Understanding probability using games

Outline for today

Questions about worksheet 5?

Probability models

- Game 1: Big league baseball
- Game 2: All-star baseball

Announcement: DataFest Info session tonight

Information session about DataFest on Wednesday 7pm in ASH 221

More information about DataFest can be found at: http://www.science.smith.edu/datafest/

Questions about worksheet 5?

Probability

Probability is a way of measuring the uncertainty of the outcome of an event

Probability assigns a number between 0 and 1 to the outcome of an event occurring

Pr(event) = 0 if there is no chance of an event occurring

- E.g., The probability that a David Ortiz will hit 100 home runs in a game

Pr(event) = 1 if the event will definitely occur

- E.g., The probability that a strike will be thrown in the 2017 baseball season

Pr(event) ϵ [0, 1] if there is some possibility that an event will occur

- E.g. the probability that Mookie Betts will hit a home run on this next plate appearance

Probability

One way to interpret probability is in terms of the **relative frequency** of an event

If we repeat an experiment N times, we get:

$$Pr(event) \approx \frac{number\ of\ times\ an\ event\ occurs}{N}$$

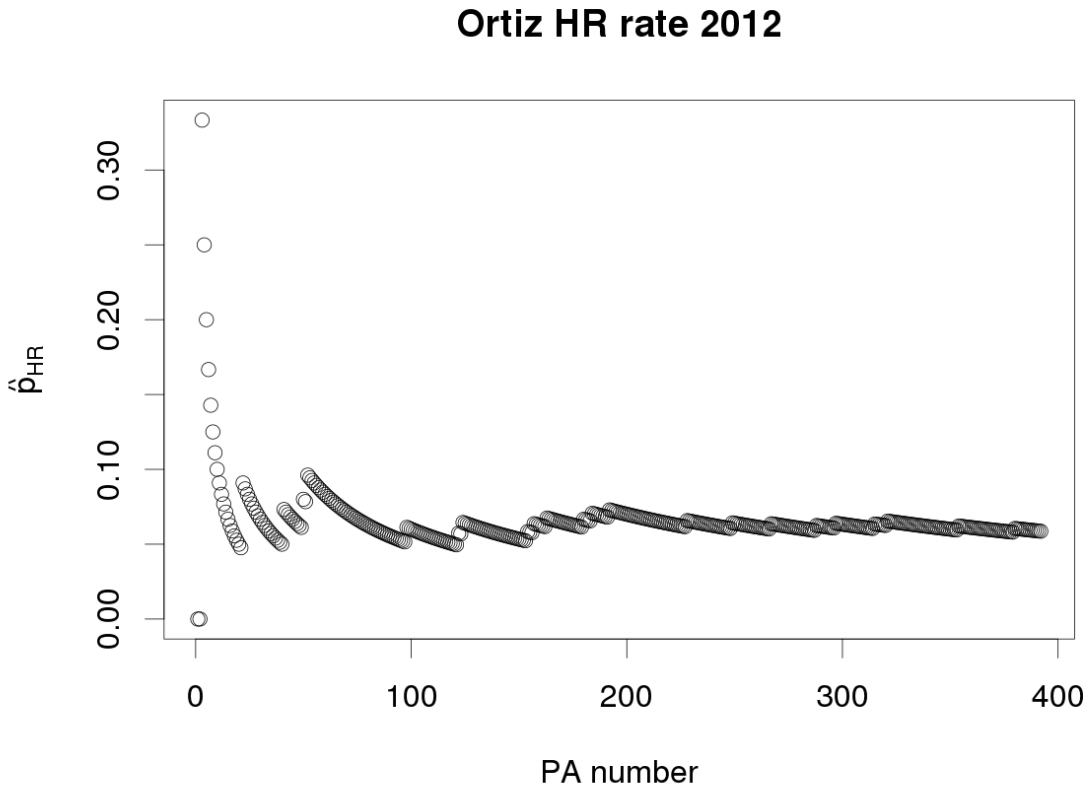
For example, we can estimate the probability that David Ortiz will hit a home run by:

$$Pr(HR) \approx$$

Probability

$\hat{p}_{HR}$ is only an estimate for Pr(event = HR) $\equiv \pi_{HR}$

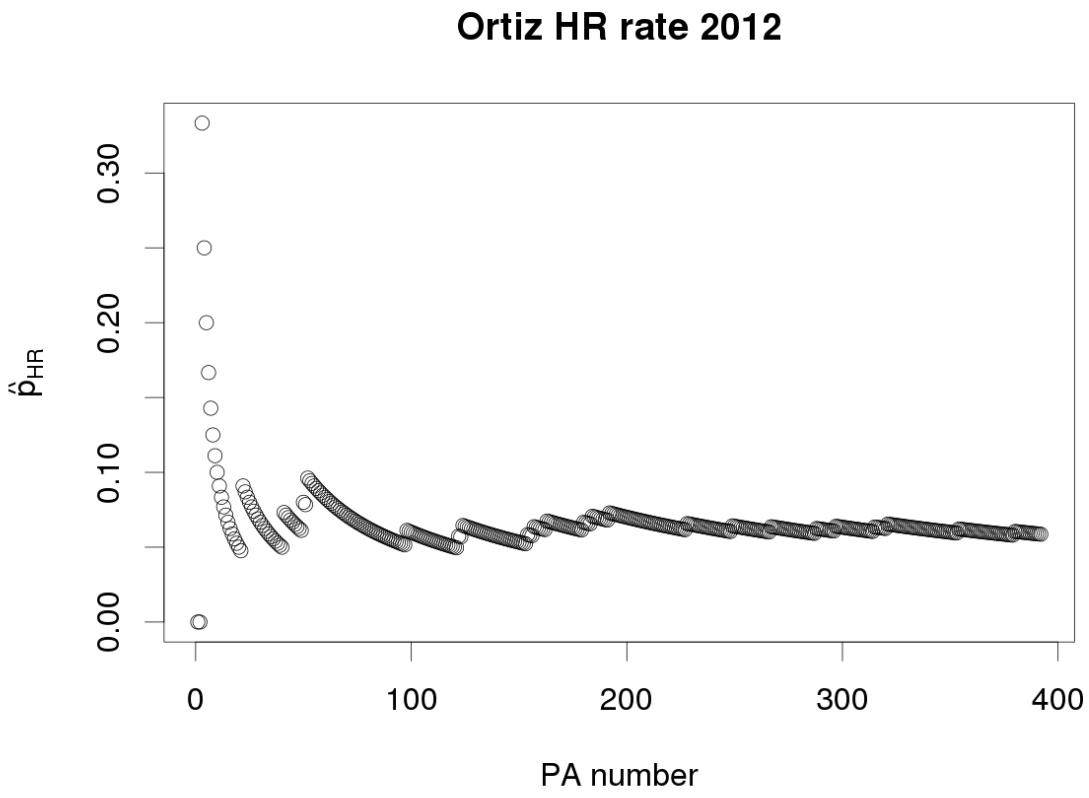
- Q: Why is this?
- A: Because $\hat{p}_{HR}$ is only based on a finite amount of data



Probability

As we collect more data $\hat{p}_{HR}$ will converge to π_{HR}

- i.e., with more data measured ***performance*** $\hat{p}_{HR}$ becomes a more accurate estimate of ***ability*** π_{HR}



True ability vs. estimated performance

The ***true ability*** of a baseball player can be viewed as some inherent property of the player

- If a player performed infinitely many times (at the same level or performance), then we could measure their true ability
- We denote ability with Greek letters, e.g., π, μ

In reality we on get to observe a player perform a finite number of times and we **estimate** the players ability based on his finite ***performance***

- We denote our estimate with Latin characters, e.g., $\hat{p}$, $\bar{x}$

The more times we observe a player, the more accurate our estimate of the players performance will be

- E.g., with more data $\hat{p}$ will get close to π

Random variables

Often we map the random events into numbers

- HR: $X = 1$
- Not a home run: $X = 0$

These random numbers are called **random variables**

- They are often denoted with capital letters such as X

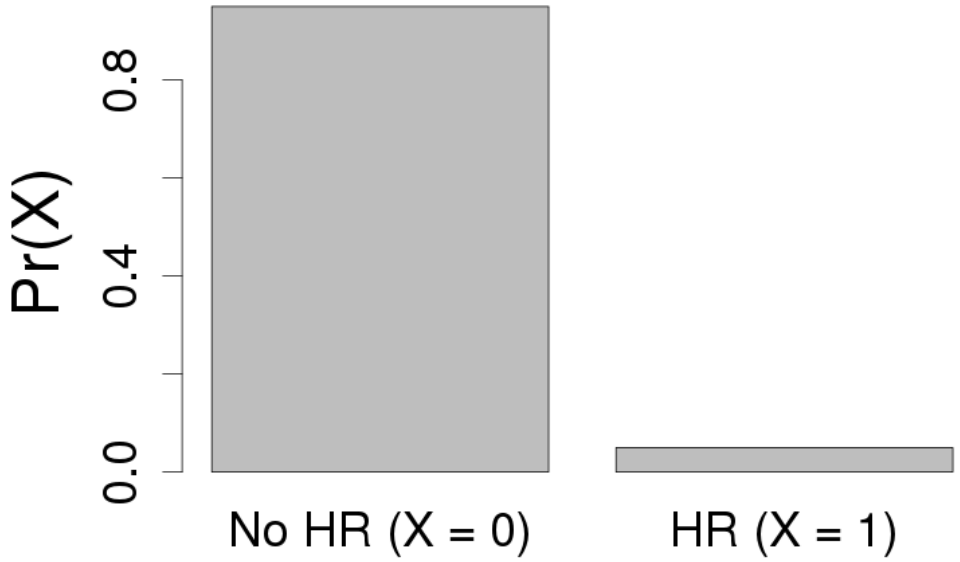
We then refer to the probability of getting a particular number

- $Pr(\text{event} = HR) = Pr(X = 1) = .05$

Parametric probability distributions

The possible outcomes are called the **sample space**

The sum of all the outcome must equal 1



Bernoulli Distribution

Example of rolling a die

Example: If we roll a fair 6-sided die:

What is the sample space?

What are the probability of the outcomes?

Roll	1	2	3	4	5	6
Probability						

Example of rolling a die

Example: If we roll a fair 6-sided die:

What is the sample space?

What are the probability of the outcomes?

Roll	1	2	3	4	5	6
Probability	1/6	1/6	1/6	1/6	1/6	1/6

Probabilities when outcomes are equally likely

When outcomes are equally likely, we have:

Probability of an event = Pr(event) =

$$\frac{\textit{Number of outcomes in the event}}{\textit{Total number of outcomes}}$$

Example: What is the probability of rolling a 1, 2 or 3?

Probability rules - Additive rule

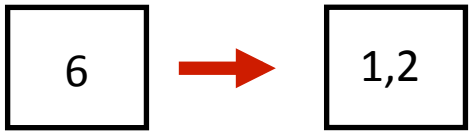
If there are two events A, and B, then the probability of A or B happening is:

$$Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A, B)$$

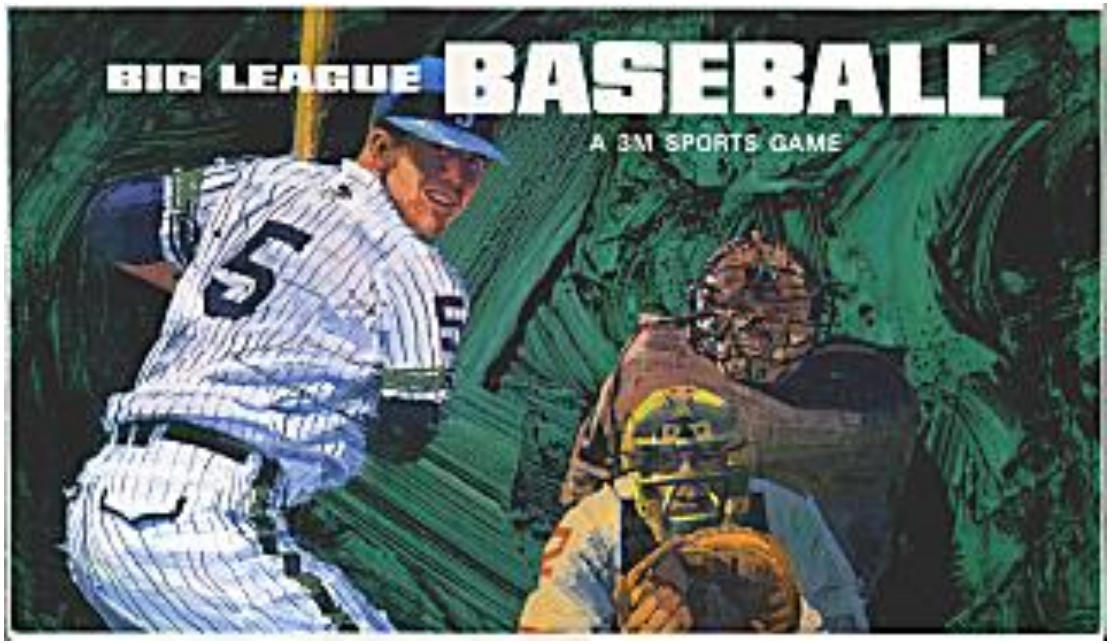
Example 1: what is the probability of rolling a 1, or 2?

Example 2:

- Suppose we replace the 6 on a die with the number 1 and 2
- What is the probability of getting a 1 or a 2?



Big League Baseball



Big league baseball was a board game from the 1960's.

Events in a baseball game were modeled as the outcome of rolling dice

Big League Baseball

Rules for Big League Baseball:

A single die is rolled:

- If a 2 or 3 occurs, a ball is pitched
- If a 4 or 5 occurs, a strike is pitched
- If a 1 or 6 occurs, a ball is hit in fair play

What is the sample space?

What is the probability of a ball being hit in fair play?

Big League Baseball

Rules for Big League Baseball:

If a fair ball is hit (1 or 6 rolled first) then two dice are rolled in sequence and the following table indicates the outcome of the play:

2nd Die

1st Die	1	2	3	4	5	6
1	Single	Out	Out	Out	Out	Error
2	Out	Double	Single	Out	Single	Out
3	Out	Single	Triple	Out	Out	Out
4	Out	Out	Out	Out	Out	Out
5	Out	Single	Out	Out	Out	Single
6	Error	Out	Out	Out	Single	Home run

Big League Baseball

If a fair ball is hit, what is the probability of a the following events:

- A home run?
- A out?
- A single?
- A Hit?

2nd Die

1st Die	1	2	3	4	5	6
1	Single	Out	Out	Out	Out	Error
2	Out	Double	Single	Out	Single	Out
3	Out	Single	Triple	Out	Out	Out
4	Out	Out	Out	Out	Out	Out
5	Out	Single	Out	Out	Out	Single
6	Error	Out	Out	Out	Single	Home run

Big League Baseball

If a fair ball is hit, what is the probability of a the following events:

- A home run? 1/36
- A out? 24/36
- A single? 7/36
- A Hit? 10/36

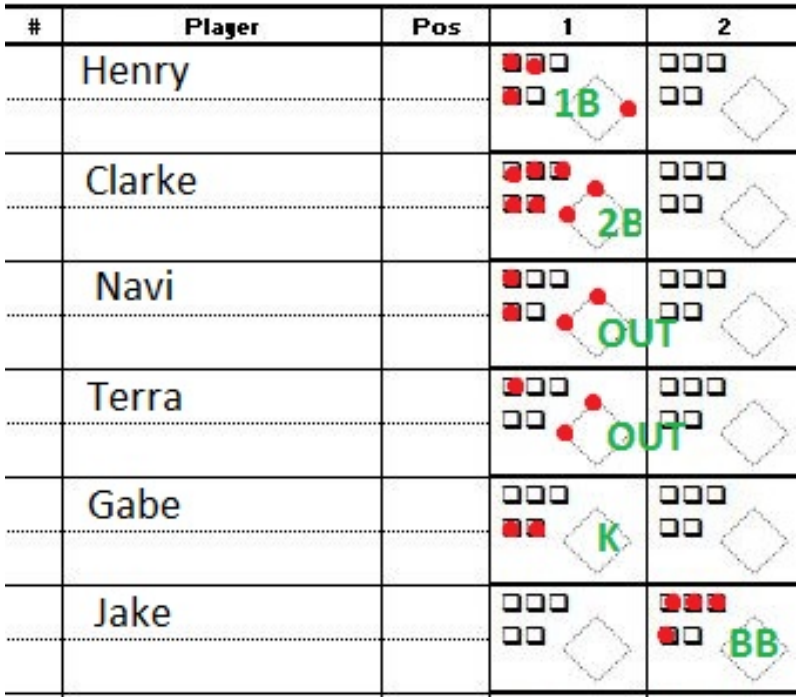
2nd Die

1st Die	1	2	3	4	5	6
1	Single	Out	Out	Out	Out	Error
2	Out	Double	Single	Out	Single	Out
3	Out	Single	Triple	Out	Out	Out
4	Out	Out	Out	Out	Out	Out
5	Out	Single	Out	Out	Out	Single
6	Error	Out	Out	Out	Single	Home run

Big League Baseball

Let's play a few innings of Big League Baseball

- How a score card works:



Big League Baseball

Let's play a few innings of Big League Baseball

If you want to contribute to class data, please enter the following statistics into the class spreadsheet

1. Total number of innings played
2. Total number of pitches (rolls of the first die)
3. Total number of strikes
4. Total number of strike outs
5. Total number of balls
6. Total number of walks
7. Total number of balls hit in play (times you rolled both dice)
8. Total number of singles
9. Total number of doubles
10. Total number of triples
11. Total number of home runs
12. Total number of errors
13. Total number of outs from balls hit in play
14. Total number of runs

You can also keep track of the game using this app:

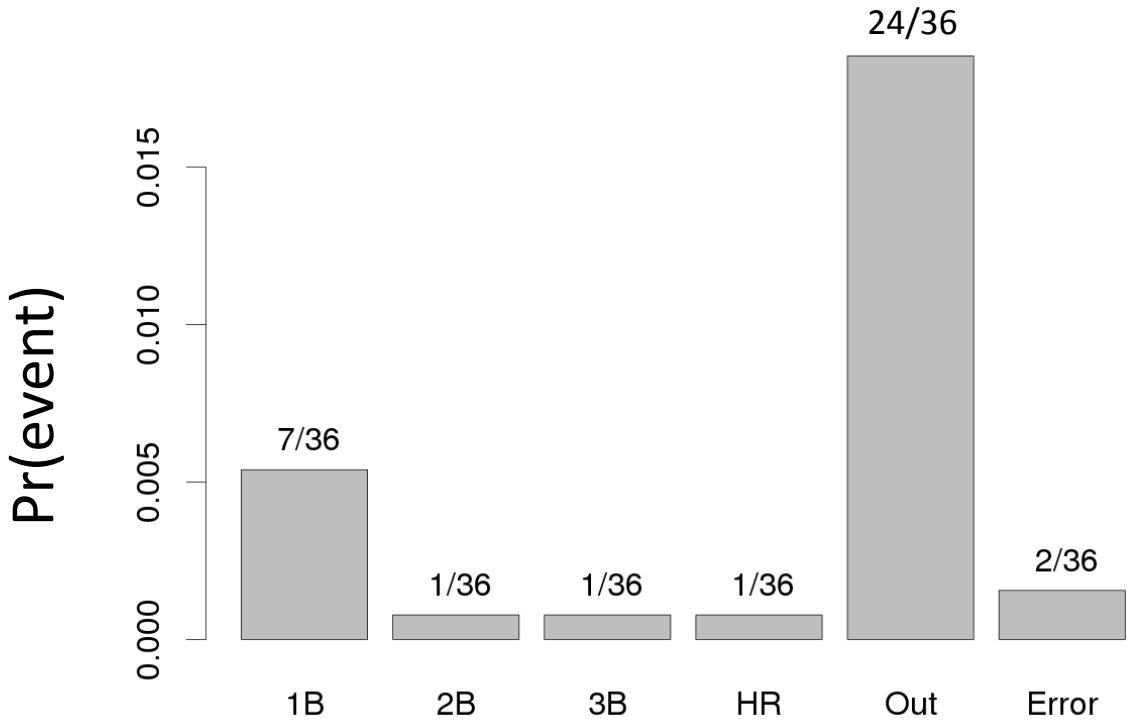
http://asterius.hampshire.edu:3838/baseball/big_league_baseball/

Limitations of Big League Baseball

A few limitations of Big League Baseball?

- All batters have the same ability
- All pitchers have the same ability
- Balls and strikes are equally likely
  - In real baseball many more strikes are thrown

Big League baseball probability distribution if a ball is hit in play



Q: why is this a valid probability distribution?

A: All values are between 0 and 1 and the sum is 1

Big League Baseball

For any one pitch (not assuming that the play is in play), what is the probability of a the following events?

- A home run?
- A out?
- A single?
- A Hit?

2nd Die

1st Die	1	2	3	4	5	6
1	Single	Out	Out	Out	Out	Error
2	Out	Double	Single	Out	Single	Out
3	Out	Single	Triple	Out	Out	Out
4	Out	Out	Out	Out	Out	Out
5	Out	Single	Out	Out	Out	Single
6	Error	Out	Out	Out	Single	Home run

Big League Baseball

For any one pitch (not assuming that the play is in play), what is the probability of a the following events?

- A home run? 1/36 * 1/3
- A out? 24/36 * 1/3
- A single? 7/36 * 1/3
- A Hit? 10/36 * 1/3

2nd Die

1st Die	1	2	3	4	5	6
1	Single	Out	Out	Out	Out	Error
2	Out	Double	Single	Out	Single	Out
3	Out	Single	Triple	Out	Out	Out
4	Out	Out	Out	Out	Out	Out
5	Out	Single	Out	Out	Out	Single
6	Error	Out	Out	Out	Single	Home run

Big League Baseball

What is the probability one would strike out on three straight pitches?

- Answer: $(1/3)^3 = 1/27$

At the start of an inning, what is the probability one would strike out on 4 pitches?

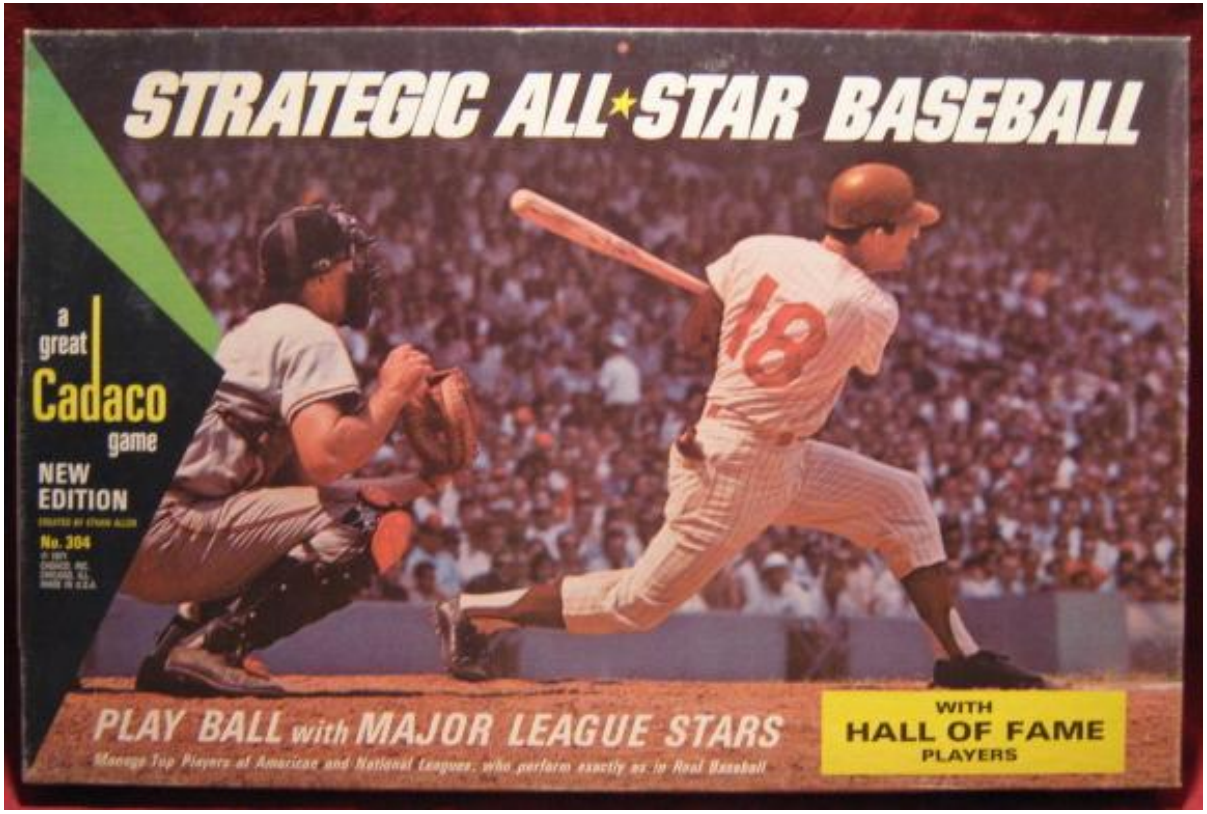
- Probability of a strike is 1/3
- Probability of a ball is 1/3
- Number of ways to strike out on 4 pitches is:
  - BSSS, SBSS, SSBS = 3
- Probability is: $3 \cdot (1/3)^4 = 1/27$

Probability through estimating the relative frequency of events

Sometimes it is not possible to mathematically calculate the probability distribution

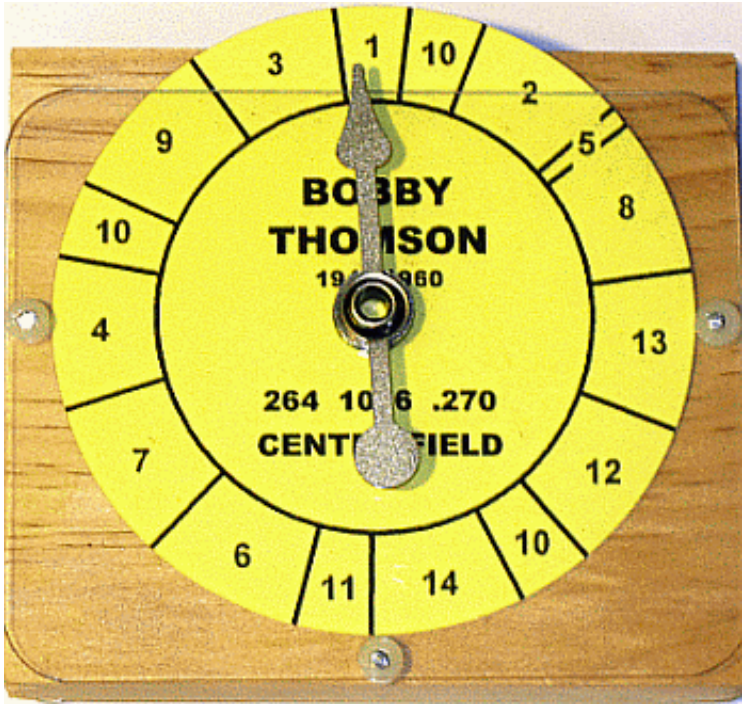
In such cases, we can estimate the probability of events by counting their relative frequencies

All-star baseball



Game that was popular in the 1960’s and 70’s

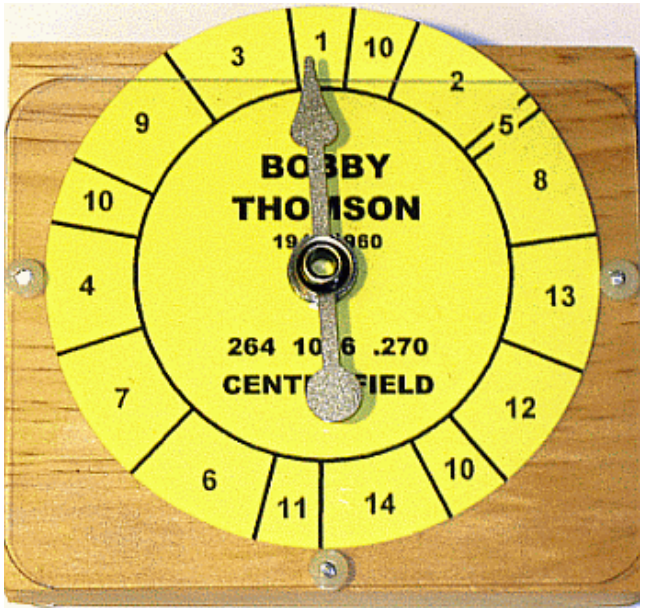
All-star baseball



Each player represented by a spinner

- 1 = HR; 2, 13 = 1B, 11 = 2B, etc.

All-star baseball



For all-star baseball is it not possible to analytically calculate the probability of different events

- Instead we would just have to repeat the event many times can calculate the relative frequencies

Let's try to implement this game in R!

Functions in R:

```
source('/home/shared/baseball_stats_2017/spinner_simulation.R')

generate_play(firstName, lastName, year)

generate_play("David", "Ortiz", 2015)
```

Record your game:
http://asterius.hampshire.edu:3838/baseball/big_league_baseball/

Limitations of All-star Baseball?

MLB 15

Evolution of baseball video games

Next class

More probability and games

Quote and Reaction to Big Data Baseball chapter 7