Course Project Preproposal: Incorporating Tennis Match structure in Modelling

Geoffrey Liu 03/07/2022

Bayesian models have been used widely across sports data to both predict outcomes and rank players. Most Bayesian models for tennis data choose a specific outcome level to model. For example, a point-level model predicts the total point won per match as a binomial distribution of N, the total number of points played and p, the probability of winning each point (Ingram, 2019). A match-level model instead models the win or loss (Angelini, 2022). Tennis matches are naturally hierarchical, i.e. points leads to games, which lead to sets, which leads to matches. The person who wins the most points does not necessarily win the match. The research question is to formulate and fit a model that incorporates both match and point-level data by formulating a hierarchical structure (optionally, could extend to games and sets). This will require understanding and replicating previous Bayesian hierarchical models and experimentation. I would like to use and learn NumPyro/Jax for this project while previous examples I have seen have been in Stan (e.g. https://github.com/martiningram/tennis_bayes_point_based).

(Angelini, 2022) https://www.sciencedirect.com/science/article/abs/pii/S0377221721003234

(Ingram, 2019) https://www.degruyter.com/document/doi/10.1515/jqas-2018-0008/html