Stat532 - Group Assignment (part 2, questions listed in Assign 8)

Demorest, Lemire and Wilson

11/25/2021

Group component

1. Use your project article to answer the following questions.

a. Briefly explain what computational methods were used by the authors of your group's chosen article to summarize or approximate the posterior distributions of interest. If it is not something we have covered in class yet, is it discussed elsewhere in the text book?

The authors considered becoming a professional sports player as an event and a number of them can be regarded as count of events. A poisson regression model was used to analyze this count data which made it possible to estimate the magnitude of the effects of several factors influencing the likelihood of playing a sport professionally.

The authors ran a MCMC algorithm for 25,000 iterations with five chains where the first 5,000 samples were discarded as burn-in (*note: this seems quite high?*). Only every 100 iteration was saved for a total of 1,000 MCMC samples. Although no code is presented in the paper, JAGS was used for computing the posterior (Plummer, 2012) and result consistency was confirmed with Stan (Stan Development Team, 2013).

b. Briefly explain what methods the authors used to check convergence of the computational algorithms they used to approximate the posterior distribution.

The authors used the Gelman-Rubin statistic (Gelman & Rubin, 1992), $\hat{R}$, which assess convergence by comparing the estimated between-chains and within-chain variances for each model parameter. Additionally, the authors used the Geweke's convergence diagnostic (Geweke, 1992) to check the convergence of the MCMC algorithms. The Geweke convergence diagnostic is a test for equality of the means of the first and last part of a Markov chain, typically the first 10% and the last 50%. If the samples are drawn from a stationary distribution of the chain, then the two means are equal and Geweke's statistic has a standard normal distribution. The test statistic is a standard Z-score: the difference between the two sample means divided by its estimated standard error. The standard error is estimated from the spectral density at zero, and so takes into account any autocorrelation and the Z-score is calculated under the assumption that the two parts of the chain are asymptotically independent. Both of these tests are available in the `coda` package in R.

References

Gelman, A., & Rubin, D. (1992). *Inference from iterative simulation using multiple sequences.* Statistical Sciences, 7, 457–472.

Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to calculating posterior moments.* In J. Berdardo, J. Berger, A. David, & A. Smith (Eds.), Bayesian statistics 4. Oxford: Claredon Press.

Plummer, M. (2012). *Just another Gibbs sampler (ver. 3.4.0).* Retrieved from http://mcmc-jags.sourceforge.net/

Stan Development Team. (2013). *Stan: A C++ library for probability and sampling (ver. 1.3.0).* Retrieved from http://mc-stan.org/