Probability and simulations using games part I

Overview

Lab 3 discussion

Discussion of chapter 3 of Astroball

Discussion of probability

Big League Baseball

Simulating big league baseball

Object oriented programming

Lab 3: questions?

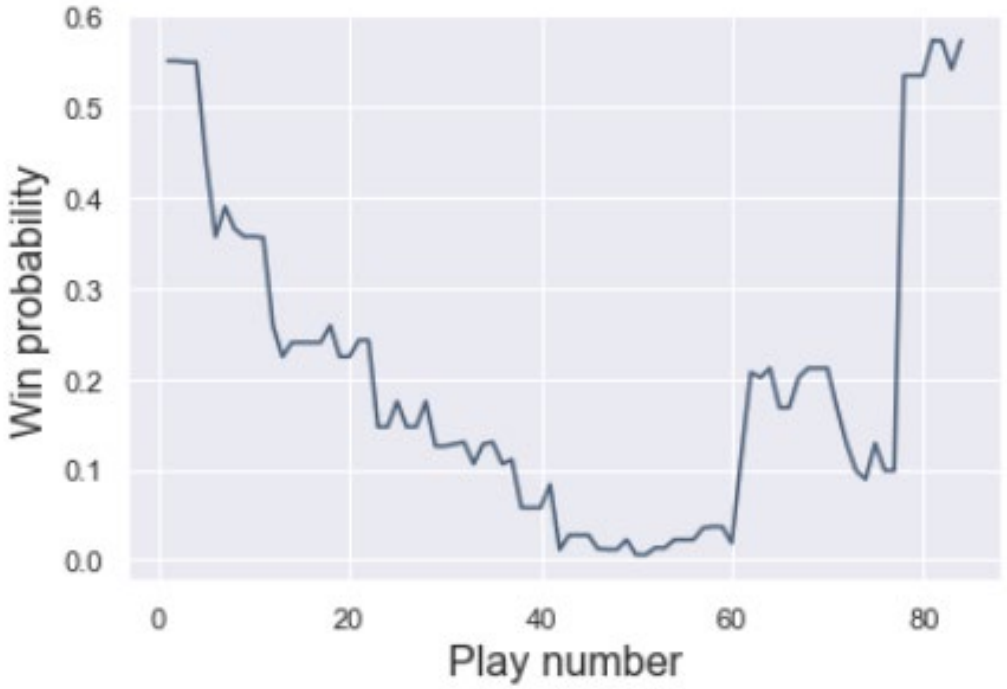
How did it go?

Please be sure to mark your pages on Gradescope

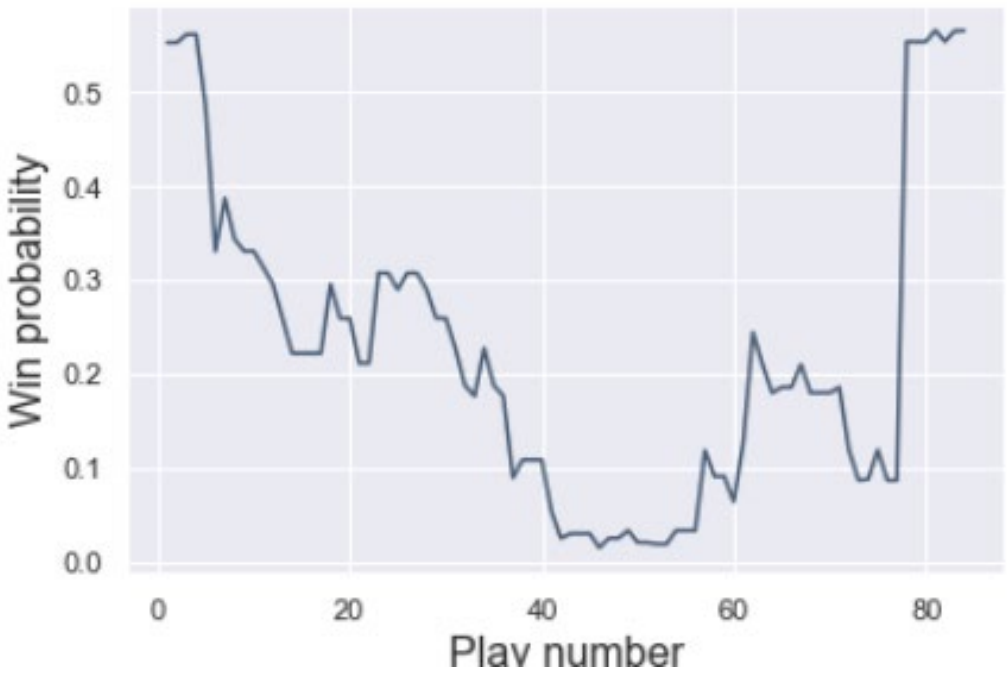
- A few points will be deducted if pages are not marked

Lab 3: questions?

Using data from 2019
to get probabilities



Using data from 1988
to get probabilities



Probability of a come back from down by 6 in the 6th using data from 1988 is: 3.28% vs. 2.2% for 2019

Astroball discussion

Let’s discuss the chapter for 6 minutes in breakout rooms and then have a larger conversation as a group

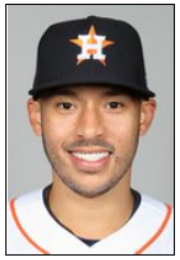
- Discuss your quote and reaction to chapter 3

Thoughts on the chapter 3 of Astroball?

Drafting Correa 2012

Correa projected statistics, pg 70

- BA: .272-.287
- HR: 35-40 HR



Carlos Correa

Position: Shortstop
Bats: Right • **Throws:** Right
6-4, 220lb (193cm, 99kg)
Team: Houston Astros (majors)

More bio, uniform, draft, salary info ▾

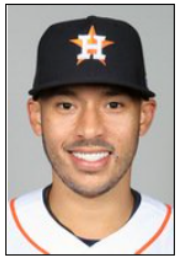
SUMMARY	WAR	AB	H	HR	BA	R	RBI	SB	OBP	SLG	OPS	OPS+
2020	1.8	201	53	5	.264	22	25	0	.326	.383	.709	92
Career	26.3	2269	626	107	.276	334	397	33	.353	.480	.833	126

Year	Age	Tm	Lg	G	PA	AB	R	H	2B	3B	HR	RBI	SB	CS	BB	SO	BA	OBP	SLG	OPS	OPS+	TB	GDP	HBP	SH	SF	IBB	Pos	Awards
2015	20	HOU	AL	99	432	387	52	108	22	1	22	68	14	4	40	78	.279	.345	.512	.857	135	198	10	1	0	4	2	6	MVP-24,**RoY-1**
2016	21	HOU	AL	153	660	577	76	158	36	3	20	96	13	3	75	139	.274	.361	.451	.811	124	260	12	5	0	3	5	*6/H	
2017 ★	22	HOU	AL	109	481	422	82	133	25	1	24	84	2	1	53	92	.315	.391	.550	.941	155	232	12	2	0	4	5	6/D	AS,MVP-17
2018	23	HOU	AL	110	468	402	60	96	20	1	15	65	3	0	53	111	.239	.323	.405	.728	99	163	17	2	0	***11***	3	*6/D	
2019	24	HOU	AL	75	321	280	42	78	16	1	21	59	1	0	35	75	.279	.358	.568	.926	137	159	8	2	0	4	0	6	
2020	25	HOU	AL	58	221	201	22	53	9	0	5	25	0	0	16	49	.264	.326	.383	.709	92	77	4	3	0	1	2	*6/H	
6 Yrs				**604**	**2583**	**2269**	**334**	**626**	**128**	**7**	**107**	**397**	**33**	**8**	**272**	**544**	**.276**	**.353**	**.480**	**.833**	**126**	**1089**	**63**	**15**	**0**	**27**	**17**		
162 Game Avg.				**162**	**693**	**609**	**90**	**168**	**34**	**2**	**29**	**106**	**9**	**2**	**73**	**146**	**.276**	**.353**	**.480**	**.833**	**126**	**292**	**17**	**4**	**0**	**7**	**5**		

Drafting Correa 2012

Correa projected statistics, pg 70

- BA: .272-.287
- HR: 35-40 HR



Carlos Correa

Position: Shortstop
Bats: Right • **Throws:** Right
6-4, 220lb (193cm, 99kg)
Team: Houston Astros (majors)

More bio, uniform, draft, salary info ▾

SUMMARY	WAR	AB	H	HR	BA	R	RBI	SB	OBP	SLG	OPS	OPS+
2020	1.8	201	53	5	.264	22	25	0	.326	.383	.709	92
Career	26.3	2269	626	107	.276	334	397	33	.353	.480	.833	126

Byron Buxton

Position: Centerfielder
Bats: Right • **Throws:** Right
6-2, 190lb (188cm, 86kg)
Team: Minnesota Twins (majors)

More bio, uniform, draft, salary info ▾

SUMMARY	WAR	AB	H	HR	BA	R	RBI	SB	OBP	SLG	OPS	OPS+
2020	2.1	130	33	13	.254	19	27	2	.267	.577	.844	124
Career	11.9	1380	329	51	.238	204	172	62	.289	.430	.719	91

Kevin Gausman

Position: Pitcher
Bats: Left • **Throws:** Right
6-2, 190lb (188cm, 86kg)
Team: San Francisco Giants (majors)

More bio, uniform, draft, salary info ▾

SUMMARY	WAR	W	L	ERA	G	GS	SV	IP	SO	WHIP
2020	1.3	3	3	3.62	12	10	0	59.2	79	1.106
Career	11.2	50	66	4.26	203	164	0	985.1	934	1.329

Drafting Correa 2012

More info on the 2012 draft

- Wikipedia
- Baseball reference

Enos Cabell

Positions: Third Baseman, First Baseman and Outfielder
Bats: Right • **Throws:** Right
6-4, 170lb (193cm, 77kg)
Born: October 8, 1949 (Age: 71-142d) in Fort Riley, KS

More bio, uniform, draft, salary info ▾

SUMMARY	WAR	AB	H	HR	BA	R	RBI	SB	OBP	SLG	OPS	OPS+
Career	11.3	5952	1647	60	.277	753	596	238	.308	.370	.678	93

Byron Buxton

Position: Centerfielder
Bats: Right • **Throws:** Right
6-2, 190lb (188cm, 86kg)
Team: Minnesota Twins (majors)

More bio, uniform, draft, salary info ▾

SUMMARY	WAR	AB	H	HR	BA	R	RBI	SB	OBP	SLG	OPS	OPS+
2020	2.1	130	33	13	.254	19	27	2	.267	.577	.844	124
Career	11.9	1380	329	51	.238	204	172	62	.289	.430	.719	91

Kevin Gausman

Position: Pitcher
Bats: Left • **Throws:** Right
6-2, 190lb (188cm, 86kg)
Team: San Francisco Giants (majors)

More bio, uniform, draft, salary info ▾

SUMMARY	WAR	W	L	ERA	G	GS	SV	IP	SO	WHIP
2020	1.3	3	3	3.62	12	10	0	59.2	79	1.106
Career	11.2	50	66	4.26	203	164	0	985.1	934	1.329

Trusting a model…

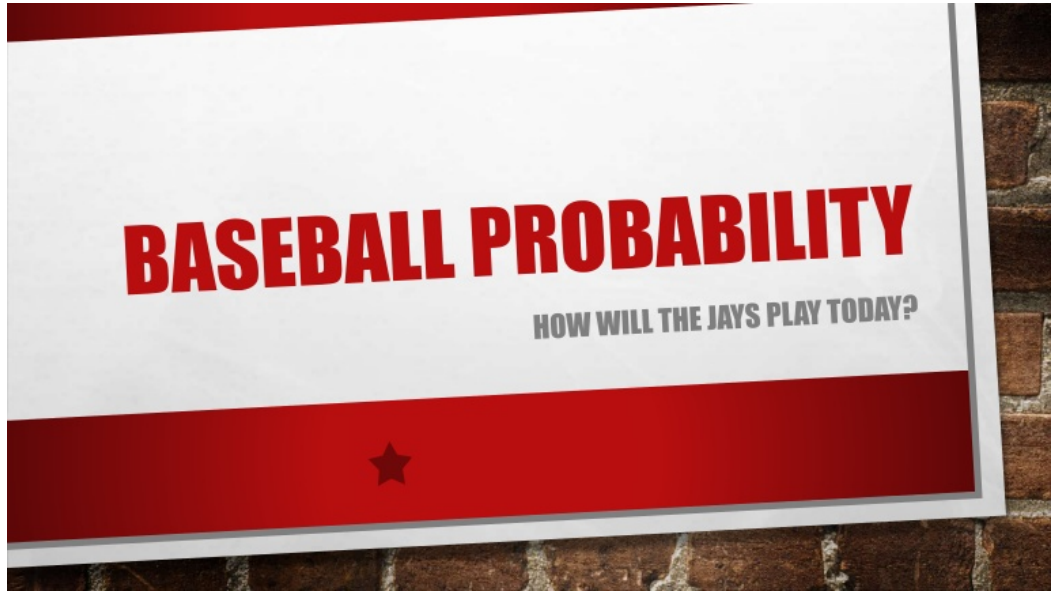
"It's one thing to get somebody with college-level statistics to create a model" Sig said. "It's another for the decision-makers to really use it. I don't mean use it as a tiebreaker, or to throw the analyst a bone late in the draft. I mean really *use* it, from the first pick. That's what Jeff did."

Probability

Probability is a way of measuring the likelihood that an event will occur

Probability models can be used to:

- Determine the likelihood of an event
- Simulate events



Probability

Probability models 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 Mookie Betts 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 2021 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 his first plate appearance this season

Probability

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

If we repeat an experiment N times, we can get an estimate of the probability:

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

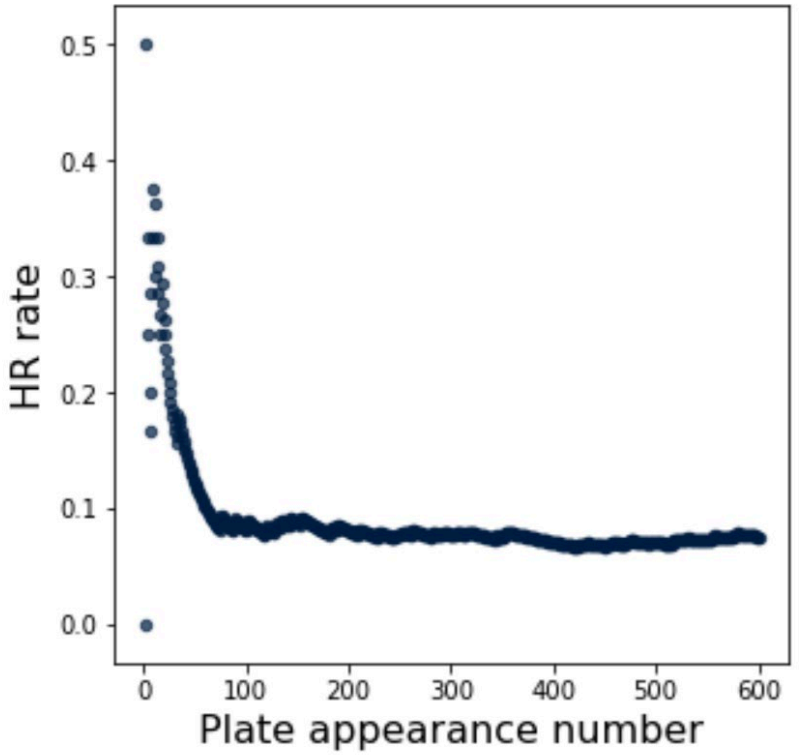
If we repeated this infinitely many times, we would get the true probability of the event

- i.e., N → ∞

Probability

For example, we can estimate the probability that Mike Trout will hit a home.

$$Pr(HR) \approx \frac{Number\ of\ HR}{Number\ of\ PA}$$



Random variables

Often we map the random outcomes 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

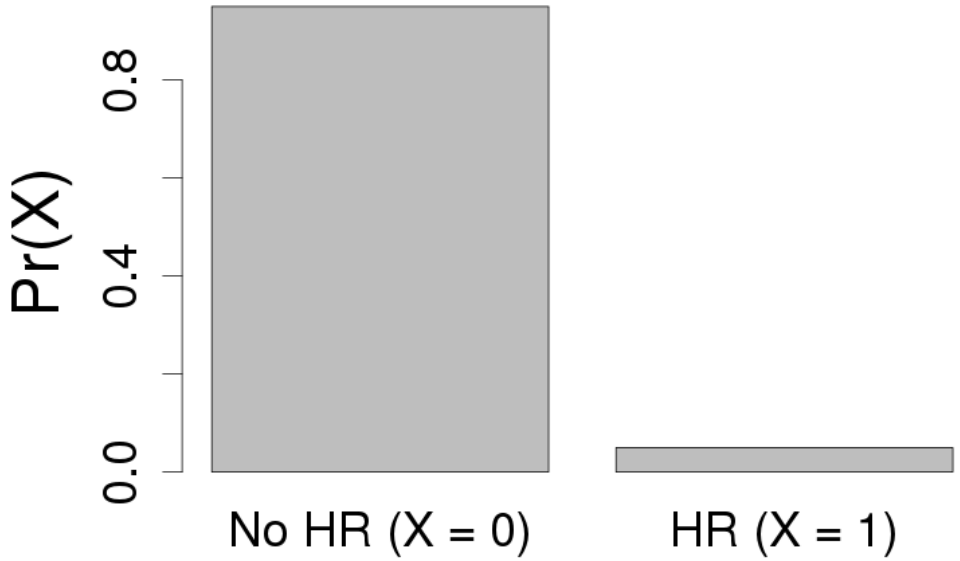
We then refer to the probability of getting a particular number

- Pr(HR) = Pr(X = 1) = .10

Parametric probability distributions

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

The sum of the probability measures over all outcomes 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: X	1	2	3	4	5	6
Probability: Pr(X)						

Example of rolling a die

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

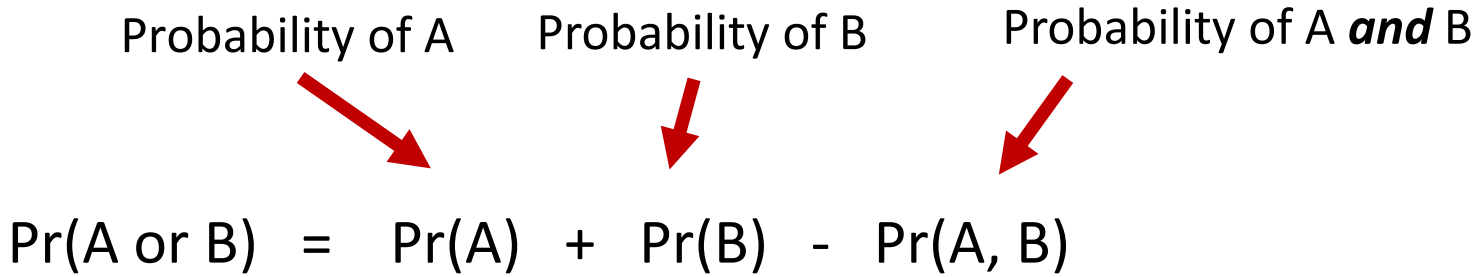
What is the sample space?

What are the probability of the outcomes?

Roll: X	1	2	3	4	5	6
Probability: Pr(X)	1/6	1/6	1/6	1/6	1/6	1/6

Probability rules - Additive rule

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



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

- Pr(A) = Pr(rolling X = 1) = 1/6
- Pr(B) = Pr(rolling X = 2) = 1/6
- Pr(A and B) = Pr(rolling X = 1 and 2) = 0

Probability rules - Additive rule

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

Probability of A Probability of B Probability of A ***and*** B

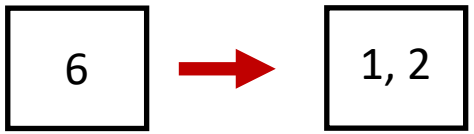
$$Pr(A \text{ or } B) = Pr(A) + Pr(B) - Pr(A, B)$$

Example 2:

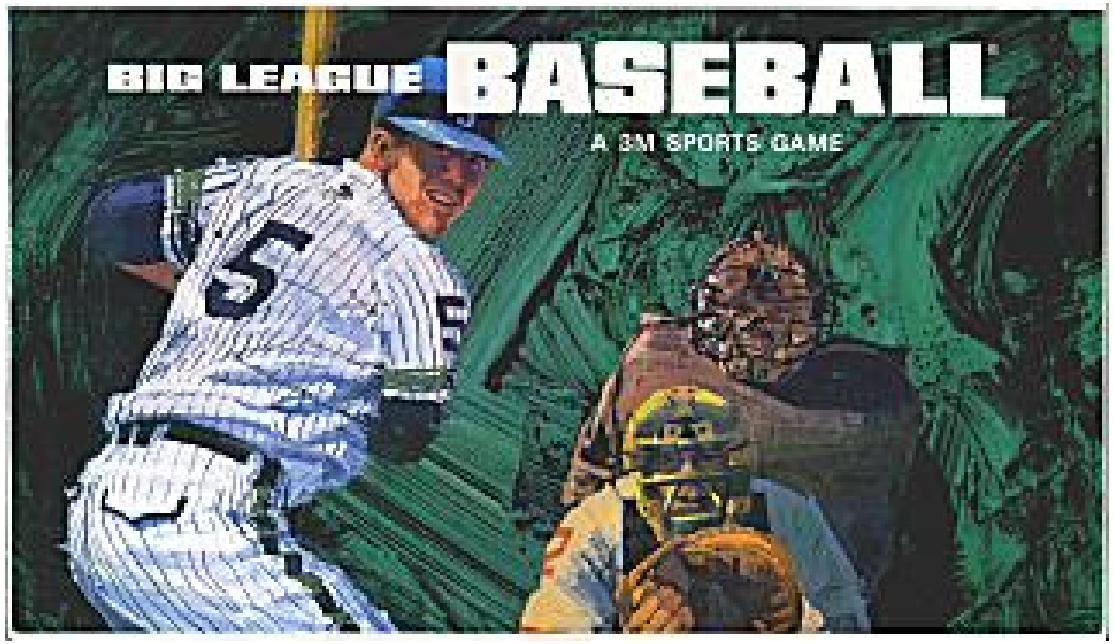
- Suppose we replace the 6 on a die with the number 1 and 2
- What is the probability of getting a 1 or a 2?

$$Pr(X = 1) + Pr(X = 2) - Pr(X = 1 \text{ and } 2)$$

$$2/6 + 2/6 - 1/6 = 3/6 = 1/2$$



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 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 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 an inning of Big League Baseball

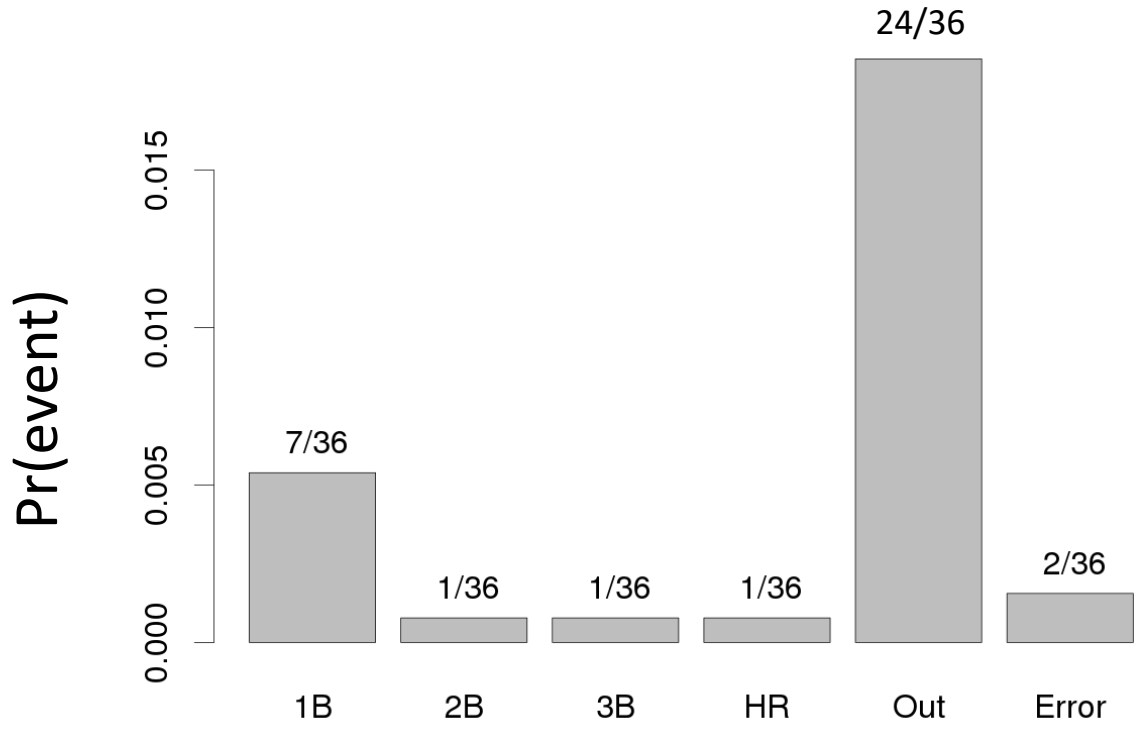
- Keep score with the scorecard: https://baseball-scorecard.pdffiller.com/
- Dice rolling simulator: https://www.random.org/dice/
- Rules of Big League Baseball and a pdf of a scorecard are also on Canvas

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

Suppose we want to calculate the probability of a home run, ***without assuming that a ball was hit in play***

What are the sequences of events that would lead to a home run?

1) We need to hit a ball in play (role a 1 or a 6)
   - Pr(ball in play) = 1/3

2) We need to hit a home run (role two 6's)
   - Pr(home run | ball in play) = 1/36

We somehow need to combine these events to get a total probability

Multiplicative Rule

The probability of two events A ***and*** B occurring is:

- The probability of A occurring given that B has occurred… times …
- The probability of B occurring

$$\Pr(A, B) = \Pr(A|B) \times \Pr(B)$$

E.g., Suppose we draw 2 cards from a 52 card deck. What is the probability they are both diamonds?

$$\Pr(D1, D2) = \Pr(D1) \times \Pr(D2 \mid D1)$$
$$= 13/52 \times 12/51 = 0.12$$

Special Case: Multiplicative Rule for independent events

Two events are independent if:

$$Pr(A, B) = Pr(A) \times Pr(B)$$

Q: what is the probability of getting two strikes in a row in Big League baseball at the start of a plate appearance?

$$Pr(\text{Strike, Strike}) = Pr(\text{Strike}) \times Pr(\text{Strike})$$

$$= 1/3 \times 1/3 = 1/9$$

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:
  - B,S,S,S or
  - S,B,S,S or
  - S,S,B,S
- Probability is: $(1/3)^4 + (1/3)^4 + (1/3)^4 = 1/27$

Simulating Big League Baseball

Simulating rolls of the dice in Python:

```
import numpy as np

np.random.choice(np.arange(1, 7))
```

Simulating Big League Baseball

Conditional statements:

```
if (X == 1) or (Y == 2):
    # do stuff

elif (X == 3) or (Y == 4):
    # do different stuff

else:
    # do this stuff
```

Simulating Big League Baseball

A dictionary maps a set of “keys” to a set of “values”.

There are a few different ways to create dictionaries in Python, but for our purposes we will use this syntax:

```
my_dictionary = {}
my_dictionary['key1'] = 'value1'
my_dictionary['key2'] = 'value2'
…
```

Simulating Big League Baseball

A dictionary maps a set of “keys” to a set of “values”.

There are a few different ways to create dictionaries in Python, but for our purposes we will use this syntax:

```
my_dictionary = {}
my_dictionary[(v1, v2)]= 'value1'
my_dictionary['key2'] = 'value2'
…
```

Simulating Big League Baseball

Let’s take a quick look at Lab 4, Exercise 2 where you will generate plays using the rules of Big Data Baseball…

Keeping track of the state of a game

If we want to do a full simulation of a Big League Baseball game it will be useful to keep track of the game after every event

We can use Object-oriented programming (OOP) to do this!

Object-oriented programming

Object-oriented programming (OOP) is a programming paradigm based on the concept of "objects", which can contain:

1. data: in the form of fields
   - often known as attributes or properties

2. methods: functions that operate on the data

Have we seen any objects in this class yet?

```
my_table = Table()                    # constructor
my_table.select("my_column")    # method
```

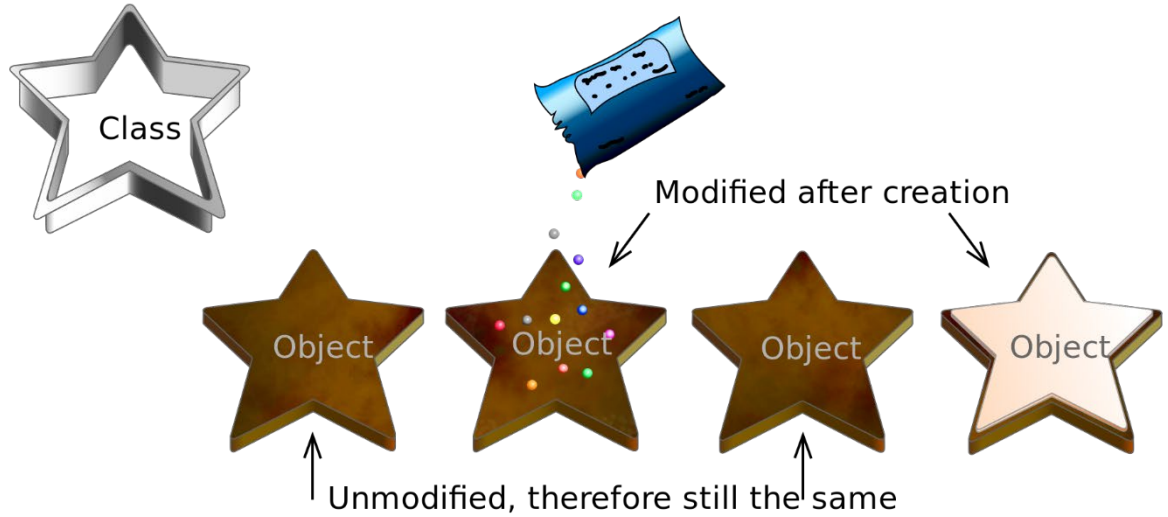
Object-oriented programming

A **class** defines what the object is

- i.e., classes have code that lists the types of data stored and the methods

An **object** is a realization of classes (often called an instance of a class)

- Stores actual data that can be manipulated with methods
- E.g., my_table.where()



Object-oriented programming

A **constructor** defines what should happen when an object is created

- In Python the constructor is defined with the **__init__(self)** method

```
class Baseball_Game:

    # Constructor
    def __init__(self):

        self.end_of_game = False
        self.inning = 1
```

```
# create an instance of an object
my_object = Baseball_Game()
```

Object-oriented programming

The __str__ method defines what should happen when the print() function is called on the object

```
# str method
def __str__(self):
    'I am an object’
```

```
# create an instance of an object
my_object = Baseball_Game()

# print the object
print(my_object)
```

Object-oriented programming

Let’s begin on Exercise 3 where you will create an object to keep track of the state of the baseball game

For exercise 3, spend 1 hour on it

- If you are totally stuck, I will send you the answer so you can go on to part 4