Bayesian Statistics
Exercises 2.

1. Read the article at `http://www.nytimes.com/2011/01/11/science/`
   `11esp.html`. The following is a quotation from this article:

   > Consider the following experiment. Suppose there was reason
   > to believe that a coin was slightly weighted toward heads. In
   > a test, the coin comes up heads 527 times out of 1,000. Is
   > this significant evidence that the coin is weighted? Classical
   > analysis says yes. With a fair coin, the chances of getting
   > 527 or more heads in 1,000 flips is less than 1 in 20, or 5
   > percent, the conventional cutoff. To put it another way: the
   > experiment finds evidence of a weighted coin with 95 percent
   > confidence. Yet many statisticians do not buy it. One in 20
   > is the probability of getting any number of heads above 526
   > in 1,000 throws. That is, it is the sum of the probability of
   > flipping 527, the probability of flipping 528, 529 and so on.
   > But the experiment did not find all of the numbers in that
   > range; it found just one 527. It is thus more accurate, these
   > experts say, to calculate the probability of getting that one
   > number 527 if the coin is weighted, and compare it with
   > the probability of getting the same number if the coin is fair.
   > Statisticians can show that this ratio cannot be higher than
   > about 4 to 1, according to Paul Speckman, a statistician,
   > who, with Jeff Rouder, a psychologist, provided the example.

    i) Formalize and confirm the statement "Statisticians can show that
       this ratio cannot be higher than about 4 to 1".

   ii) In the above setting, if your prior probability that the coin is
       weighted is 0.5, what is the largest your posterior probability (that
       the coin is weighted) could be?

  iii) Explain how the "largest" value of the ratio described above can be
       considered to be a Bayes Factor in favor of an alternative hypoth-
       esis with a data-dependent prior for the binomial proportion. Pro-
       vide an alternative analysis that computes a Bayes Factor using

instead a Beta(α, β) prior for p under the alternative. Compute the Bayes Factor for $\alpha = \beta = 10$, $\alpha = \beta = 1$, and $\alpha = \beta = 0.5$. Which of these priors would you prefer in this setting? Give reasons for your answer.

2. Suppose $X_i | \theta_i \sim N(\theta_i, 1)$ (for $i = 1, \ldots, n$ with n large). Then $X_i \pm 1.96$ is the conventional 95% CI for θ_i.

   i) Show that, for any fixed value of θ_i, the conventional 95% CI has the correct frequentist coverage properties. i.e. the interval will contain θ_i in 95% of trials.

   ii) Suppose that the true values of θ_i are uniformly distributed on $[-a, a]$ for some a. Derive the posterior distribution for $\theta_i | X_i, a$. Is it true that, for any fixed x, $\Pr(\theta_i \in x \pm 1.96 | X_i = x, a) \to 0.95$ as $a \to \infty$?

   ii) Now suppose that the true values of θ_i are normally distributed: $\theta_i \sim N(0, 1)$. Suppose that our interest focuses on which of the θ_i are positive. We will say we are "confident" that θ_i is positive if its 95% CI contains only positive values. Among all the occasions i in which we are confident that θ_i is positive, what proportion will we be wrong? Provide a mathematical argument, and provide R code that confirms your result by simulation., a

   iii) Comment briefly on how the results from this are related to discussions we had about confidence intervals vs credible intervals in class.

3. Suppose $X_i | \theta_i \sim N(\theta_i, 1)$ (for $i = 1, \ldots, n$ with n large).

4. Consider $X_1, X_2, \cdots \sim \text{Bernoulli}(\pi)$.

   i) If the prior distribution for π is $\pi \sim \text{Beta}(\alpha, \beta)$, derive the posterior distribution for π given $X_1, \ldots, X_n$. Show that the predictive distribution $X_{n+1} | X_1, \ldots, X_n$ is $\text{Bernoulli}((\alpha + \sum_{i=1}^n X_i)/(\alpha + \beta + n))$.

   ii) Now assume that the prior distribution for π is instead a *mixture* of Beta distributions, with prior density

$$p(\pi | w, \alpha, \beta) = \sum_{j=1}^k w_j \text{Beta}(\pi; \alpha_j, \beta_j) \qquad (1)$$

where $w_1 + \cdots + w_k = 1$ are the mixing weights and $\text{Beta}(\cdot; \alpha_j, \beta_j)$ denotes the density of a Beta random variable with parameters α_j, β_j. Show that the posterior distribution for π given $X_1, \ldots, X_n$ is also a mixture of Beta distributions. Give expressions for the (posterior) mixing weights and (posterior) hyperparameters.

5. In problem set 1 you were asked to Complete the exercise in the commented text at the end of `exercises/seeb/train_test.R`.

   (a) Swap your code and results with someone else from the class (henceforth referred to as your "partner"). [Your partner should be someone with whom you have not previously compared code/results for this question - if you need help identifying a partner, let me know.]

   (b) Run your partner's code yourself, and check you get the same answer that they report. If not, communicate with your partner to resolve the difference, and report the outcome.

   (c) Compare the results of your code with the results of your partner's code. Are they the same? If not, communicate with your partner to resolve the difference, and report the outcome.

   (d) Compare the way that your partner tackled the problem with the way you tackled the problem. Comment on any similarities and differences. Is their code easier to read or harder to read than yours? Did they use any R tricks or functions that you were unfamiliar with? If possible improve your code (and correct it if necessary) as a result of what you have learned. Hand in your improved code and its output.