Title: Bayesian Baseball

A multi-level regression model (MLM) for predicting the outcome of a baseball game is presented.

The dataset contains each game played in the 2018 MLB regular season. There are two levels to the data: game-level data and team-level data. Game-level data includes the scores for each game as well as batting and pitching parameters. Team-level data contains fielding parameters for each team.

The model is presented in a Bayesian framework, so many aspects of the model are explained from this perspective. The structure of the likelihood function (ie. regression model) is explained with emphasis on the choice of group-level parameters. Prior distributions for both the model coefficients and correlation matrix are clarified. Regularization and shrinkage are visualized in order to show the benefit of a multi-level structure.

Software and sampling methodology are mentioned, including a visualization of Hamiltonian Monte Carlo simulation. The model is evaluated using WAIC and the insights of a model comparison are presented. Posterior distributions of outcomes are visualized and analyzed in order to predict the probability that a given team wins future games.