FINAL EXAM

STAT 3303: Bayesian Analysis and Statistical Decision Making

Due: Friday, April 27, 2018 (submit on Carmen before 5pm)

INSTRUCTIONS

This is an exam and should be treated as such. **DO NOT discuss any aspect of this exam with anyone other than Professor Calder.** This includes, but is not limited to, not discussing report structure, coding problems, instructor expectations, how long the exam took you, whether you finished it, etc. You are responsible for ensuring that other students do not have access to your exam. Any violation of these instructions constitutes academic misconduct and will be reported to the university's Committee on Academic Misconduct.

In response to growing concerns about a new strain of influenza (K9C9) that has been identified in humans in 10 countries across the world, medical researchers have developed an inexpensive diagnostic test (named "EZK") and a treatment option (a drug named "Neonicon") for individuals infected with the virus. Unfortunately, the EZK diagnostic test is not perfect (i.e., results in false positives and false negatives) and treatment with Neonicon is expensive. For this exam, you will analyze data related to both the new diagnostic test and treatment.

PART I

In an effort to quickly assess the diagnostic ability of EZK, the World Health Organization sponsored a small clinical trial run in each of 10 countries where the K9C9 virus in endemic. Using a highly accurate (very expensive) diagnostic test, 100 randomly selected subjects in each country were tested for K9C9 – this test does not perfectly diagnose infection status of the subjects, but is believed to be far more accurate than the less expensive EZK test. Each subject was then administered the EZK test. As expected, not all of the results of the EZK test agreed with the highly accurate diagnostic results. In the data set `flu.txt` (available on Carmen), the following variables contain the results of the clinical trial:

`Infected` binary indicator of whether the subject is infected (1) or not infected (0) according the highly accurate diagnostic test

`EZK` binary indicator of whether the subject's EZK test was positive (1) or negative (0)

Country country of residence of the subject, where the countries are labeled A-J

Propose a Bayesian hierarchical model for K9C9 status, with `Infected` as the outcome. Model the probability of a subject having the virus as determined by the highly accurate test (`Infected`) as a parametric function of the results of the EZK test. Explain how your model can be used to assess the diagnostic ability of the EZK test. Allow model parameters to vary across country using a hierarchical model structure to account for potential genetic variation in the virus (i.e., the virus and accuracy of the tests may be slightly different

in across countries). Fit your model using rjag/JAGS and provide appropriate and interesting summaries of your results.

In writing your response to this part, be sure to:

1. define all variables and model parameters in the context of the problem
2. specify your model in detail, including conditional independence and prior assumptions (providing your JAGS model files is not adaquate)
3. provide details on model fitting (what were your starting values, how many iterations did your algorithm run, how did you diagnose convergence of the model fitting algorithm)
4. provide interpretations of the results of your statistical analysis in the context of the problem

PART II

An insurance provider in Country D would like to use your results from Part I to decide whether it is cost effective to treat patients who test positive for K9C9 using the the EZK test. (Rolling out the highly accurate test for use on the general population is known to be cost prohibitive.) The cost to the insurance company of treating an individual with Neonicon, regardless of whether the individual has K9C9 or not, is $457, and if an infected individual is treated he/she is not expected to generate any additional costs to the insurance company as a result of complications from the infection. If an infected individual does not receive Neonicon, the cost to the insurance company is $1,490, which takes into account the potential medical complications that may arise from not treating the individual. There is no cost to the insurance company of uninfected individuals not being treated. Using inferences derived from your hierarchical Bayesian model and a decision theoretic criterion, should the insurance company approve Neonicon presciption for individuals who test positive for K9C9 on the EZK test. Be sure to explain how you come to your conclusion about what the insurance company should do, assuming the insurance company's goal is only to minimize costs.

When preparing your final report, please be sure to:

- Your report should be typed. You may use R Markdown if you wish, but DO NOT include any code in the main body of your report.
- Carefully proofread and spell check your report. Write in complete sentences and in paragraphs, not bulleted lists.
- Define all mathematical notation in the text of the report.
- Make sure all figures/tables are straightforward to understand, have captions, and are referenced in the text.
- Include commented R/JAGS code in an appendix.
- You may assume that the reader is familiar with Bayesian statistics, but not that he/she is familiar with the content of STAT 3303. For example, do not refer to specific examples that we have discussed in class (e.g., "the schools model" or "the cats data").

Your report should be **no longer than five pages** double-spaced, including figures and tables. (Text, figures, and tables that are after five pages may not be considered by the instructor.) Your appendix with R/JAGS code does not count toward the five page limit.

Submit your final exam report as a **single PDF file** on Carmen before the deadline.