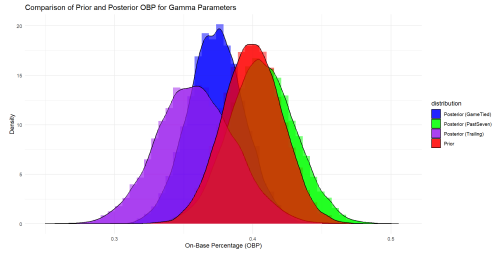
Executive Summary

Bayesian MLB Hitting - Grant and Rafey



Introduction

Taking inspiration from the study “Semi-parametric Bayesian Inference for Multi-Season Baseball Data,” which focuses on analyzing binary sequences for every player in the MLB over four seasons using an auto-regressive logistic model and the Dirichlet process, we analyzed batting performance in a similar way using in-game covariates for Aaron Judge and Shohei Ohtani.

Compared to frequentist methodologies, Bayesian statistics allows to account for the probability of the hypothesis; when new evidence is presented, the probabilities are subsequently updated. Utilizing OBP, we were particularly interested in finding out the “clutch” performance of Aaron Judge and Shohei Ohtani. We define clutch performance to be a positive posterior distribution of the ‘On Base Percentage’ (OBP). More than that, we wanted to go in depth and explore particular moments in the game when these players perform better relative to their career OBP, which we used as the prior for theta in our models.

Results:

Our overall analysis of Shohei Ohtani pointed to him not being “clutch” within our definition of the term. Although, we did see with Ohtani that his OBP posterior did not diminish as much in certain situations (GameTied, PastSeven) when compared to Judge. So we can at least say that Ohtani plays at a relatively constant level across games inside and outside of these “clutch” events. There is some evidence to suggest that Ohtani is “clutch” when the bases are loaded and the game is tied past the seventh inning, but the data does not inform a strong posterior estimation. How much we believe Ohtani to be clutch is dependent on how optimistic we are regarding some of the extremely wide posteriors generated that point to OBP moving in a positive direction in these “clutch” situations.

In terms of the “clutch” factor, Judge and Ohtani have very similar posterior distributions for both the 9th and 11th inning. However, Judge’s distribution is centered and distributed closer to the prior for the 10th inning. Surprisingly, that is the only factor Judge leads over Ohtani.

Ohtani is much better in scenarios where the bases are loaded and the game is tied, as well as in those critical instances when the bases are loaded, it's past the seventh inning, and the game remains tied. The only situation that comes close to these in terms of intensity is when the bases are loaded and the game has stretched beyond the seventh inning. Hence, we conclude that Ohtani is more “clutch.”