for a larger Phase III study is based on a simulation where one would try to assess how the trial would perform in the larger future study. Suppose that a larger study would enroll 500 total new patients, equally assigned to each of the two arms. Summarize the results of the simulation of the larger study, by means of the mean and 95% credible intervals of the posterior predictive distributions for both arms.

Multi-arm Adaptive Phase trials

In recent year, Bayesian multi-arm adaptive phase trials have received attention from the news. For an example, see

`http://www.nytimes.com/2013/07/14/opinion/sunday/do-clinical-trials-work.html`

The trials seek to quickly identify effective drugs and combinations for specific subtypes of a disease. In brief, such trials are conducted in multiple stages. At the end of each stage, an evaluation is made for each arm of the trial (treatment group):

a) the arm is dropped and the treatment is discontinued if there's strong evidence that there is little hope for success in a future trial

b) the arm is graduated and the treatment is recommended for further study in a larger Phase III study if the results obtained so far are very promising, meaning that the treatment have good chances to be successful in a future Phase III study

c) the arm is continued and the treatment keeps being administered and evaluated in the next stages of the current Phase II trial, if its chances in a future study are yet inconclusive

1. In the multi-arm trial FOCUS (Seymour et al, 2007, *The Lancet*, `https://www.ncbi.nlm.nih.gov/pubmed/17630037`) trial patients were randomised to five treatment plans A, B, C, D and E.

   Consider a (simple) Bayesian multi-arm adaptive trial, where treatments are investigated for 4 specific subgroups of patients, e.g. the 4 groups are defined based on some specific genetic characteristics of the patients. For now, assume only a treatment is tested for each subgroup of patients. Also, assume that patients can be assigned to the standard of care (control) treatment, irrespectively of the specific subgroup they belong to (i.e. you have a total of 5 trial arms). Write a hierarchical model (in formulas, i.e. not with rjags or rstan) to analyze the data from this trial.

2. How would you change your model if you had two treatments for each subgroup of patients (i.e. a total of 8 treatments across 4 specific subgroups + standard of care=9 arms)

3. Let's consider again the case of a single treatment in each subgroup of patients (for a total of 5 arms). Let's suppose that we are at the end of one of stages of the trials, and you need to set up a decision rule for either (a) stopping, (b) graduating or (c) continuining a treatment arm. Provide your best attempt at a decision rule for such purpose, based on the information above. *[Of course, there's no set answer. However, I am interested to see if you understand the main parts of the problem, and how to set up a decision rule like a Bayesian might do. You should have all the key ingredients to successfully answer this challenging question, but clearly motivating your thought process is the key here.]*

4. In the discussion about the design of these multi-arm adaptive clinical trials, there has been a big controversy if it were OK to drop the standard-of-care arm, or instead the standard of care arm should have always remained in the set of available treatments. What would you suggest? *[Again, there's no set answer, but it is important to see how you motivate your answer. You should provide compelling motivation for the decision, whatever that might be.]*

Problem 4

We consider a problem related to the game of baseball. During a year of games, different players have different numbers of opportunities at bat, and on some of these opportunities a player might actually hit the ball. In American Major League Baseball, the ball is pitched very fast, sometimes at speeds exceeding 90 miles (145 km) per hour, and batters typically hit the ball on about only 23% of their opportunities at bat. That ratio, of hits divided by opportunities at bat, is called the batting average of each player . We can think of it as an indicator of the underlying probability that the player will hit the ball for any opportunity at bat.We would like to estimate that underlying probability, as one indicator of a player's ability.

Players also must play a position in the field when the other team is at bat. Different fielding positions have different specialized skills, and those players are expected to focus on those skills, not necessarily on hitting. In particular, pitchers typically are not expected to be strong hitters, and catchers might also not be expected to focus so much on hitting. Most players have a primary fielding position, although many players perform different fielding positions at different times. For purposes of simplifying the present example, we will categorize each player into a single primary fielding position.

The data consist of records from 948 players in the 2012 regular season of Major League Baseball. You can find the data in the file Batting Averages.csv on Canvas.

For player s, we have his (it was an all-male league) number of opportunities at bat, $N_{s|c}$, his number of hits, $z_{s|c}$, and his primary position when in the field, c_s, which was one of nine possibilities (e.g., pitcher, catcher, and first base).

All players for whom there were zero at-bats were excluded from the data set. To give some sense of the data, there were 324 pitchers with a median of 4.0 at-bats, 103 catchers with a median of 170.0 at-bats, and 60 right fielders with a median of 340.5 at-bats, along with 461 players in six other positions.

1. We want to study how the estimated battling abilities differ between selected pairs of players. We also want to ask, how much batting abilities differ based on the position of the players. Based on this information, propose a model to analyze this data (in formulas, i.e. not with rjags or rstan).

2. Write the previous model in Rjags (or rstan)

3. Discuss your inference on the batting abilities between Pitchers and Catchers. Make sure to use the information from the relevant posterior distributions at best.

4. Discuss your inference on the batting abilities between Pitchers and First Base players. Make sure to use the information from the relevant posterior distributions at best.

5. Discuss your inference on the batting abilities between Wellington Castillo (catcher) and Matt Wieters (catcher). Make sure to use the information from the relevant posterior distributions at best. Do you see any evidence of shrinkage in the estimates?

6. Discuss your inference on the batting abilities between Andrew McCutchen (Center Field) and Jason Castro (catcher). Make sure to use the information from the relevant posterior distributions at best. Do you see any evidence of shrinkage in the estimates?