Applied Bayesian Statistics
Final examination

Robin Ryder

March 8, 2022

Documents: you may use notes, books and other documents, as well as access the Internet. Any attempt to use any form of e-mail or messenging or to post on a forum will result in immediate disqualification.
No phones allowed.
Duration: 3 hours.
Students may answer in English or French. All code must be written in the R language.
At the end of the examination, you must hand in your answers written on paper AND send your R code to `ryder@ceremade.dauphine.fr`.
Please contact the examiner if you wish to hand in your answers early. Please make sure that your R code has been correctly received before leaving the room.
Make sure to save your code on a regular basis. Loss of data following computer failure shall not entitle you to extra time.

A chess game involves two players: one with the white pieces, the other with the black pieces. The game may end in either a win for white, a win for black, or a draw. At high level chess, draws are common, but white has an advantage.

Your data corresponds to the 2016 Candidate's tournament, during which 8 players faced off for a chance to challenge World champion Magnus Carlsen; Sergey Karjakin won the tournament. Each player played each other player twice (once as white, once as black), for a total of 14 rounds. We denote by Y_{ij} the result of the game with candidate i playing white and candidate j playing black. Each row of the data corresponds to one game, and includes:

- round number $1 \leq t \leq 14$
- player i with white pieces
- player j with black pieces
- result Y_{ij}: 1 if white wins, 0 if black wins, $\frac{1}{2}$ for a draw

The aim of this examination is to model these data. Each section corresponds to a model. The sections are mostly independent; they are ordered by increasing complexity.

1 Simplistic model

In this section, we assume that there is no advantage for white and that all players have equal strength: there is a parameter $p \in [0, 1]$ such that for all i, j:

$$\mathbb{P}[Y_{ij} = 1] = \mathbb{P}[Y_{ij} = 0] = \frac{p}{2} \qquad \mathbb{P}\left[Y_{ij} = \frac{1}{2}\right] = 1 - p.$$

1. Write the likelihood associated with this model.
2. Calculate Jeffrey's prior for p.
3. Verify that the Beta family of priors is conjugate for this model.
4. Pick a prior and explain your choice. For this prior, give the posterior distribution.
5. How sensitive are the posterior mean and variance to your choice of prior?
6. Is this model a good fit to the data? Justify your answer, using the method of your choice.
7. Compute the marginal likelihood associated with this model.

2 Advantage to white

We now assume that the probability of winning is greater for white than for black. In this model, we have parameters p_1, p_2 and p_3 such that for all i, j

$$\mathbb{P}[Y_{ij} = 1] = p_1 \qquad \mathbb{P}[Y_{ij} = 0] = p_2 \qquad \mathbb{P}\left[Y_{ij} = \frac{1}{2}\right] = p_3$$

and $p_1 + p_2 + p_3 = 1$.

In particular, if $p_1 > p_2$ then White has an advantage over Black.

We shall use a Dirichlet prior. The *Dirichlet*$(\alpha_1, \alpha_2, \alpha_3)$ distribution is the distribution over the simplex $\Delta_2 = \left\{(p_1, p_2, p_3) \in \mathbb{R}_+^3 : \sum p_i = 1\right\}$ with density

$$\frac{\Gamma(\sum_i \alpha_i)}{\prod_i \Gamma(\alpha_i)} \prod_i p_i^{\alpha_i - 1}.$$

8. Check that the model of section 1 is a special case of this model.
9. Verify that the Dirichlet family of priors is conjugate for this model.
10. Find a package which allows to sample from the Dirichlet distribution. Choose and justify a value for the prior parameter α.
11. Get a sample from the posterior distribution.
12. What is the posterior probability of $\{p_1 > p_2\}$? How sensitive is it to your choice of prior?
13. Compute the marginal likelihood associated with this model.
14. Compare this model to the model of section 1. Which model do you prefer?
15. Is this model a good fit to the data? Justify your answer, using the method of your choice.

3 Davidson-Bradley-Terry model

We now use Davidson's generalization of the Bradley-Terry model. We assign a latent variable Z_i to each player i, denoting their strength. In this model,

$$\begin{aligned}
\mathbb{P}[Y_{ij} = 1] &= \frac{\exp(Z_i)}{\exp(Z_i) + \exp(Z_j + \gamma) + \exp\left(\delta + \frac{Z_i + Z_j + \gamma}{2}\right)} \\
\mathbb{P}[Y_{ij} = 0] &= \frac{\exp(Z_j + \gamma)}{\exp(Z_i) + \exp(Z_j + \gamma) + \exp\left(\delta + \frac{Z_i + Z_j + \gamma}{2}\right)} \\
\mathbb{P}\left[Y_{ij} = \frac{1}{2}\right] &= \frac{\exp\left(\delta + \frac{Z_i + Z_j + \gamma}{2}\right)}{\exp(Z_i) + \exp(Z_j + \gamma) + \exp\left(\delta + \frac{Z_i + Z_j + \gamma}{2}\right)}
\end{aligned}$$

where γ is a parameter which measures the disadvantage of playing black, and δ is a parameter which controls the probability of a draw.

16. Check that the model of section 2 is a special case of this model.
17. Propose a prior for the parameters.
18. Implement an MCMC strategy to sample from the joint posterior of $(\gamma, \delta, (Z_i)_i)$.
19. Describe how you have checked that the MCMC has converged. What is your effective sample size?
20. What is the posterior probability of $\{\gamma < 0\}$?
21. Are there any pairs of players (i, j) for which you are confident that $Z_i < Z_j$?
22. What is the posterior probability of the event that Karjakin is the player with the maximum strength?
23. Is this model a good fit to the data? Justify your answer, using the method of your choice.

4 Further extensions

Propose other modifications to the model, and check whether they impact your conclusions. A suggestion is given below, but you may propose a new one.

Suggestion: as the tournament advances, player strategy may evolve. Players may take more risks in the later part of the tournament, decreasing the probability of a draw. Propose a model which takes this into account, and compare it to the previous models.