Tree diagrams and the binomial distribution

Outline for today

Better know a player: Honus Wagner

Questions about worksheet 6?

Review of concepts in probability: additive and multiplicative rules

Tree diagrams and the analysis of Strat-o-matic

The binomial distribution

Announcement: class final projects

Final project proposal due on Wednesday

Should focus on research question

- Doesn't have to be about baseball, but need to find data set that you can use to answer the question

1-2 paragraph proposal is due the Wednesday March 29th

Projects presentation are on May 3rd – good to start working on them soon!

Better know a player: Honus Wagner

Probability

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

Definitions:

Sample space

- All possible outcomes

An Event

- Subset of the sample space

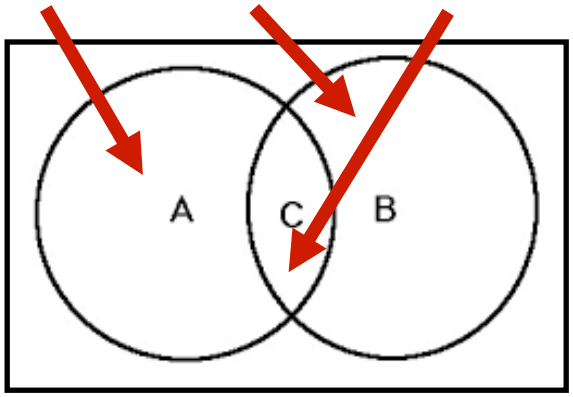
Probability key properties:

- $0 \leq \Pr(X) \leq 1$
- $\Sigma \Pr(X = x) = 1$

Probability rules - Additive rule

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

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

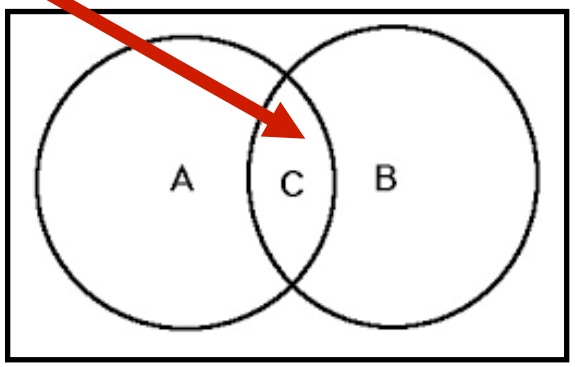


Events are called **mutually exclusive** if events A and B can not both occur - i.e., Pr(A, B) = 0

Q: What would mutually exclusive events look like in the Venn diagram?
A: The circles would not overlap

Multiplicative Rule

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



Probability of B happening

×

Probability of A happening given B happened

Two events are independent if:

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

i.e., if the occurrence of B does not effect the probability of A happening

Big League Baseball

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

- Probability of a single: $1/3 \cdot 7/36$ +
- Probability of a double: $1/3 \cdot 1/36$ +
- Probability of a triple: $1/3 \cdot 1/36$ +
- Probability of a home run: $1/3 \cdot 1/36$ = 10/108

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

Strat-o-matic

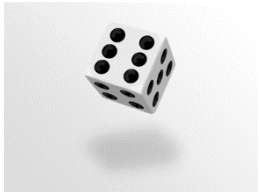
Much more complex board games

- Takes into account Hitters and Pitchers
- Advanced version accounts for additional factors (e.g., ball parks etc.)

Strat-o-matic rules and tree diagrams

Each player is represented by a card

1. A white single die is rolled to determine whether to use hitter or pitcher's card:
   - 1-3 -> hitters card
   - 4-6 -> pitcher's card



2. Then, two dice are rolled and their sum determines which play in the card should be used

3. For some plays additionally a 20 sided determine the final outcome
   - and other tables/rules often need to be c

Strat-o-matic: analysis

Let's calculate the probability of different events…

Calculating the probability of getting a particular column pitcher or hitter's card is pretty simple

- Answer?

Calculating the sum of the two dice is a little more involved…

- What is the sample space here?
  - i.e., how many possible outcomes are there?
- Can you calculate the probability distribution?

Strat-o-matic: analysis

Fill in the table below with the sum of the two dice and then calculate the probability of rolling a 2 to a 12

2nd Die

1st Die	**1**	**2**	**3**	**4**	**5**	**6**
1						
2						
3						
4			7			
5						
6						

Strat-o-matic: analysis

Fill in the table below with the sum of the two dice and then calculate the probability of rolling a 2 to a 12

1st Die \ 2nd Die	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Strat-o-matic: analysis

2	3	4	5	6	7	8	9	10	11	12
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

2nd Die

1st Die	**1**	**2**	**3**	**4**	**5**	**6**
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Strat-o-matic: analysis

2	3	4	5	6	7	8	9	10	11	12
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

What is the probability of rolling a 1 on the white die **and then** getting a sum of 7 on the two red dice?

- $1/6 \cdot 6/36 = 6/216$

What is the probability of rolling a:

- 2 on the white die…. **and then** getting…
- Sum of 8 on the two red dice… **and then…**
- A number for 1-8 on the 20 sided die?
  - $1/6 \cdot 5/36 \cdot 8/20 = .00926$

Strat-o-matic: analysis

2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

What is the probability of rolling a 1 on the white die **or** a 6 on the white die?

- 1/6 + 1/6 = 2/6

What is the probability of rolling a:

- 5 on the white die… **and then** getting …
- a sum of 8 **or** a sum of 10 on the two red dice?
- $1/6 \cdot (4/36 + 3/36) = .0324$

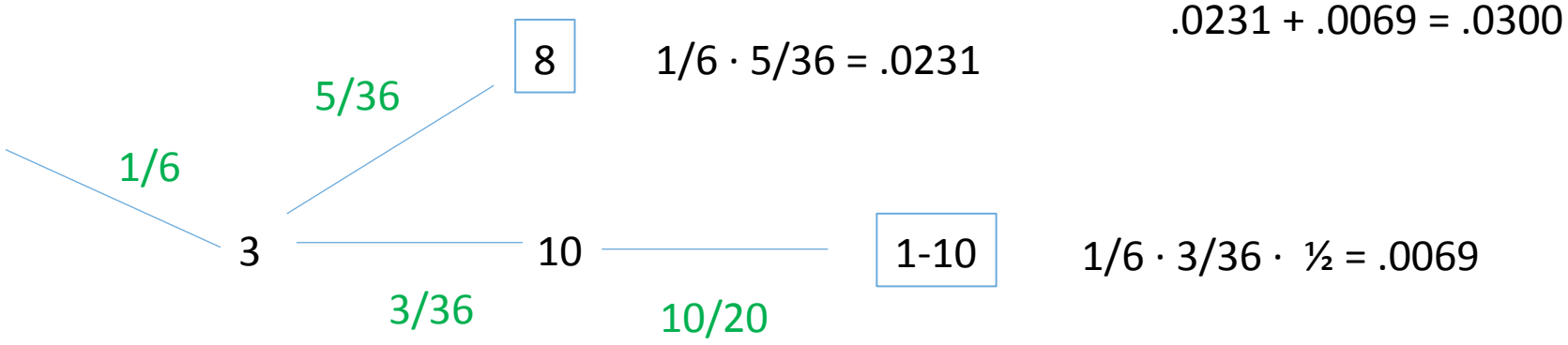
Strat-o-matic: analysis

2	3	4	5	6	7	8	9	10	11	12
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

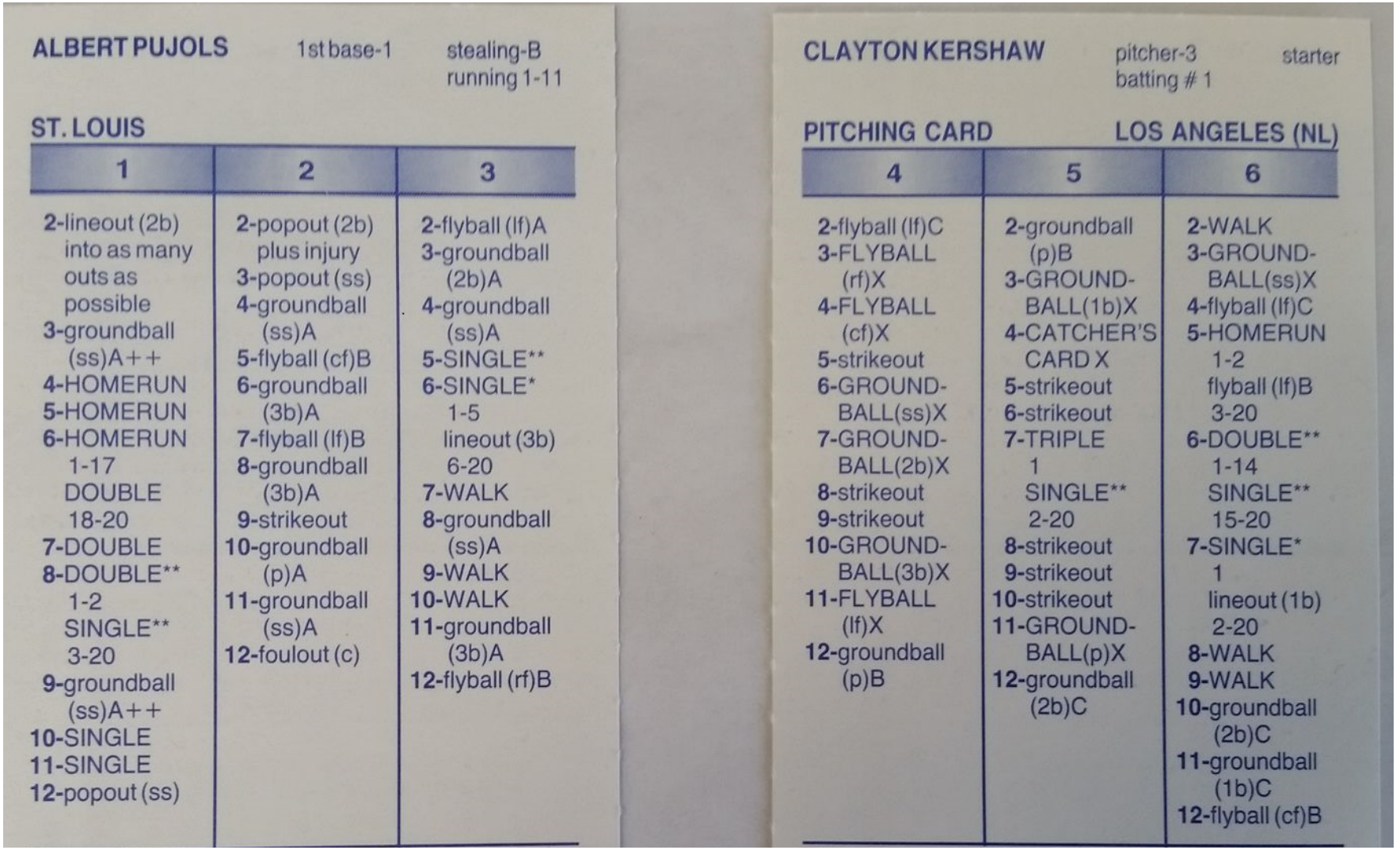
What is the probability of rolling a 3 on the white die

- and then getting a sum of 8 two red dice
- or a sum of 10 on the two red dice
  - and then a 1-10 on the 20 sided die?

Tree diagram!



Strat-o-matic: Pujols vs. Kershaw



ALBERT PUJOLS 1st base-1 stealing-B running 1-11

ST. LOUIS

1	2	3
2-lineout (2b) into as many outs as possible 3-groundball (ss)A++ 4-HOMERUN 5-HOMERUN 6-HOMERUN 1-17 DOUBLE 18-20 7-DOUBLE 8-DOUBLE** 1-2 SINGLE** 3-20 9-groundball (ss)A++ 10-SINGLE 11-SINGLE 12-popout (ss)	2-popout (2b) plus injury 3-popout (ss) 4-groundball (ss)A 5-flyball (cf)B 6-groundball (3b)A 7-flyball (lf)B 8-groundball (3b)A 9-strikeout 10-groundball (p)A 11-groundball (ss)A 12-foulout (c)	2-flyball (lf)A 3-groundball (2b)A 4-groundball (ss)A 5-SINGLE** 6-SINGLE* 1-5 lineout (3b) 6-20 7-WALK 8-groundball (ss)A 9-WALK 10-WALK 11-groundball (3b)A 12-flyball (rf)B

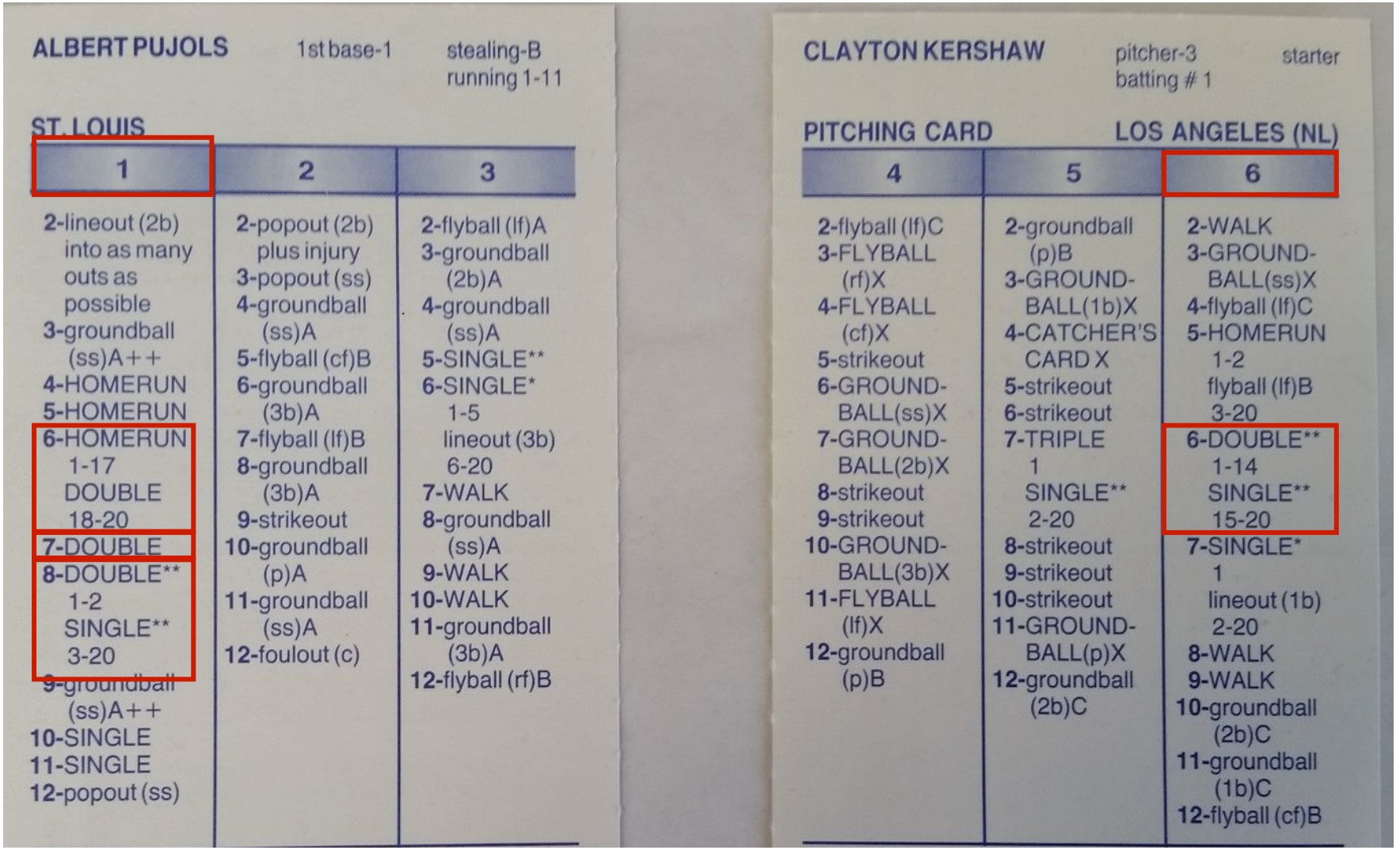
CLAYTON KERSHAW pitcher-3 batting # 1 starter

PITCHING CARD **LOS ANGELES (NL)**

4	5	6
2-flyball (lf)C 3-FLYBALL (rf)X 4-FLYBALL (cf)X 5-strikeout 6-GROUND-BALL(ss)X 7-GROUND-BALL(2b)X 8-strikeout 9-strikeout 10-GROUND-BALL(3b)X 11-FLYBALL (lf)X 12-groundball (p)B	2-groundball (p)B 3-GROUND-BALL(1b)X 4-CATCHER'S CARD X 5-strikeout 6-strikeout 7-TRIPLE 1 SINGLE** 2-20 8-strikeout 9-strikeout 10-strikeout 11-GROUND-BALL(p)X 12-groundball (2b)C	2-WALK 3-GROUND-BALL(ss)X 4-flyball (lf)C 5-HOMERUN 1-2 flyball (lf)B 3-20 6-DOUBLE** 1-14 SINGLE** 15-20 7-SINGLE* 1 lineout (1b) 2-20 8-WALK 9-WALK 10-groundball (2b)C 11-groundball (1b)C 12-flyball (cf)B

What is the probability of a hitting a double?

Strat-o-matic: Pujols vs. Kershaw



ALBERT PUJOLS 1st base-1 stealing-B running 1-11

ST. LOUIS

1	2	3
2-lineout (2b) into as many outs as possible 3-groundball (ss)A++ 4-HOMERUN 5-HOMERUN 6-HOMERUN 1-17 DOUBLE 18-20 7-DOUBLE 8-DOUBLE** 1-2 SINGLE** 3-20 9-groundball (ss)A++ 10-SINGLE 11-SINGLE 12-popout (ss)	2-popout (2b) plus injury 3-popout (ss) 4-groundball (ss)A 5-flyball (cf)B 6-groundball (3b)A 7-flyball (lf)B 8-groundball (3b)A 9-strikeout 10-groundball (p)A 11-groundball (ss)A 12-foulout (c)	2-flyball (lf)A 3-groundball (2b)A 4-groundball (ss)A 5-SINGLE** 6-SINGLE* 1-5 lineout (3b) 6-20 7-WALK 8-groundball (ss)A 9-WALK 10-WALK 11-groundball (3b)A 12-flyball (rf)B

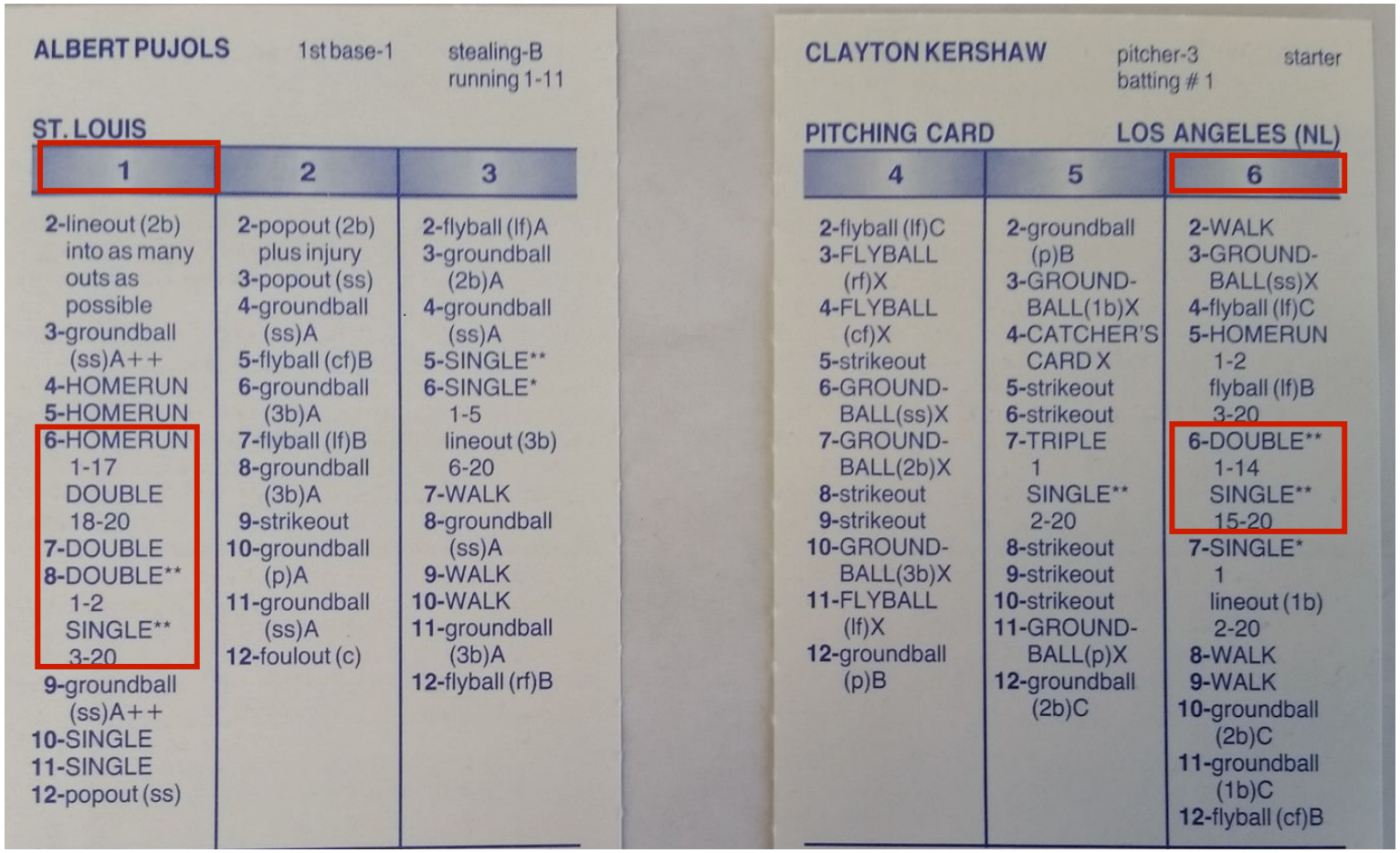
CLAYTON KERSHAW pitcher-3 batting # 1 starter

PITCHING CARD LOS ANGELES (NL)

4	5	6
2-flyball (lf)C 3-FLYBALL (rf)X 4-FLYBALL (cf)X 5-strikeout 6-GROUND-BALL(ss)X 7-GROUND-BALL(2b)X 8-strikeout 9-strikeout 10-GROUND-BALL(3b)X 11-FLYBALL (lf)X 12-groundball (p)B	2-groundball (p)B 3-GROUND-BALL(1b)X 4-CATCHER'S CARD X 5-strikeout 6-strikeout 7-TRIPLE 1 SINGLE** 2-20 8-strikeout 9-strikeout 10-strikeout 11-GROUND-BALL(p)X 12-groundball (2b)C	2-WALK 3-GROUND-BALL(ss)X 4-flyball (lf)C 5-HOMERUN 1-2 flyball (lf)B 3-20 6-DOUBLE** 1-14 SINGLE** 15-20 7-SINGLE* 1 lineout (1b) 2-20 8-WALK 9-WALK 10-groundball (2b)C 11-groundball (1b)C 12-flyball (cf)B

What is the probability of a hitting a double?

Strat-o-matic: Pujols vs. Kershaw



ALBERT PUJOLS 1st base-1 stealing-B running 1-11

ST. LOUIS

1	2	3
2-lineout (2b) into as many outs as possible	2-popout (2b) plus injury	2-flyball (lf)A
3-groundball (ss)A++	3-popout (ss)	3-groundball (2b)A
4-HOMERUN	4-groundball (ss)A	4-groundball (ss)A
5-HOMERUN	5-flyball (cf)B	5-SINGLE**
6-HOMERUN 1-17 DOUBLE 18-20	6-groundball (3b)A	6-SINGLE* 1-5 lineout (3b) 6-20
7-DOUBLE	7-flyball (lf)B	7-WALK
8-DOUBLE** 1-2 SINGLE** 3-20	8-groundball (3b)A	8-groundball (ss)A
9-groundball (ss)A++	9-strikeout	9-WALK
10-SINGLE	10-groundball (p)A	10-WALK
11-SINGLE	11-groundball (ss)A	11-groundball (3b)A
12-popout (ss)	12-foulout (c)	12-flyball (rf)B

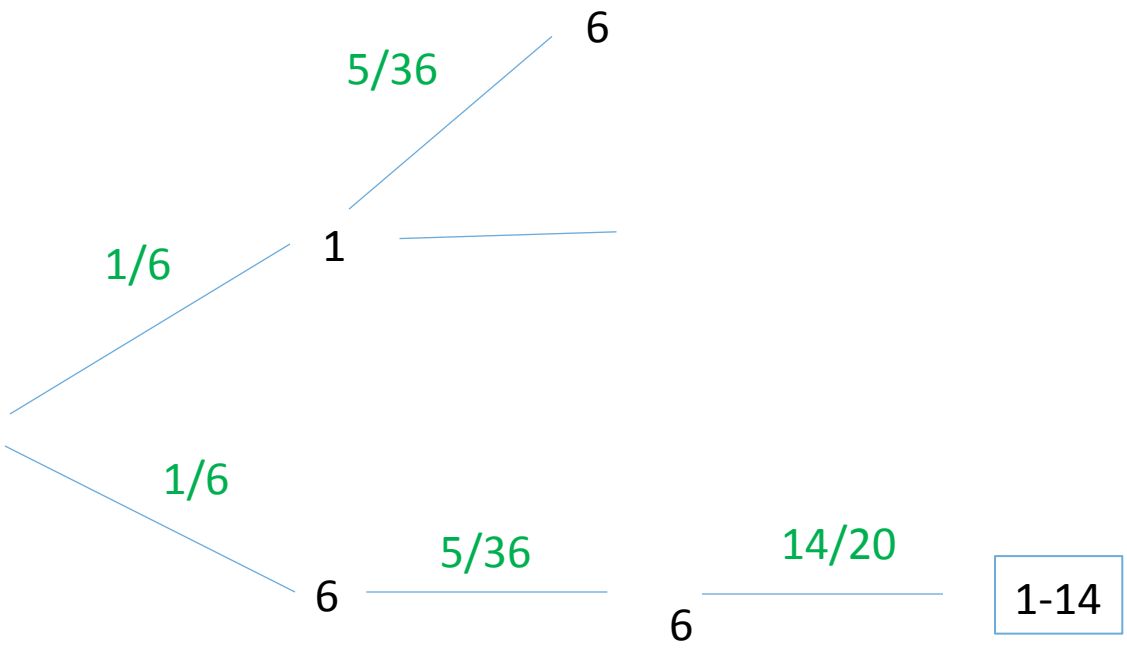
CLAYTON KERSHAW pitcher-3 batting # 1 starter

PITCHING CARD **LOS ANGELES (NL)**

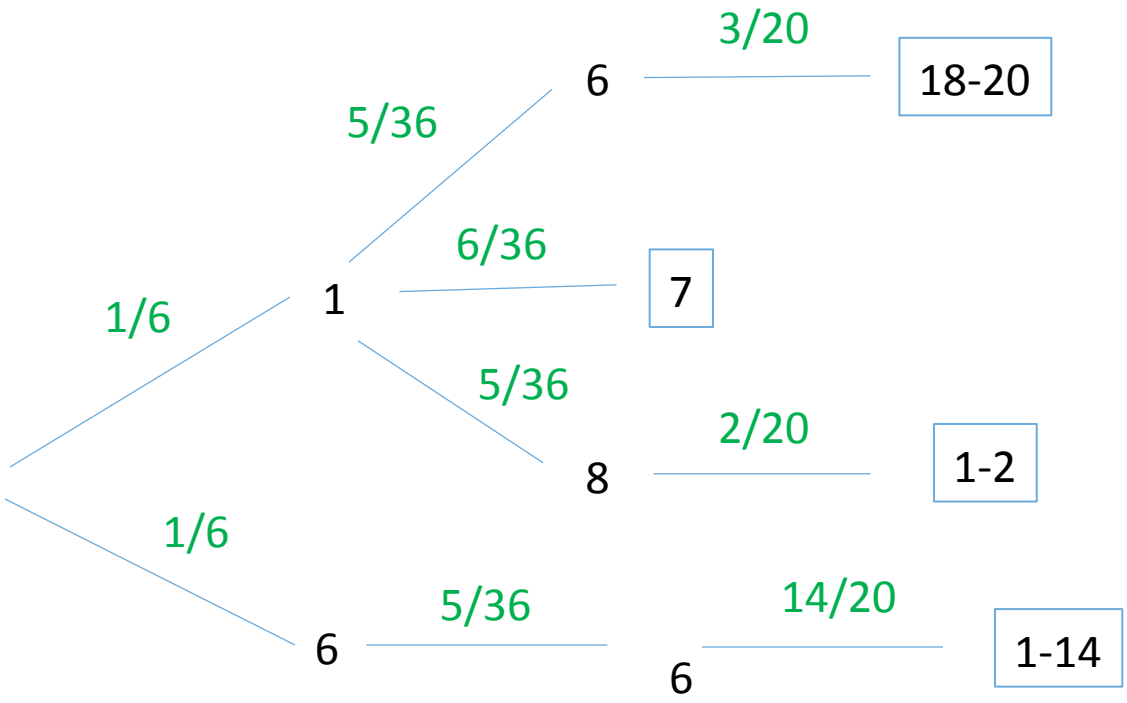
4	5	6
2-flyball (lf)C	2-groundball (p)B	2-WALK
3-FLYBALL (rf)X	3-GROUND-BALL(1b)X	3-GROUND-BALL(ss)X
4-FLYBALL (cf)X	4-CATCHER'S CARD X	4-flyball (lf)C
5-strikeout	5-strikeout	5-HOMERUN 1-2 flyball (lf)B 3-20
6-GROUND-BALL(ss)X	6-strikeout	6-DOUBLE** 1-14 SINGLE** 15-20
7-GROUND-BALL(2b)X	7-TRIPLE 1 SINGLE** 2-20	7-SINGLE* 1 lineout (1b) 2-20
8-strikeout	8-strikeout	8-WALK
9-strikeout	9-strikeout	9-WALK
10-GROUND-BALL(3b)X	10-strikeout	10-groundball (2b)C
11-FLYBALL (lf)X	11-GROUND-BALL(p)X	11-groundball (1b)C
12-groundball (p)B	12-groundball (2b)C	12-flyball (cf)B

2	3	4	5	6	7	8	9	10	11	12
1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Tree diagram

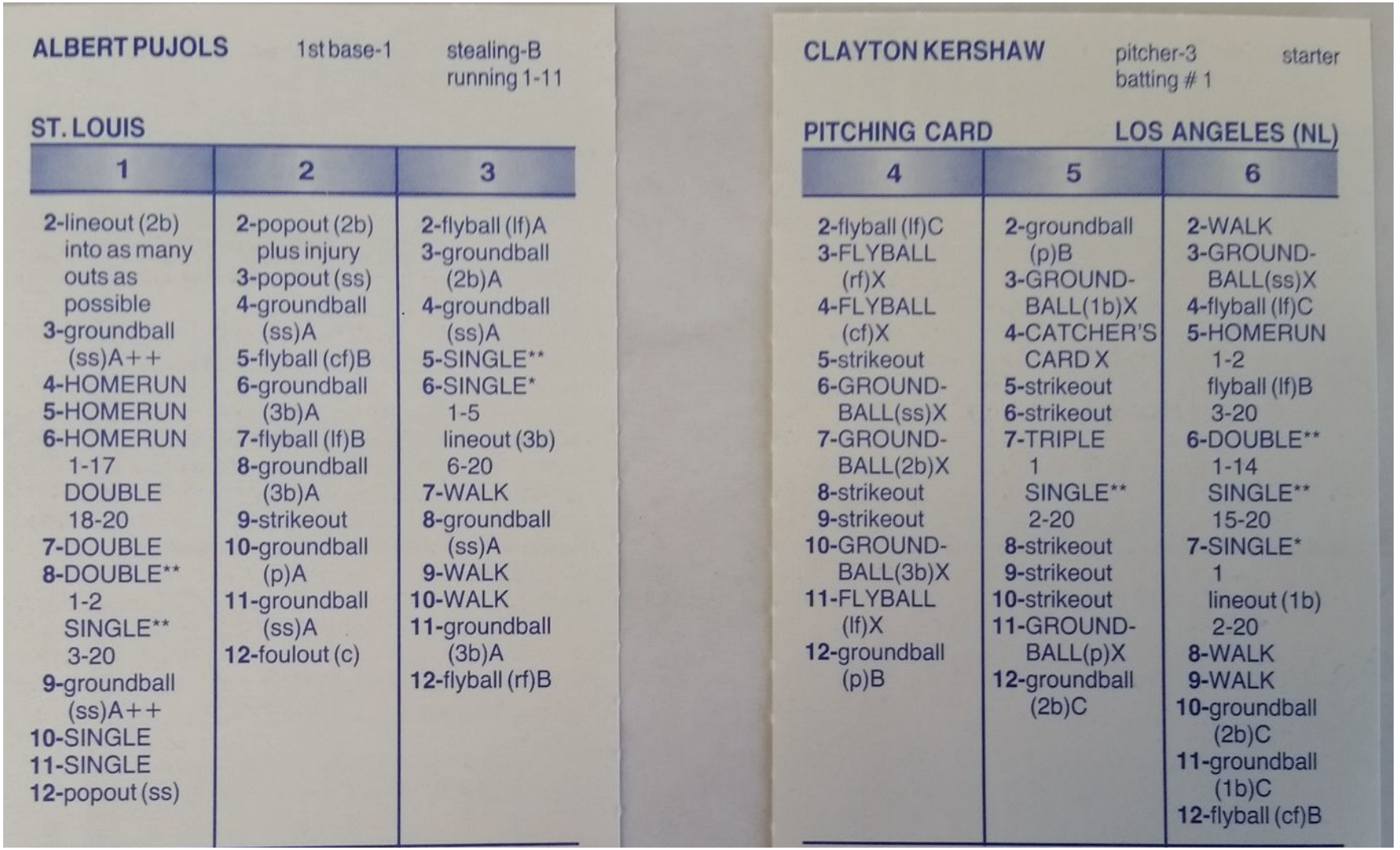


Tree diagram



$1/6 \cdot (5/36 \cdot 3/20 + 6/36 + 5/36 \cdot 2/20) + 1/6 \cdot 5/36 \cdot 14/20 = .04977$

Strat-o-matic: Pujols vs. Kershaw



ALBERT PUJOLS 1st base-1 stealing-B running 1-11

ST. LOUIS

1	2	3
2-lineout (2b) into as many outs as possible	2-popout (2b) plus injury	2-flyball (lf)A
3-groundball (ss)A++	3-popout (ss)	3-groundball (2b)A
4-HOMERUN	4-groundball (ss)A	4-groundball (ss)A
5-HOMERUN	5-flyball (cf)B	5-SINGLE**
6-HOMERUN 1-17 DOUBLE 18-20	6-groundball (3b)A	6-SINGLE* 1-5 lineout (3b) 6-20
7-DOUBLE	7-flyball (lf)B	7-WALK
8-DOUBLE** 1-2 SINGLE** 3-20	8-groundball (3b)A	8-groundball (ss)A
9-groundball (ss)A++	9-strikeout	9-WALK
10-SINGLE	10-groundball (p)A	10-WALK
11-SINGLE	11-groundball (ss)A	11-groundball (3b)A
12-popout (ss)	12-foulout (c)	12-flyball (rf)B

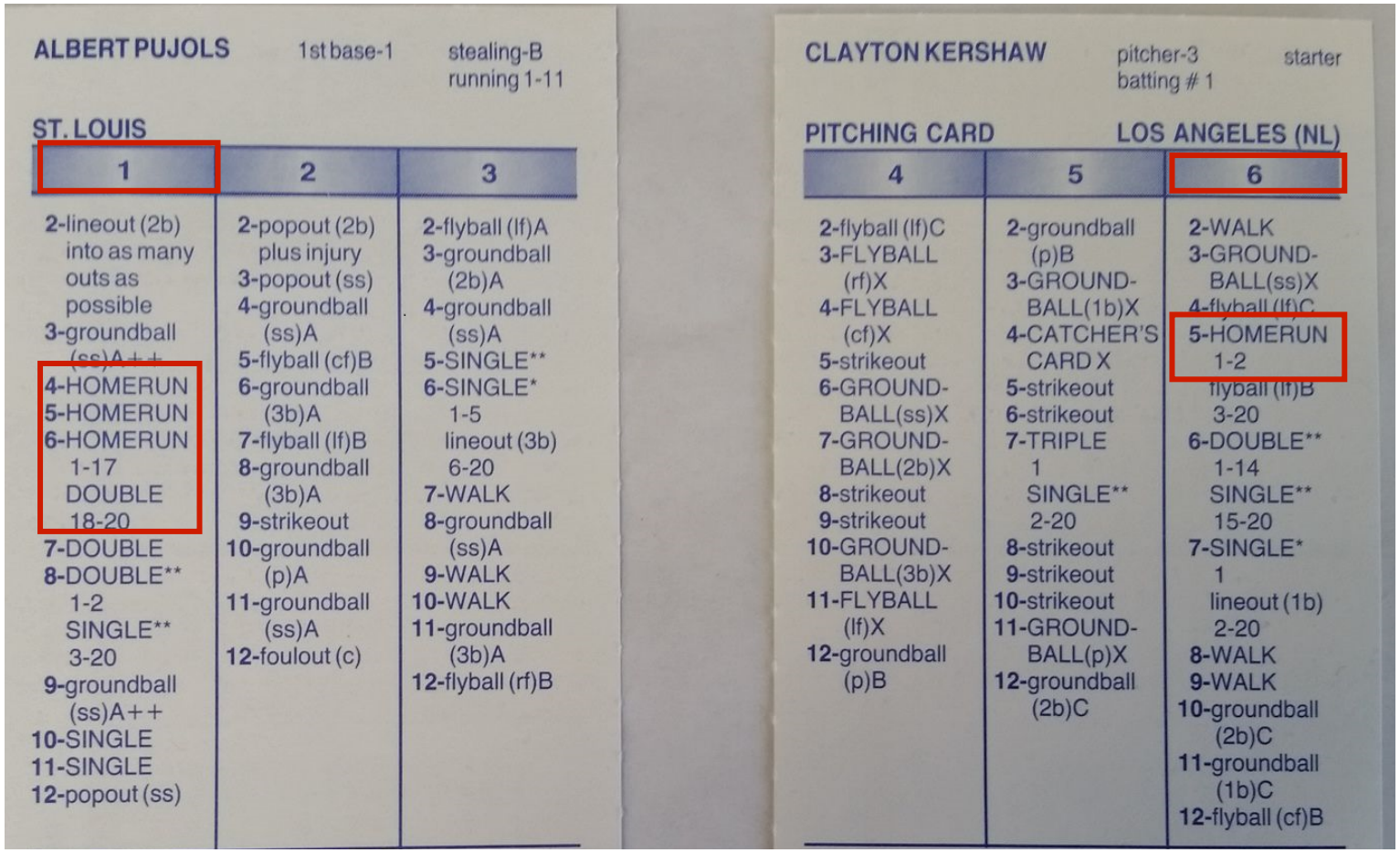
CLAYTON KERSHAW pitcher-3 batting # 1 starter

PITCHING CARD **LOS ANGELES (NL)**

4	5	6
2-flyball (lf)C	2-groundball (p)B	2-WALK
3-FLYBALL (rf)X	3-GROUND-BALL(1b)X	3-GROUND-BALL(ss)X
4-FLYBALL (cf)X	4-CATCHER'S CARD X	4-flyball (lf)C
5-strikeout	5-strikeout	5-HOMERUN 1-2 flyball (lf)B 3-20
6-GROUND-BALL(ss)X	6-strikeout	6-DOUBLE** 1-14 SINGLE** 15-20
7-GROUND-BALL(2b)X	7-TRIPLE 1 SINGLE** 2-20	7-SINGLE* 1 lineout (1b) 2-20
8-strikeout	8-strikeout	8-WALK
9-strikeout	9-strikeout	9-WALK
10-GROUND-BALL(3b)X	10-strikeout	10-groundball (2b)C
11-FLYBALL (lf)X	11-GROUND-BALL(p)X	11-groundball (1b)C
12-groundball (p)B	12-groundball (2b)C	12-flyball (cf)B

What is the probability of a hitting a home run?

Strat-o-matic: Pujols vs. Kershaw



ALBERT PUJOLS 1st base-1 stealing-B running 1-11

ST. LOUIS

1	2	3
2-lineout (2b) into as many outs as possible	2-popout (2b) plus injury	2-flyball (lf)A
3-groundball (ss)A++	3-popout (ss)	3-groundball (2b)A
4-HOMERUN	4-groundball (ss)A	4-groundball (ss)A
5-HOMERUN	5-flyball (cf)B	5-SINGLE**
6-HOMERUN 1-17 DOUBLE 18-20	6-groundball (3b)A	6-SINGLE* 1-5 lineout (3b) 6-20
7-DOUBLE	7-flyball (lf)B	7-WALK
8-DOUBLE** 1-2 SINGLE** 3-20	8-groundball (3b)A	8-groundball (ss)A
9-groundball (ss)A++	9-strikeout	9-WALK
10-SINGLE	10-groundball (p)A	10-WALK
11-SINGLE	11-groundball (ss)A	11-groundball (3b)A
12-popout (ss)	12-foulout (c)	12-flyball (rf)B

CLAYTON KERSHAW pitcher-3 batting # 1 starter

PITCHING CARD **LOS ANGELES (NL)**

4	5	6
2-flyball (lf)C	2-groundball (p)B	2-WALK
3-FLYBALL (rf)X	3-GROUND-BALL(1b)X	3-GROUND-BALL(ss)X
4-FLYBALL (cf)X	4-CATCHER'S CARD X	4-flyball (lf)C
5-strikeout	5-strikeout	5-HOMERUN 1-2 flyball (lf)B 3-20
6-GROUND-BALL(ss)X	6-strikeout	6-DOUBLE** 1-14 SINGLE** 15-20
7-GROUND-BALL(2b)X	7-TRIPLE 1 SINGLE** 2-20	7-SINGLE* 1 lineout (1b) 2-20
8-strikeout	8-strikeout	8-WALK
9-strikeout	9-strikeout	9-WALK
10-GROUND-BALL(3b)X	10-strikeout	10-groundball (2b)C
11-FLYBALL (lf)X	11-GROUND-BALL(p)X	11-groundball (1b)C
12-groundball (p)B	12-groundball (2b)C	12-flyball (cf)B

What is the probability of a hitting a home run?

Tree diagram!

$1/6 \cdot [3/36 + 4/36 + 5/36 \cdot 17/20] + 1/6 \cdot (4/36 \cdot 2/20) = .0594$

Parametric probability models

We have explored probability using:

- Probability rules to calculate the probability of an event
- Dice/spinners

We often use mathematical formulas, called probability distributions, to calculate the probability of different events

Random variates

We can think of a **random variate** as a random number

- Typically random variates are denoted with capital letters, e.g., X
- X can take on values in the sample space
  - Which in this case is a set of numbers

We can use probability distributions to assess the probability that a random variate X will have a value between two other numbers

- Notation: $Pr(a < X < b)$

Random variates

Random variates can be either:

- Discrete: X takes on integer values
- Continuous: X takes on real values

Same properties of probability distributions apply:

- $Pr(a < X < b) \geq 0$
- $\Sigma Pr(X = x_i) = 1$

Bernoulli Distribution

Models the probability of two outcomes:

- Success: $X = 1$
- Failure: $X = 0$
- E.g., getting a head ($X = 1$) or a tail ($X = 0$) for flipping a coin

Model has one parameter π, which is the probability of getting a 1

- E.g., the probability of getting head on a coin flip
- $Pr(X = 1) = ?$ $Pr(X = 0) = ?$

Bernoulli Distribution

Probability mass functions tell the probability of each outcome in the sample space of a discrete distribution

For Bernoulli Distributions we can write the probability mass function as:

$$f(k;\pi) = \begin{cases} \pi & \text{if } k = 1 \\ 1 - \pi & \text{if } k = 0 \end{cases}$$

$$f(k;\pi) = \pi^k (1-\pi)^{1-k} \quad \text{for } k \in \{0, 1\}$$

Parameter π

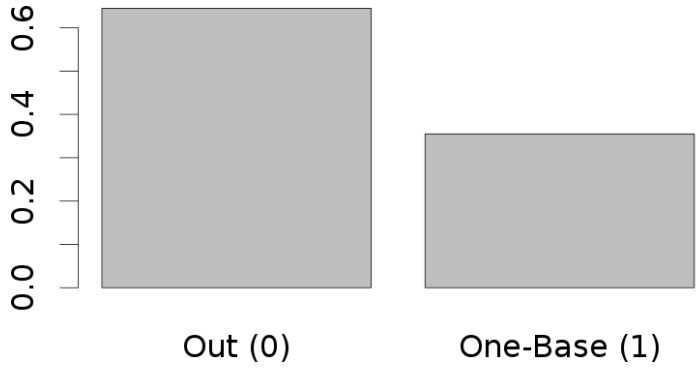
We can also plot the probability mass function

- If π = .5 what would it look like?
- If π = .9 what would it look like?

Bernoulli Distribution

Can you think of baseball example?

- Example: The probability that a player get on base for a given at bat
- If we were modeling David Ortiz, what would a good estimate of π be?
  - OBP = .355, so a good number to use would be π = .355



Binomial distribution

Models the probability of having k successes out of the n trials

- Probability of success on each trial is π

Example: if a player comes to bat 4 times in a game, what is the probability of getting on-base k times

- What is the sample space?

Assumes the same probability of getting on-base each plate appearance (π)

- Are the assumptions reasonable for this model?
  - No streakiness

Binomial distribution

Probability of getting ***k*** successes out of ***n*** trials is:

Parameters: π and n

$$Pr(X = k)$$

where $$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

What values can X take on (i.e., what is the sample space)?

- k ranges from 0, 1, …, n

n choose k

N choose k function tells us how many ways there are to order ***k*** items out of ***n*** total

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Q: How many ways are there to choose 3 things out of a total of 8?

A: $(8 \cdot 7 \cdot 6)/ (3 \cdot 2) = 56$

R: choose(n, k)