NC STATE UNIVERSITY

2018 Undergraduate Research Symposium
Department of Statistics
Daniel Dulaney, Patrick Snook
Dr. Jason Osborne

A Markov Lineup Tool for Baseball

URL : `http://shiny.stat.ncsu.edu/jaosborn/nineinnings/`

Introduction

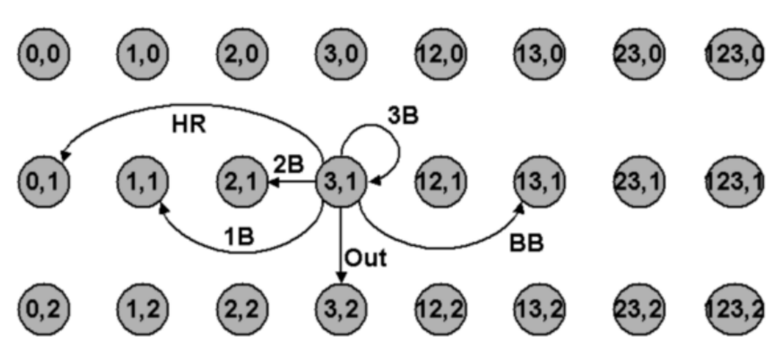
How to order hitters in the lineup is a favorite topic among baseball pundits, fans and "sabermetricians."

Conventional lineups bat basestealers leadoff and sluggers cleanup. More recently, many teams have favored putting their best hitters in the third spot.

We develop a lineup tool that enables users to choose their own lineup and compute a run expectancy. They can choose any team from major league baseball for any of the seasons between 2014 and 2018, or they can choose the 2017 NC State Baseball club.

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:

- A walk would bring us to (13,1)
- A home run would bring us (0,1) (and score 2 runs)
- A triple would bring us back to (3,1) (and score 1 run)
- An out would bring us to (3,2)

Probability of transition from state i to j an element of **transition probability matrix** (TPM), $\mathbf{P}$.

For example: Joey Votto on third, with one out, Zack Cozart hitting:

$$P_{12,14} = P(3,1 \longrightarrow 13,1) = \widehat{P}(\text{walk}) = \frac{62}{507} = .12$$
$$P_{12,9} = P(3,1 \longrightarrow 0,1) = \widehat{P}(\text{home run}) = \frac{24}{507} = .047$$

Markov chains simplify calculation of state probabilities in the future: Let $\mathbf{P^1}, \mathbf{P^2}, \mathbf{P^3}, \ldots, \mathbf{P_9}$ denote TPM for batters in a lineup.

Then the probability that an inning starting with batter i will be in j^{th} state after k^{th} batter hits given by

$$\mathbf{P^i}_{1,}\mathbf{P^{i+1}} \ldots \mathbf{P^{i+k}}$$

Runs associated with any transition given by the equation

$$r_{ij} = 1 + (|B_i| + O_i) - (|B_j| + O_j)$$

Run expectancy calculated as average of all r_{ij} weighted by corresponding transition probabilities.

Not hard to extend this methodology to nine-inning game.

Team: CIN
Year: 2017
What inning? 1
Baserunners Bases empty
Outs: 0
Batting Order A(spots separated by spaces) 7 9 2 6 5 8 4 1 18
Batting Order B(spots separated by spaces) 2 4 8 3 5 6 1 11 18

- **SHINY app** requires only a browser. User selects
  - team, year and baserunner/out/inning situation (often choosing beginning of game)
  - two batting orders to compare
- **SHINY app** runs R software on NCSU server . . .
  - uses statistics grabbed from web to estimate $\mathbf{P}$ for every player.
  - calculates `wOBA` to rank players:
  - calculates *probabilistic run expectancy* according to user-specified batting order over the remainder of a nine-inning game.
- Cincinnati's 2017 most frequent lineup averages 3.687 runs. Instead of batting speedy Billy Hamilton leadoff, move All-Star Joey Votto there, followed by Zack Cozart and gain an average of 0.5 runs per game, or **81 runs per season**!

Results:
Order A, expected number of runs in the rest of the game: **3.687**
Order B, expected number of runs in the rest of the game: **4.197**

Lineup | Cumulative Run Exp. Plot

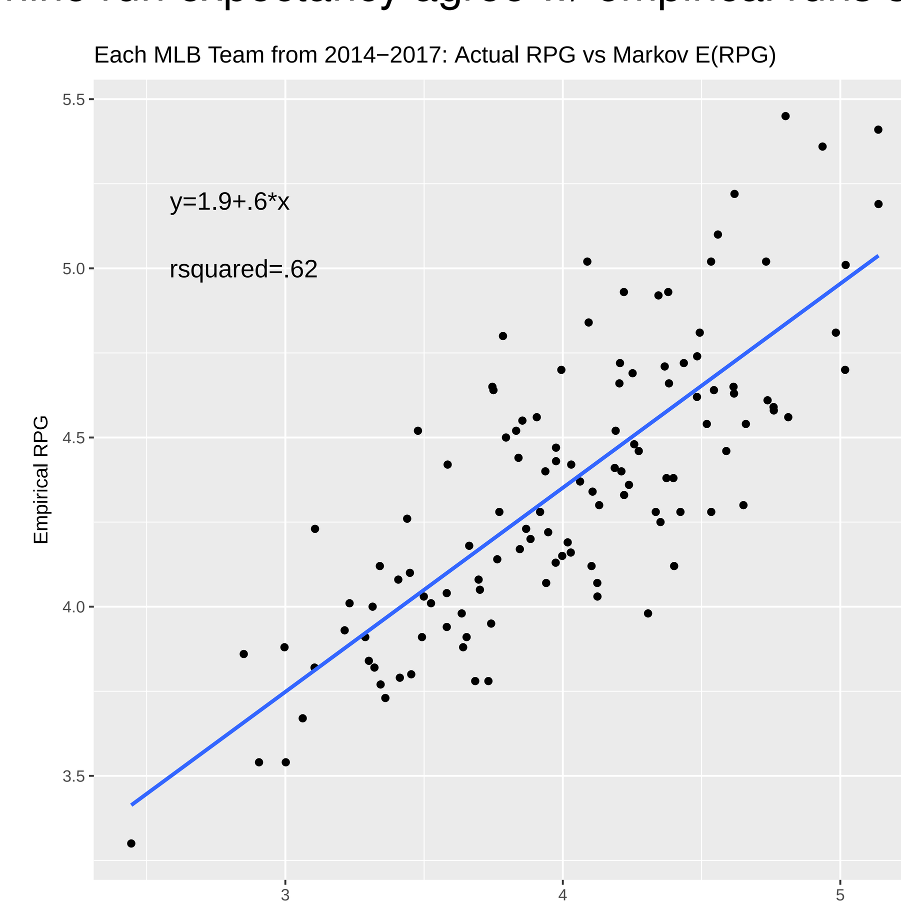
Show 10 entries — Search:

Spot	Pos	Name	HR	Trip	Doub	Hits	BB	PA	OBP	wOBA
1	C	Tucker Barnhart#	7	2	24	100	42	423	0.347	0.323
2	1B	Joey Votto*	36	1	34	179	134	707	0.454	0.438
3	2B	Scooter Gennett*	27	3	22	136	30	497	0.342	0.377
4	SS	Zack Cozart	24	7	24	130	62	507	0.385	0.402
5	3B	Eugenio Suarez	26	2	25	139	84	632	0.367	0.364
6	LF	Adam Duvall	31	3	37	146	39	647	0.301	0.336
7	CF	Billy Hamilton#	4	11	17	144	44	633	0.299	0.283
8	RF	Scott Schebler*	30	2	25	110	39	531	0.307	0.340
9	MI	Jose Peraza	5	4	9	126	20	518	0.297	0.278

$wOBA$ ranks hitters by weighting events "appropriately" (Tango et al., 2006).

$$wOBA = \frac{.69uBB + .72HBP + .89S + 1.27D + 1.62T + 2.1HR}{AB + uBB + SF + HBP}$$

Does Markov Machine run expectancy agree w/ empirical runs scored?

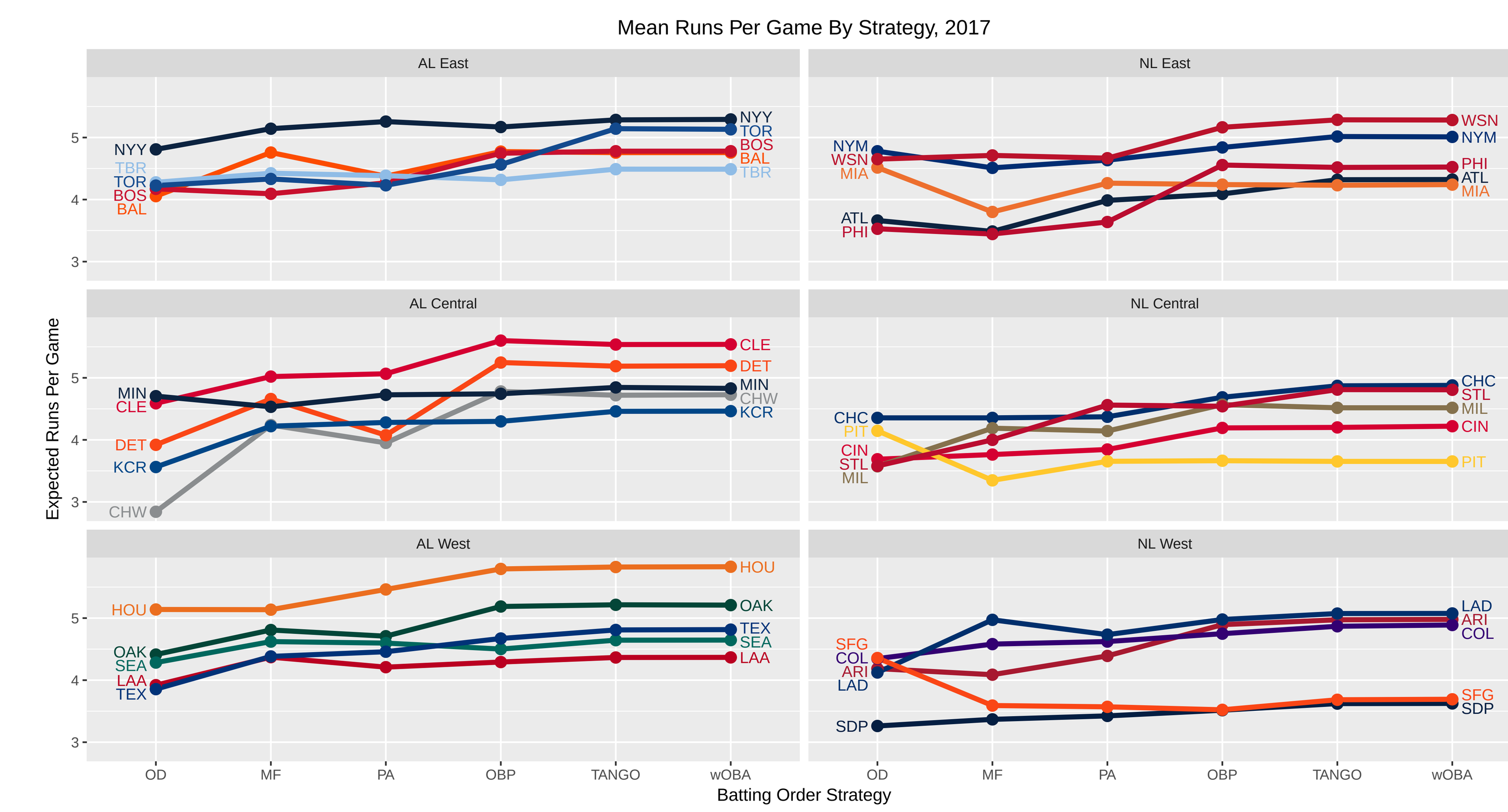


Lineup Strategies

Six lineup strategies evaluated:

- Opening Day (OD)
- Most Frequent (MF)
- By plate appearance (PA)
- By OBP
- TANGO using wOBA
- By wOBA

Tango: best hitters 1,2 and 4, next best in 3^{rd} and 5^{th} slots, then 6-9.



Takeaways

- Run expectancies estimated by the "Markov machine", with most frequently used lineups as input give good agreement with empirical runs scored over the four-year period of investigation.
- If run expectancies believable, then teams could improve on strategies they currently employ in choosing lineups.
  - opening day (OD) lineup almost always the worst. (Lineups evolve/improve over season)
  - Tango ordering almost always better than opening day and most frequent(MF)
  - machine mirrors higher empirically higher AL scoring.

Future work

- Including double play (GIDP) events
- Incorporating TPM changes for
  - 2 outs with/without full count
  - runner on third < 2 outs
  - handedness of pitcher
- Optimization ($9! = 362880$)
- Probability distribution: $P(R = 0), P(R = 1), P(R = 2), \ldots$
  - more informative for in-game decision-making
- Equating run expectancy with wins.

Limitations

- Stolen bases, threats of stolen bases, runners in motion
- Doesn't consider in-game batting strategy (sacrifice/hit-and-run/moving runners over)
- Managerial alternate hitter handedness to confound relief pitchers

References

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.