NC STATE UNIVERSITY

A SHINY Markov Machine for decision-making in MLB

2019 New England Symposium on Statistics in Sports, Sept, 2019
Jason A. Osborne (NCSU), Melody Wen (NC School of Sci & Math),
Daniel Dulaney (Pittsburgh Pirates)

URL : http://shiny.stat.ncsu.edu/jaosborn/MLBdecideR

Introduction

Some important decisions in baseball:

(a) batting order

(b) whether to attempt to steal a base

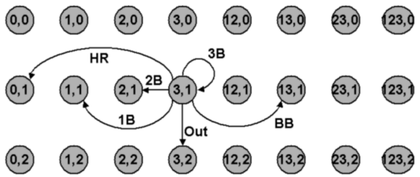
(c) whether to sacrifice a runner.

Good answers to (b) & (c) depend on answer to (a)

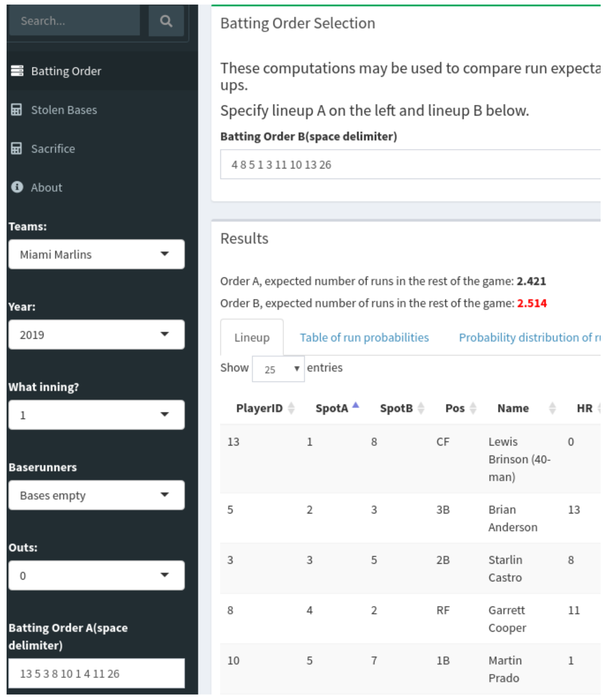
A SHINY app implements a Markov model and allows users to select teams from 2014-2019 and order lineups and specify game situations to obtain estimates of $\mathbf{P(R = r)}$ for $r = 0, 1, 2, \ldots, 20$. These distributions, including means and $P(R > 0)$ can be compared across outcomes to assist in decision-making.

Methodology

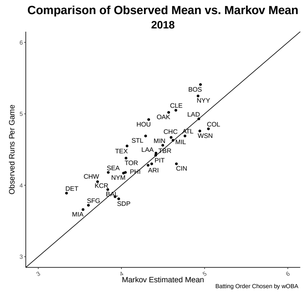
A *Markov Chain* model for one half-inning of a baseball game has 24 *transient states*, or combos of baserunners and outs (Diagram from (Sokol, 2004)):

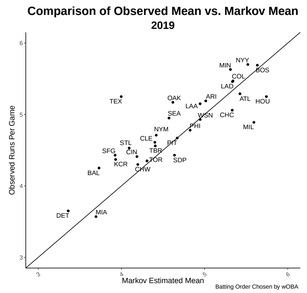


As game progress, it *transitions* between states.
For example, suppose a runner's on 3^{rd} with 1 out (3,1). A BB transitions game to (13,1), a HR to (0,1) (and scores 2 runs) and so on. $\mathbf{P(R = r)}$ governed by *Transition Probability Matrix* estimated by substitution of empirical frequencies from MLB data selected by user. This SHINY app implements an algorithm by (Bukiet et al., 1997) that enables user to consider full nine-inning games.

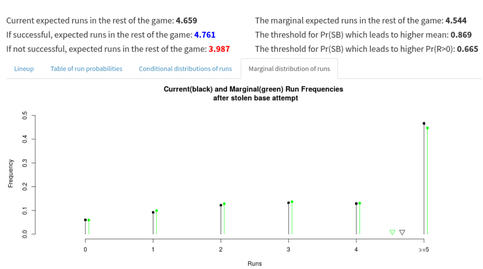


Agreement between observed and Markov means not bad, $r_{2018} = 0.85, r_{2019} = 0.84$:





- 2019 Marlins opening day lineup was terrible ($\widehat{E}(R) = 2.42$).
- So far in 2019, $\overline{\mathbf{RS}} = \mathbf{3.57}, \overline{\mathbf{RA}} = \mathbf{4.57}$. Rearranging according to wOBA (Tango et al., 2006) leads to .09 more runs per game, or $162 \times .09 = \mathbf{14.6}$ more runs per season, or, from Pythagorean calculus, 2 more season wins. Lineup from July 24 even better ($\widehat{\mathbf{E}}(\mathbf{R}) = \mathbf{3.131}$).



Attempt a Steal?

Whether or not to steal is a question. Marginal 2018 league success rate was 0.72 (app default). Announcers sometimes give .7 or .75 as the threshold beyond which it makes sense to steal. Main Point: threshold for $\mathbf{P(SB)}$ where marginal run expectancy after a stolen base attempt exceeds that when not attempting a stolen base depends on many things . . .

- who's up, who's on deck, and the entire lineup!
- how many outs, what inning!

The same is true for consideration of $P(R > 0)$.
Consider 2019 Braves' Ronald Acuña.
SB attempts as leadoff hitter in 65 games: 24
SB attempts as cleanup hitter in 36 games: 2
The table below uses $\widehat{\mathbf{P(SB)}} = \mathbf{21/25} = \mathbf{.84}$ and gives $\mathbf{E(R)}$ in remainder of game for 1^{st} inning and $\mathbf{P(R > 0)}$ for 9^{th} inning and corresponding thresholds for $\mathbf{P(SB)}$ that increase the marginal measures:

Lineup	Inning 1	Inning 9
Leadoff	Current mean 4.659	$P(R > 0 \vert \text{no att}) = .441$
	Marginal mean **4.637**	$P(R > 0 \vert \textbf{att}) = .491$
	Mean if SB **4.761**	$P(R > 0 \vert SB) = .546$
	Mean if CS **3.987**	$P(R > 0 \vert CS) = .206$
threshold	$\mathbf{P(SB) > .87}$	$\mathbf{P(SB) > .69}$
Cleanup	Current mean 4.82	$P(R > 0 \vert \text{no att}) = .358$
	Marginal Mean **4.834**	$P(R > 0 \vert \textbf{att}) = .432$
	Mean if SB **4.945**	$P(R > 0 \vert SB) = .484$
	Mean if CS **4.255**	$P(R > 0 \vert CS) = .156$
threshold	$\mathbf{P(SB) > .82}$	$\mathbf{P(SB) > .617}$

Conclusions: despite Acuna's high success rate, this analysis does not support running when reaching first base as the leadoff batter in the game, but it does support running if he bats cleanup.
There is also an app to consider sacrificing.

References

Bukiet, B., Harold, E. R. and Palacios, J. L. (1997), 'A markov chain approach to baseball', *Operations Research* **45**, 14–23.

Sokol, J. S. (2004), 'An intuitive markov chain lesson from baseball', *INFORMS Transactions on Education* **5**, 47–55.

Tango, T., Lichtman, M. and Dolphin, A. E. (2006), *The Book: Playing the Percentages in Baseball*, TMA Press.