

System ◇○

Slim Lim

EECS Department
UC Berkeley

OPLSS 2025

Problem

Baseball is poorly marketed to the average OPLSS attendee

Goals of this talk

Objective 1

Convince you that baseball is a rich, fascinating sport

Objective 2

Equip you to understand baseball games as a spectator

Outline

- Motivation
- Rules
- Discussion
- Future work

What's interesting about baseball?

Two months ago, I knew nothing about baseball and honestly thought it was pretty stupid.

- Not interested in sports, American patriotism
- Games have a reputation for being long and uneventful
- Seems to be: “which guy can smack the ball the hardest”

What else is there?

In a nutshell: entropy

Baseball is a sequence of **discrete probabilistic outcomes**

- Surprisingly bounded state space
- Individual plays are highly unpredictable
- Game structure causes nondeterminism to **compound**

Well-modeled by discrete-time Markov processes¹²

¹Bukiet, B., Harold, E.R., Palacios, J.L. (1997) A Markov Chain Approach to Baseball. *Operations Research* 45(1):14-23.

²D'Angelo, J.P. (2010) Baseball and Markov Chains: Power Hitting and Power Series. *Notices of the AMS* 57(4), April 2010.

Variance

Extremely **high variance** compared to other professional sports

- 162 games/season
- Each matchup is a **series** of 3 games, over consecutive days

Transitivity often violated

- *Dodgers > ... > Rockies* in overall standings
- *Rockies >³ Marlins >³ Giants* in series sweeps

Variance

Extremely **high variance** compared to other professional sports

- 162 games/season
- Each matchup is a **series** of 3 games, over consecutive days

Transitivity often violated

- *Dodgers > ... > Rockies* in overall standings
- *Rockies >³ Marlins >³ Giants* in series sweeps
- *Giants vs. Dodgers* individual games:
 - 1 *Giants, 6-2*
 - 2 *Dodgers, 5-11*
 - 3 *Dodgers, 4-5*

Variance: Home game win probability

*When the best team plays at home in the NBA, it is always favored to win at least 60% of the time [...] ranging from a 68% probability to an 84% probability. Meanwhile, even with a home advantage, it is **rare that the best MLB team is ever given a 70% probability of winning** [...] ranging from **57% to 63%**.³*

³Lopez, M.J., Matthews, G.J., Baumer, B.S., "How often does the best team win? A unified approach to understanding randomness in North American sport," Ann. Appl. Stat. 12(4), (December 2018).

Variance: Baseball has high n , but not enough

Sport	League	Games/season
Baseball	MLB	162
Hockey	NHL	82
Basketball	NBA	82
Football	Brasileirão	68.7
Football	Premier League	46.9
American football	NFL	17

⁴Ben-Naim, E., Hengartner, N.W., Redner, S. et al. "Randomness in Competitions." J Stat Phys 151 (2013).

Variance: Baseball has high n , but not enough

Sport	League	Games/season
Baseball	MLB	162
Hockey	NHL	82
Basketball	NBA	82
Football	Brasileirão	68.7
Football	Premier League	46.9
American football	NFL	17

Would need to play 256 games to ensure the best team wins⁴

⁴Ben-Naim, E., Hengartner, N.W., Redner, S. et al. "Randomness in Competitions." J Stat Phys 151 (2013).

Regression to the mean

- Yet somehow, statistics work
- **Sabermetrics:** many hobbyists; teams employ multiple PhDs for their analytics teams
- *Example:* Batters perform better the **third time** they face a pitcher in a game
 - Use a backup pitcher for the first inning, then send in your main starting pitcher
 - On average, starter can get ~1 more inning

Other interesting bits

- **Signals:** pitching strategy and on-field coordination
 - Teams obfuscate their systems
 - Feasible because of limited state machine
- Home runs are **not standardized**: “HR in $n/30$ ballparks”
 - Defined as “going out-of-bounds into the outfield”
 - But stadium dimensions vary
- MLB team names
 - Canadian team (Toronto Blue Jays)
 - Most are cities, except: Texas, Minnesota, Colorado, Arizona
 - Some cities have two teams (Los Angeles, Chicago)
- **Hand dominance** matters a lot
 - Only 10% of general population is left-handed, but 25-33% of MLB players

Hand dominance: First order effects

- Infield positions strongly favor **throwing right-handed**, due to direction of play
 - Exception: 1B, and maybe catcher
- **Batting left-handed** is slightly closer to first base, easier to hit into gaps

Hand dominance: Second order effects

- **Same-handed matchups** favor the pitcher; easier to hit balls from the opposite side
- Effect is **asymmetric**: hitting LH-on-LH is harder than hitting RH-on-RH
- **Switch-hitters** learn to hit from both their non-dominant side
 - Much easier than switch-pitching

Left-handed players are heavily overrepresented

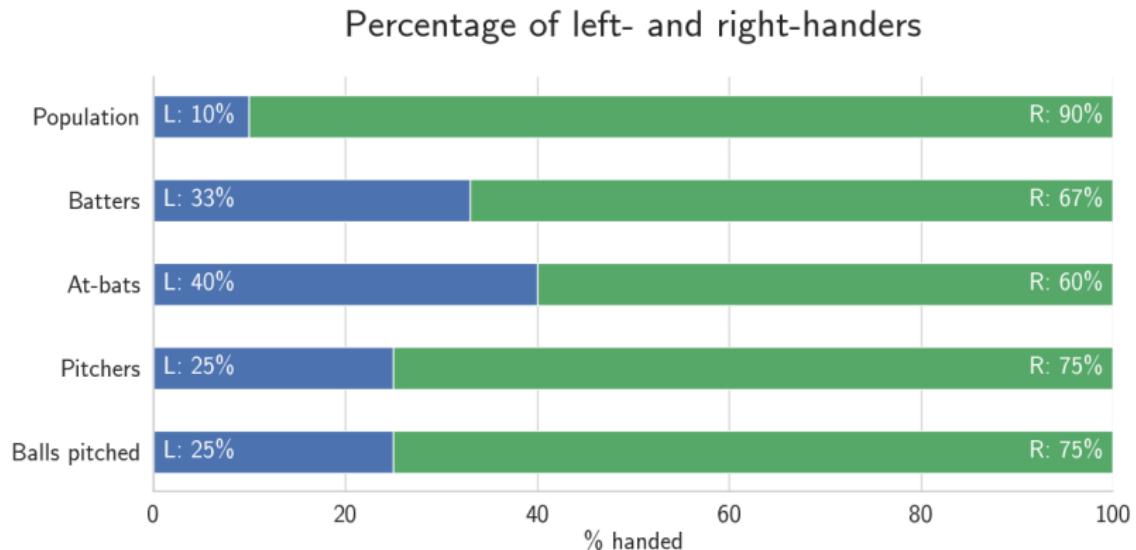


Figure 1: Dominant hand, by position

Gameplay: High-level intuition

- Hit the ball to get on base
- Sprint through all the bases to score a **run**
- Winner = team with most runs after 9 innings

This leaves a lot of gaps

- What happens for the other 95% of the game?
- Are the pitcher and batter on the same team, opposing teams, or are the pitchers a separate third faction?
 - What is the point of the pitcher, anyway?
- What is a “ball”?
- What is Minor League Baseball and why do all of the teams have such fundamentally unserious names?
 - Akron RubberDucks, Pensacola Blue Wahoos, Rocket City Trash Pandas, Albuquerque Isotopes

Field layout

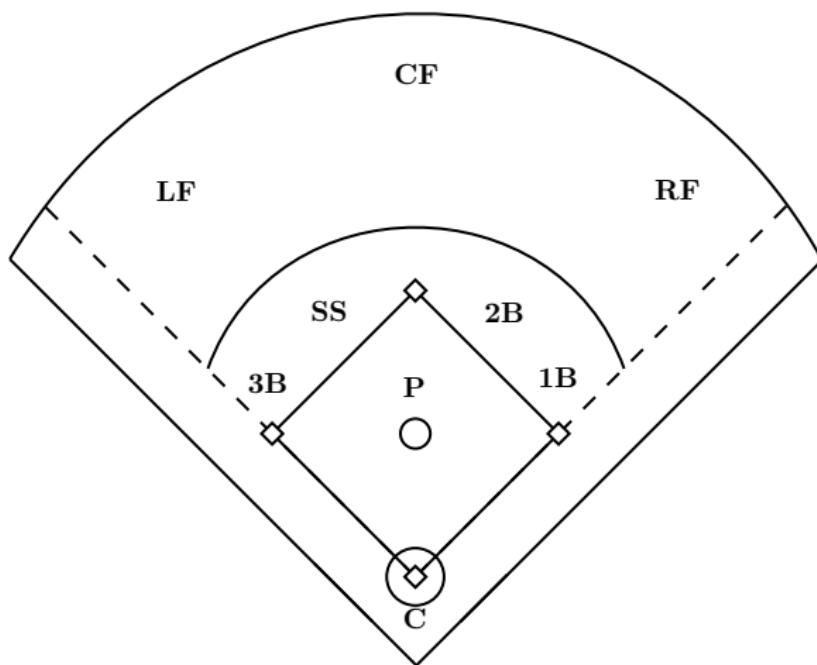


Fig. 1: Defensive positions

Field layout: Infield

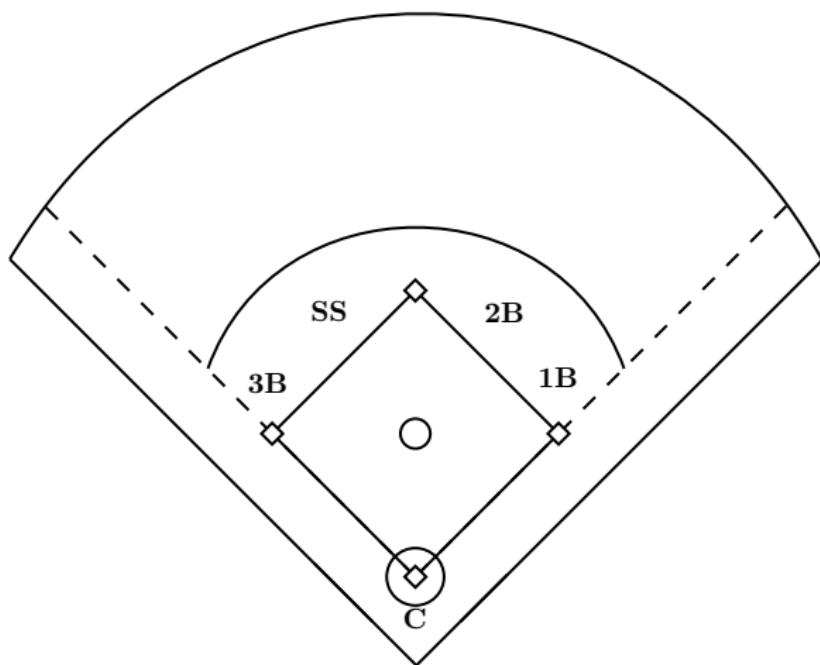


Fig. 2: Defensive positions, infield

Field layout: Outfield

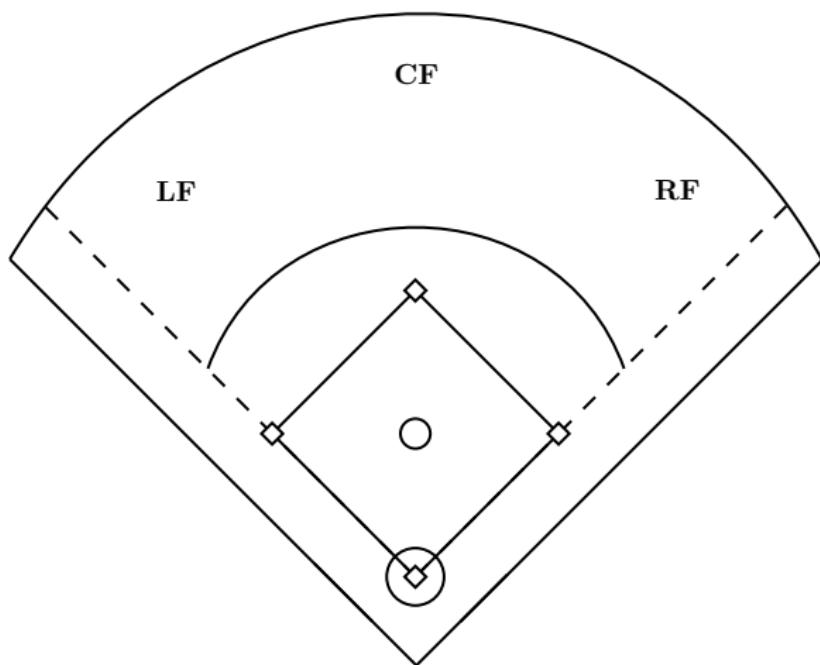


Fig. 3: Defensive positions, outfield

Pitchers vs. position players

Roster divided in ~half

- **Pitchers** are the defensive anchor of the team
- **Position players** switch between offense and defense
- Disjoint skillsets; most players can only do one
 - **Two-way players** (Shohei Ohtani) rare, highly valuable

Half-inning

- ① Next batter goes up for plate appearance
 - **Walk** (pitcher fail): batter gets on-base for free
 - **Strikeout** (batter fail): +1 out, next batter
 - Ball **in play**
- ② During play:
 - Offense runs bases, balancing risk/reward
 - Defense tries to catch the ball, pass around to tag runners out
- ③ Repeat until **three outs**

Half-inning

- ① Next batter goes up for plate appearance
 - **Walk** (pitcher fail): batter gets on-base for free
 - **Strikeout** (batter fail): +1 out, next batter
 - **Ball in play**
- ② During play:
 - Offense runs bases, balancing risk/reward
 - Defense tries to catch the ball, pass around to tag runners out
- ③ Repeat until **three outs**

Outs

- **Three strikes** in one plate appearance
- **Ball caught** before touching ground
- **Runner tagged** before safely reaching base

Understanding the at-bat

Where a lot of the action happens!

Understanding the at-bat

Where a lot of the action happens!

How I used to think it worked

- ① Pitcher throws
- ② Batter tries to hit
 - ① If they hit, try to run to first base
 - ② If they miss, strike → bad for batter
 - ③ If ???, ball → bad for pitcher
- ④ 3 strikes = out (batter fail)

Understanding the at-bat

How it actually works (approx.)

- ① Pitcher throws **aiming for invisible rectangle** ("strike zone")
- ② Batter **predicts trajectory and decides to hit**
 - ① If they hit, try to run to first base → good for batter
 - ② If they miss, strike → bad for batter
 - ③ If they don't swing, and **ball misses strike zone**, ball → bad for pitcher
- ③ 3 strikes = out (batter fail)
- ④ 4 balls = **walk** (pitcher fail, batter gets free pass to first base)

Plate appearance

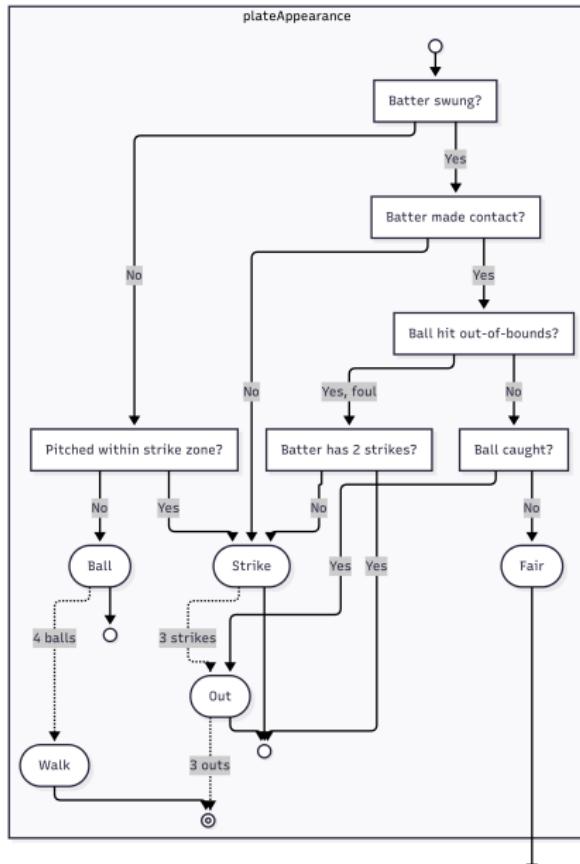
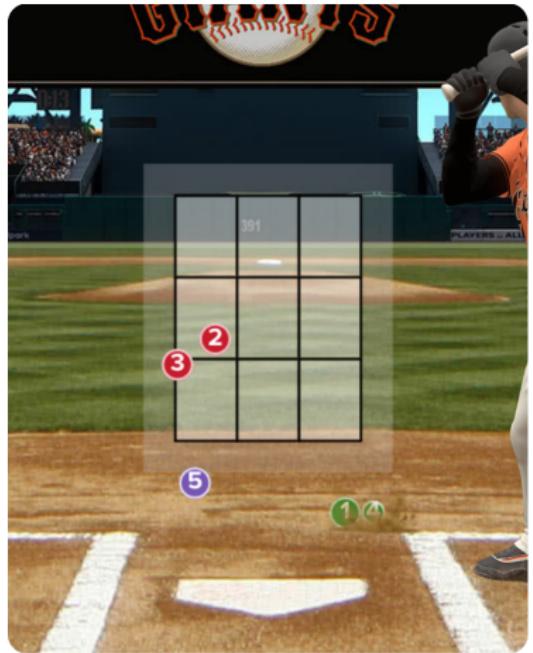


Plate appearance

Checklist

- Did the pitcher throw to strike zone?
- Did batter make contact?
- Did ball go out of bounds?
- Did ball get caught?

Strike zone



1 Ball In Dirt 1 - 0

78.3 mph Curveball

2 Called Strike 1 - 1

94.9 mph Four-Seam Fastball

3 Called Strike 1 - 2

85 mph Slider

4 Ball In Dirt 2 - 2

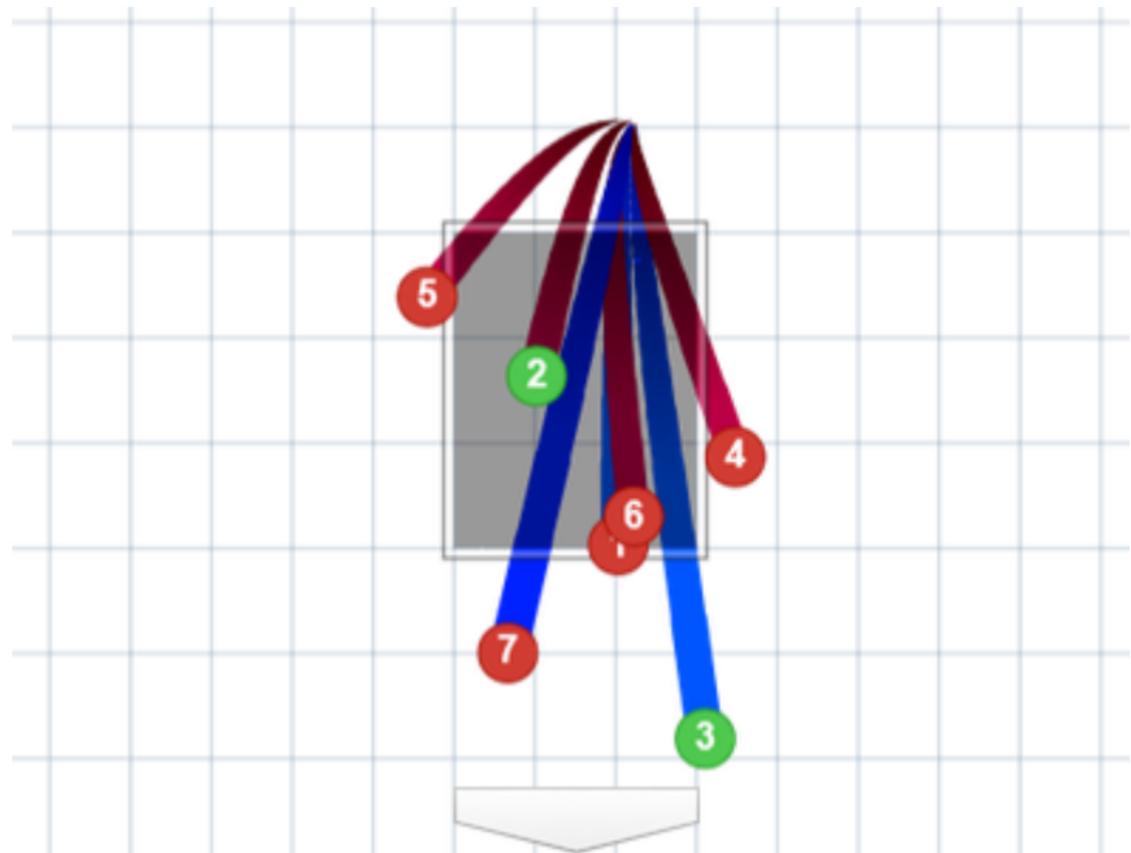
79.6 mph Curveball

5 In play, out(s)

86.1 mph Slider

Figure 3: You can't see it in the game, and nothing makes sense without it

Umpires are fallible



Pitching strategy

- **Faster** is better: 100mph fastballs are normal
 - Four-seam, cutter, sinker
- Outwitting the batter with different pitch types
 - **Breaking ball**: curveball, slider
 - **Weird speeds**: changeup, splitter
- **Location**: strikeouts, batter hot zones
 - e.g. some batters can't hit "outside" pitches
- Catcher's job is to **trick the umpire** by moving glove

First pitch strikes

“Getting ahead in the count”

Count	BA	OBP	SLG	K%	BB%	HR%
1-0	.267	.384	.474	18.9%	15.4%	3.9%
0-1	.220	.266	.364	31.3%	5.0%	2.9%

Pitcher's count (0-2, 1-2)

- Nearly half of plate appearances that start 0-2 end in a strikeout
- Can throw a “waste pitch” far from the zone

Substitution principle

Recall the **platoon advantage** (same-handed matchups favor pitcher)



Offensive Substitution

Offensive Substitution: Pinch-hitter Daniel Johnson replaces Christian Koss.



Pitching Substitution

Pitching Change: Justin Wilson replaces Zack Kelly.



Offensive Substitution

Offensive Substitution: Pinch-hitter Tyler Fitzgerald replaces Daniel Johnson.

Figure 5: Back-to-back player changes

Substitution principle

$$M[x \mapsto N]$$

Substitution principle

$$M[x \mapsto N]$$

$M' = M[\text{Koss}_R \mapsto \text{Johnson}_L]$
 $[\text{Wilson}_{LP} \mapsto \text{Kelly}_{RP}]$
 $[\text{Johnson}_L \mapsto \text{Fitzgerald}_R]$

Batter	Pitcher
Koss (R)	Kelly (R)
Johnson (L)	Kelly (R)
Johnson (L)	Wilson (L)
Fitzgerald (R)	Wilson (L)

Substitution principle

$$M[x \mapsto N]$$

$M' = M[\text{Koss}_R \mapsto \text{Johnson}_L]$
 $[\text{Wilson}_{LP} \mapsto \text{Kelly}_{RP}]$
 $[\text{Johnson}_L \mapsto \text{Fitzgerald}_R]$

Batter	Pitcher
Koss (R)	Kelly (R)
Johnson (L)	Kelly (R)
Johnson (L)	Wilson (L)
Fitzgerald (R)	Wilson (L)

Substitution is *not free*: once a player exits the game, they cannot go back in

Pitching lineup

- **Starting pitchers:** 5
 - Rotate pitching every 5 days to recover
 - Usually pitch for 5-7 innings
- **Bullpen:** ~6-7 relief pitchers
 - Take over for 1-2 innings after the starter gets tired
- **Closer:** specializes in endgame pitching
 - Usually does not pitch in blow-out situations

Pitcher statistics

- **Earned Run Average (ERA)**: how much would the other team score if you pitched a whole game
 - ≤ 2 is elite, 3-4 is great to good
- **Strikeouts (K%)**: can you hit the perimeter of the rectangle
- **Walks (BB)**: can you hit the rectangle at all
- **Innings pitched (IP)**: endurance

Hitting strategy

- Example: “taking” pitches with a (3, 0) count
 - Pitcher has one last chance; hitter has three
 - **89.2%** of the time, if batter doesn’t swing, it’s called a ball
 - But slugging .847 when they do swing
- “Hit it where they ain’t”
- Reading the ball: better than 20/20 vision

Types of batted balls

- **Line drive:** usually best, straight w/ not much arc
- **Ground ball:** not much distance, hard to insta-catch
- **Fly ball:** high parabolic trajectory, easy to catch unless far
- **Pop-up:** fly ball that doesn’t travel, often leads to outs
- **Bunt:** doink

Situational hitting

Depends on the count, base configuration, and number of outs

- Example: sacrifice bunt

$\langle \underbrace{(3, 2)}_{\text{count}}, 1 \text{ out}, \spadesuit \heartsuit | \underbrace{(3, 4)}_{\text{score}}, \overbrace{8^\perp}^{\text{frame}} \rangle$

Hitting statistics

- **Batting average (BA)**: % of time you record a hit
 - No fouls, no outs
 - Elite hitters are close to .300
- **Slugging (SLG)**: harder-hit balls count for more
 - For top hitters, often ~twice their average
- **On-base plus slugging (OPS)**: slugging + batting + walks

“Three True Outcomes”

Previous style: pitchers want to pitch hard-to-hit balls; batters want to get on base

Now, trend toward all-or-nothing gameplay

- Strikeout
- Walk
- Home run

Counterexample: Ichiro

Exercises

Consider two players, B and K :⁵

- B hits only **singles** with probability p , striking out otherwise
- K hits only **home runs** with probability q , striking out otherwise

Let $B(p)$ and $K(q)$ denote the expected number of runs scored per half-inning (until 3 outs are reached), assuming only player B or K is batting in each case.

Assume 3 singles score 1 run, and all at-bats are i.i.d.

⁵Adapted from D'Angelo, J.P. (2010) Baseball and Markov Chains: Power Hitting and Power Series. Notices of the AMS 57(4), April 2010.

Exercises

- ① Set $q = p/4$. For what values of p is $B(p) \geq K(p/4)$?
- ② Set $q = 2p/5$. For what values of p is $B(p) = K(2p/5)$?
- ③ Prove that

$$K(q) = \frac{3q}{1-q}$$
$$B(p) = \frac{p^3(3p^2 - 10p + 10)}{1-p}$$

and sketch their respective graphs for $0 \leq p \leq 1$.

Possible outcomes

```
-- How far did the runner get?  
data Base = First | Second | Third | Home  
  
-- How was the ball hit?  
data BattedBall = Fly | Pop | Line | Ground | Bunt  
  
-- What was the outcome of the plate appearance?  
data Outcome = Walk  
| Strikeout  
| Out BattedBall (Maybe Nat)  
| Hit Base BattedBall (Maybe Nat)
```

Game state $\sigma = \langle c, o, b \mid s, i \rangle$

$$\underbrace{\langle (2, 2), 1 \text{ out}, \diamondsuit \diamondsuit}_{\text{count}} \mid \underbrace{\langle (6, 4) \rangle, 9^\perp}_{\text{score}} \rangle^{\text{frame}}$$

- $c \in \{0, 1, 2, 3\} \times \{0, 1, 2\}$ is the **pitch count** (balls, strikes)
- $o \in \{0, 1, 2, 3\}$ is the **number of outs**
- $b \in \{0, 1\}^3$ is the state of the **bases**
 -  = “bases loaded”
 -  = “runner on first and third”
- $s \in \mathbb{N} \times \mathbb{N}$ is the **score** (home, away)
- $i \in \{1, \dots, 9\} \times \{\top, \perp\}$ is the **inning/frame**
 - 9^\perp = “bottom of the 9th”

Possible outcomes: impact on σ

Outcome	Count	Outs	Bases	Score	Inning
Walk	✓	-	✓	(✓)	-
Strikeout	✓	✓	-	-	(✓)
Hit out	✓	✓	(✓)	(✓)	(✓)
Base hit	✓	(✓)	(✓)	(✓)	(✓)

Possible outcomes: big-step

- **Walk** $\sigma \mapsto \langle \vec{0}, o, b' \mid s', i \rangle$
 - Runner gets free pass to 1B, could drive in a run
- **Strikeout** $\sigma \mapsto \langle \vec{0}, o + 1, b \mid s, i \rangle \xrightarrow{o'=3} \langle \vec{0}, 0, \diamond\lozenge\diamond \mid s, \text{next}(i) \rangle$
 - Outs increment, possible inning change
 - Next batter comes up
- **Hit out** (type, position)
 $\sigma \mapsto \langle \vec{0}, o', b' \mid s', i \rangle \xrightarrow{o'=3} \langle \vec{0}, 0, \diamond\lozenge\diamond \mid s', \text{next}(i) \rangle$
 - Possible base running, scoring by existing runners
 - Outs increment, possible inning change
- **Base hit** (n , type, position)
 $\sigma \mapsto \langle \vec{0}, o', b' \mid s', i \rangle \xrightarrow{o'=3} \langle \vec{0}, 0, \diamond\lozenge\diamond \mid s', \text{next}(i) \rangle$
 - Bases change, possible scoring
 - Outs technically possible for inning change

Judgment

$$\boxed{\sigma \vdash A \text{ play} \dashv \sigma'}$$

Rules we didn't cover

- Double plays
- Force outs
- Infield fly rule
- Stealing bases
- Pickoff attempts
- Balks, wild pitches
- and more!

How to bring baseball into your life

Baseball is **constantly happening**: the Markov process runs every day!

- Not making progress on a key proof? → “*I wonder how the game is going*”
- Advisor micromanaging you? → “*How did my team do today?*”
- Department drama? → “*Let me read about the 2019 Houston Astros sign-stealing scandal instead*”

We are going to a Minor League baseball game **tomorrow, Wednesday June 2** at 6:30pm!

Tickets are **\$12**. See #baseball on Slack for more info.

- Practice your new game-watching skills
- You could win a free car
- Learn how the Minor League system works, and how the MLB antitrust exemption allows for unfair labor practices

How to watch baseball the “official” “real” way

Streaming

- **MLB.TV** (\$120/year) is the “official” “streaming” “service”
 - Many ways to get it for cheaper, but not needed
 - Note: Blackouts
- **Sportsurge.net** but *only* use this URL

How to follow baseball: Radio

Baseball is great on the radio!

- Action in short bursts, amenable to narration
- Pay attention when announcer's voice gets excited
- Use free online radio stations
- **MLB At Bat** (\$20-30/year, included in MLB.TV)
 - Often better than TV announcers; can set audio source on MLB.TV

How to follow baseball: Gameday website

mlb.com/gameday/red-sox-vs-giants/2025/06/20/777425/

The screenshot shows the MLB Gameday interface for the Red Sox vs. Giants game on June 20, 2025. The score is 7-5 in favor of the Red Sox. The game summary table includes columns for Innings, Outs, Hitters, Runs, Hits, Errors, and Pitchers. Below the table are tabs for Key Moments, Scoring, Video, Home Runs, Strikeouts, Hard Hit Balls, and Lead Changes. A "Top 1st" section highlights the first inning with a "Pop Out" by Jaren Duran and a "Strikeout" by Carlos Núñez. At the bottom is a video player showing a play from the game.

Figure 6: Gameday overview

The screenshot shows the pitch data for Connor Norby in the top of the second inning. The count is 3-2. The data table lists six pitches:

Pitch Type	Speed (mph)	Description	Count
Foul	88.3	Cutter	0 - 1
Swinging Strike	91.1	Four-Seam Fastball	0 - 2
Ball	92.3	Four-Seam Fastball	1 - 2
Ball	91.7	Four-Seam Fastball	2 - 2
Foul	89.7	Cutter	2 - 2
Foul	92.8	Four-Seam Fastball	2 - 2

Figure 7: Pitch data

How to follow baseball: WebSocket API

wss://ws.statsapi.mlb.com/api/v1/game/push/subscribe/gameday/777425

```
{  
    "timeStamp": "20250629_181140",  
    "gamePk": "777313",  
    "updateId": "dd076a14-20ec-49c2-9a45-b2c4f2e212b5",  
    "wait": 10,  
    "logicalEvents": [  
        "countChange",  
        "count01"  
    ],  
    "gameEvents": [  
        "swinging_strike"  
    ],  
    "changeEvent": {  
        "type": "new_entry"  
    },  
    "isDelay": true  
}
```

How to follow baseball: mlbt

mlbt (<https://github.com/mlb-rs/mlbt>) Rust TUI

Scoreboard | Gameday | Stats | Standings Help: ?

June 29, 2025 away home time [PDT] status

Giants	45-39	2	White Sox	28-56	9	11:10 am	Final
Carnivalists	69-35	1	Red Wings	60-46	9	10:35 am	Final
Mets	46-37	1	Pirates	35-59	12	10:35 am	Final
Rays	47-37	1	Orioles	36-47	5	10:35 am	Final
Blue Jays	45-38	5	Red Sox	41-44	3	10:35 am	Final
Phillies	49-35	2	Braves	38-45	1	10:35 am	Final
Athletics	37-46	5	Yankees	48-39	12	10:35 am	Final
Padres	45-38	1	Marlins	44-49	10	10:35 am	Final
Cubs	49-35	8	Astros	50-34	2	11:10 am	Final
Dodgers	53-32	5	Royals	39-45	1	11:10 am	Final
Rockies	19-65	6	Brewers	47-37	3	11:10 am	Final
Mariners	43-60	6	Rangers	61-43	4	11:35 am	Final
Nationals	39-55	1	Angels	42-59	1	11:35 am	Final
Marlins	37-45	6	D-backs	43-42	4	11:35 pm	Final
Twins	48-44	8	Tigers	53-32	3	4:30 pm	Final

1 2 3 4 5 6 7 8 9 R H E
SF 0 1 0 0 1 0 0 0 0 2 6 1
WS 1 0 0 0 0 0 4 0 0 5 11 0

player	ab	r	h	rbi	bb	so	lob	avg
1 RF Slater	4	1	1	0	0	0	2	.231
2 SS Meidroth	4	1	2	0	0	0	3	.265
3 LF Benintendi	4	0	0	1	0	1	7	.233
4 DH Vargas, M	3	1	0	0	1	1	3	.229
5 C Teel	4	1	3	2	0	0	2	.319
6 1B Sosa, L	4	0	1	0	0	2	2	.277
7 1B Sosa, L	3	0	0	0	0	0	1	.211
8 CF Taylor, M	3	0	1	0	1	1	3	.214
9 3B Rojas, J	4	0	0	0	0	0	4	.185

© TMUX +3d 1h > 1 mlbt ⌘ ⌘ 96% < 11:28 slim dyn-30-112-18-105-guest

Figure 8: mlbt scoreboard

Scoreboard | Gameday | Stats | Standings Help: ?

bottom 6
Michael A. Taylor called out on strikes.
3 outs
Ryan Noda flies out to right fielder Mike Yasutremski. 2 outs
Lenny Sosa flies out to center fielder Jung Hoo Lee. 1 out
→ Kyle Teel singles on a fly ball to right fielder Mike Yasutremski.

top 6
Rafael Devers strikes out swinging. 3 outs
outs
→ Elliot Ramos flies out to right fielder Mike Touchman. 2 outs
Brett Wiseley pops out to third baseman Josh Rojas. 1 out
→ Patrick Bailey hit by pitch.

exit velo: 70.1 | LA: 32° | distance: 218'
4 In play, no out | 2-1 | 76.7 |
Curveball | Ball | 2-1 | 76.3 |
Curveball |
2 Called Strike | 1-1 | 83.2 |
Changeup |
3 Ball | 1-0 | 76.5 |
Curveball |

© TMUX +3d 1h > 1 mlbt ⌘ ⌘ 96% < 11:21 slim dyn-30-112-18-105-guest

Figure 9: live game view

Acknowledgements

Thank you to Christopher Callahan and Connor O'Brien for patiently explaining baseball to me