DATA 500

Introduction to Probability

Discrete Probability

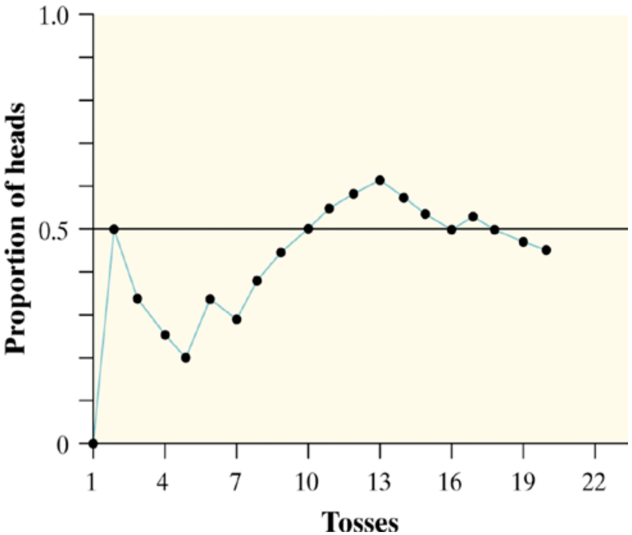
How to make money gambling

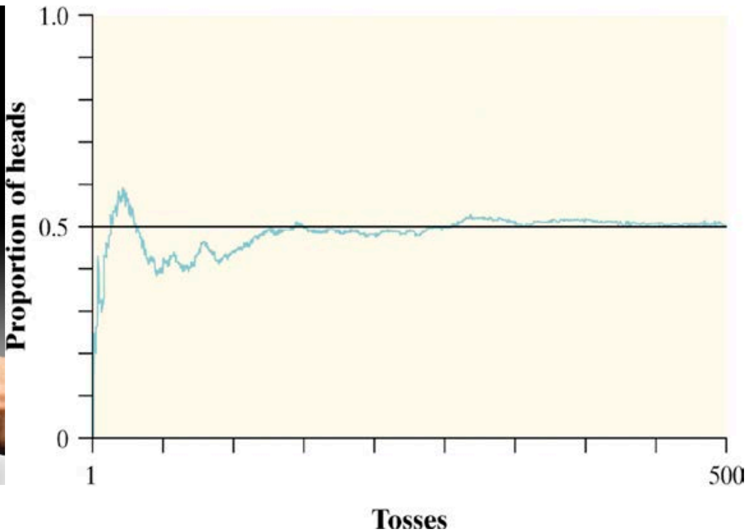
The Language of Probability

Chance behavior is unpredictable in the short run **but is regular and predictable in the long run.**

We call a phenomenon **random** if individual outcomes are uncertain **but** there is a regular distribution of outcomes in many repetitions.

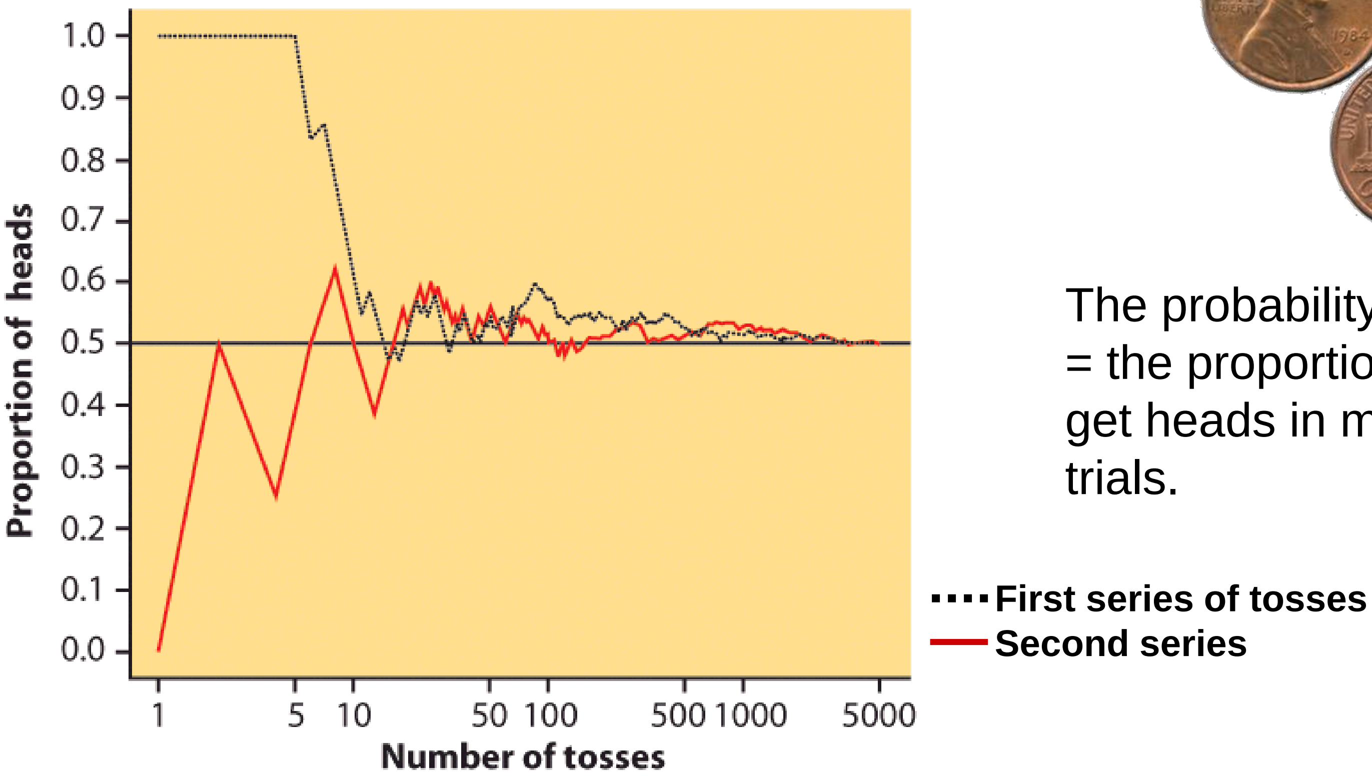
The **probability** of any outcome of a chance process is the proportion of times the outcome would occur in **many repetitions.**





Thinking About Randomness

The result of any single coin toss is random. But the result over many tosses is predictable, as long as the trials are **independent** (i.e., the outcome of a new coin flip is not influenced by the result of the previous flip).



The probability of heads is 0.5 = the proportion of times you get heads in many repeated trials.

Probability Models

Descriptions of chance behavior contain two parts: a list of possible outcomes and a probability for each outcome.

The **sample space *S*** of a chance process is the set of all possible outcomes.

An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.

A **probability model** is a description of some chance process that consists of two parts: a sample space *S* and a probability for each outcome.

Samples and Probabilities

Independent events: one outcome does not influence the next

Simple example: Coin tosses

Toss a coin twice: four outcomes, equal probabilities

Combo	***P***
H - T	**0.25**
H - H	**0.25**
T - H	**0.25**
T - T	**0.25**

P(HH) = 0.25

P(TT) = 0.25

P(one tail, one head, HT or TH) = 0.25 + 0.25 = 0.50 or 50%

one result OR another: add probabilities

Samples and Probabilities

Another example: Picking balls from a bag and replacing

****** For any proportions ******

N balls: R are red, N-R are green

Pick two balls. X is number of red balls

P(RR) or P(X=2) = (R/N) * (R/N) = $(R/N)^2$

P(GG) or P(X=0) = $((N-R)/N)^2$

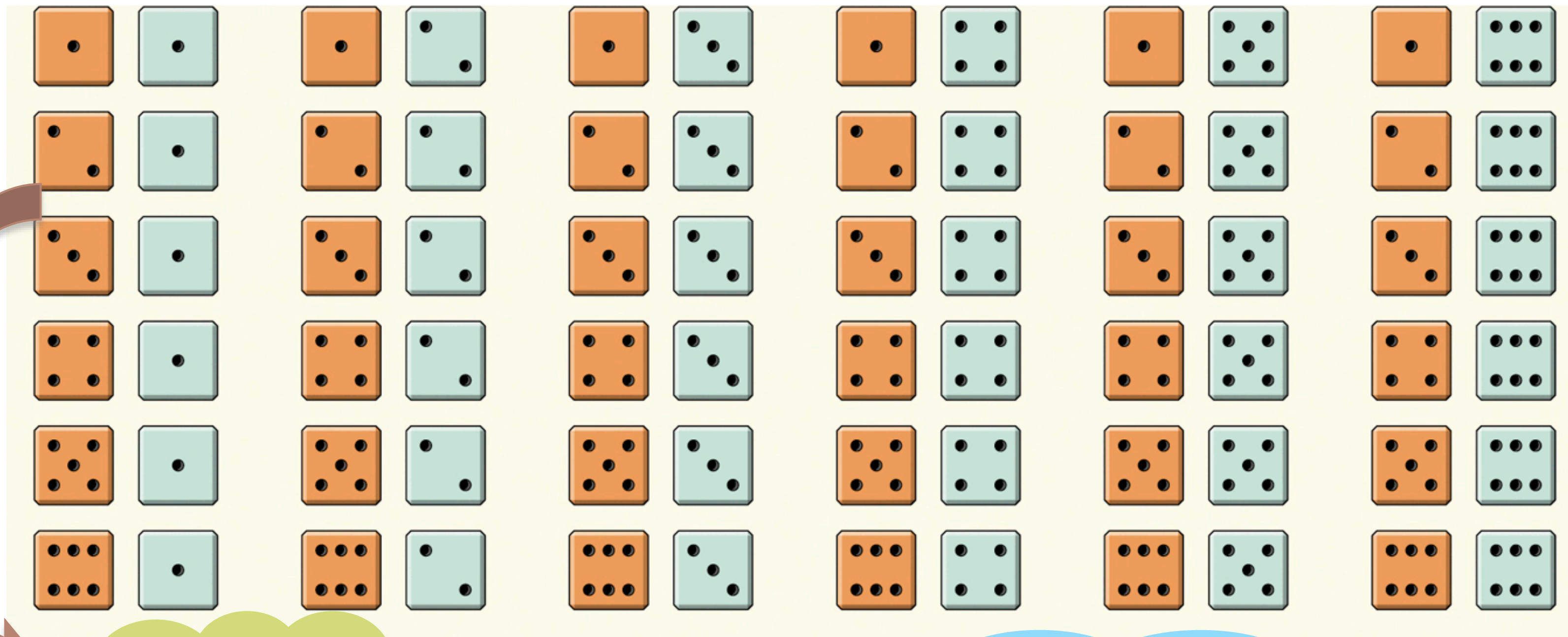
P (GR or RG) or P(X=1) = (R/N) * ((N-R)/N) * 2

Binomial probabilities, and DISCRETE probabilities (based on counts)

Independent events: one outcome does not influence the next

Probability Models

Example: Give a probability model for the chance process of rolling two fair, six-sided dice—one that's red and one that's green.



Sample Space
36 Outcomes

Since the dice are fair, each outcome is equally likely.
Each outcome has probability 1/36.

Probability Rules

1. Any probability is a number between 0 and 1.
2. All possible outcomes together must have probability 1.
3. If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities.
4. The probability that an event does not occur is 1 minus the probability that the event does occur.

Rule 1. The probability $P(A)$ of any event A satisfies $\mathbf{0 \le \boldsymbol{P}(\boldsymbol{A}) \le 1}$.

Rule 2. If S is the sample space in a probability model, then $\boldsymbol{P(S)} \mathbf{= 1}$.

Rule 3. If A and B are **disjoint**, $\boldsymbol{P(A \text{ or } B) = P(A) + P(B)}$.

This is the **addition rule for disjoint events.**

Probability Rules

Distance-learning courses are rapidly gaining popularity among college students. Randomly select an undergraduate student who is taking distance-learning courses for credit and record the student's age. Here is the probability model:

Age group (yr):	18 to 23	24 to 29	30 to 39	40 or over
Probability:	0.57	0.17	0.14	0.12

(a) Show that this is a legitimate probability model.

Each probability is between 0 and 1 and
$0.57 + 0.17 + 0.14 + 0.12 = 1$

(b) Find the probability that the chosen student is not in the traditional college age group (18 to 23 years).

$$\boldsymbol{P(\text{not 18 to 23 years}) = 1 - P(\text{18 to 23 years}) = 1 - 0.57 = 0.43}$$

Multiplication Rule for Independent Events

If two events *A* and *B* do not influence each other, and if knowledge about one does not change the probability of the other, the events are said to be **independent** of each other.

Rule 5: Multiplication Rule for Independent Events

Two events *A* and *B* are **independent** if knowing that one occurs does not change the probability that the other occurs. If *A* and *B* are independent:

$$P(A \text{ and } B) = P(A) \times P(B)$$

Random Sampling in R

R has the sample () function

`sample(dv, size=1, prob=p)`

dv = discrete values

size = number of times

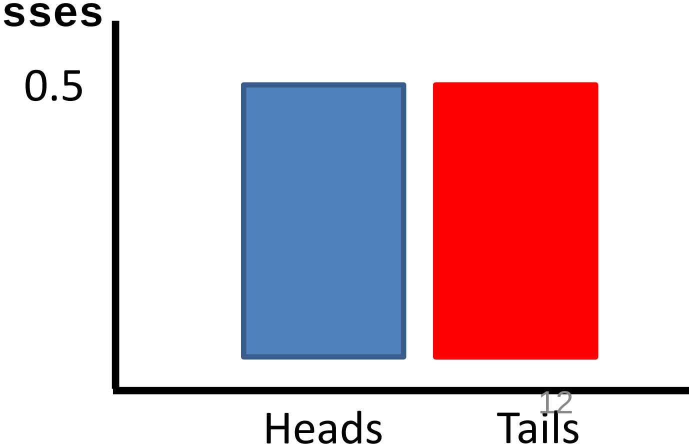
prob = probability for each value

Flip one coin:

```
# coin toss: 0 = tails, 1 = heads
# dv = 0 and 1
# probabilities (proportions) all sum to 1
p <- c(0.5,0.5)
# OR you can also use equal values for two equal probabilities:
# p <- c(1,1) # or p <- c(100,100)  #proportions are calculated
# flip a coin
sample(0:1, size=1, prob=p)
```

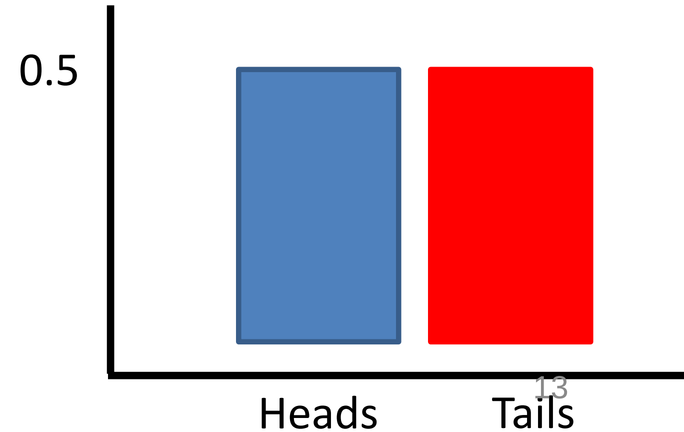
Random Sampling in R: a coin

```
# to flip a coin 100 times, we have to add replace = T
cointoss <- sample(0:1, size=100, prob=p, replace = T)
cointoss
  [1] 1 1 1 0 1 1 1 0 1 1 0 0 0 1 0 1 0 0 0 1 1 1 1 1 1 1 1 1 0 0 1 0 0 1 0 0 0 1
 [39] 1 1 0 1 1 0 0 1 1 0 1 1 0 0 0 1 0 1 0 0 0 1 1 0 0 0 1 0 0 1 1 0 1 1 1 0 0 1
 [77] 0 1 1 0 1 1 1 1 0 1 0 0 1 0 0 0 0 1 0 1 1 0 0 1
table(cointoss) # counts of each
 0  1
47 53
mean(cointoss) # mean for 0,1 is the proportion that are 1
[1] 0.53

# get the mean proportion of heads from 100 tosses
mean(sample(0:1, size=100, prob=p, replace=T))
[1] 0.54
mean(sample(0:1, size=100, prob=p, replace=T))
[1] 0.55
mean(sample(0:1, size=100, prob=p, replace=T))
[1] 0.51
```



Random Sampling in R: a coin

```
# flip a coin 100 times, use tails and heads
cointoss <- sample(c("tails","heads"), size=100, prob=p, replace = T)
cointoss
  [1] "heads" "heads" "heads" "tails" "heads" "tails" "tails" "heads" "heads"
 [10] "tails" "tails" "tails" "heads" "tails" "heads" "tails" "tails" "tails"
 [19] "tails" "heads" "tails" "heads" "tails" "heads" "tails" "tails" "tails"
 [28] "tails" "heads" "heads" "tails" "tails" "heads" "tails" "heads" "tails"
 [37] "tails" "tails" "heads" "heads" "tails" "tails" "tails" "heads" "heads"
 [46] "tails" "heads" "heads" "heads" "tails" "heads" "heads" "tails" "heads"
 [55] "heads" "heads" "tails" "tails" "heads" "tails" "heads" "heads" "heads"
 [64] "tails" "tails" "heads" "tails" "heads" "tails" "tails" "tails" "tails"
 [73] "tails" "heads" "heads" "heads" "tails" "heads" "heads" "tails" "heads"
 [82] "heads" "heads" "tails" "heads" "heads" "heads" "tails" "tails" "tails"
 [91] "heads" "heads" "heads" "tails" "tails" "heads" "heads" "heads" "tails"
[100] "heads"
table(cointoss) # get counts
cointoss
heads tails
   51    49
table(cointoss)/100  # probabilities / proportions
cointoss
heads tails
 0.51  0.49
```

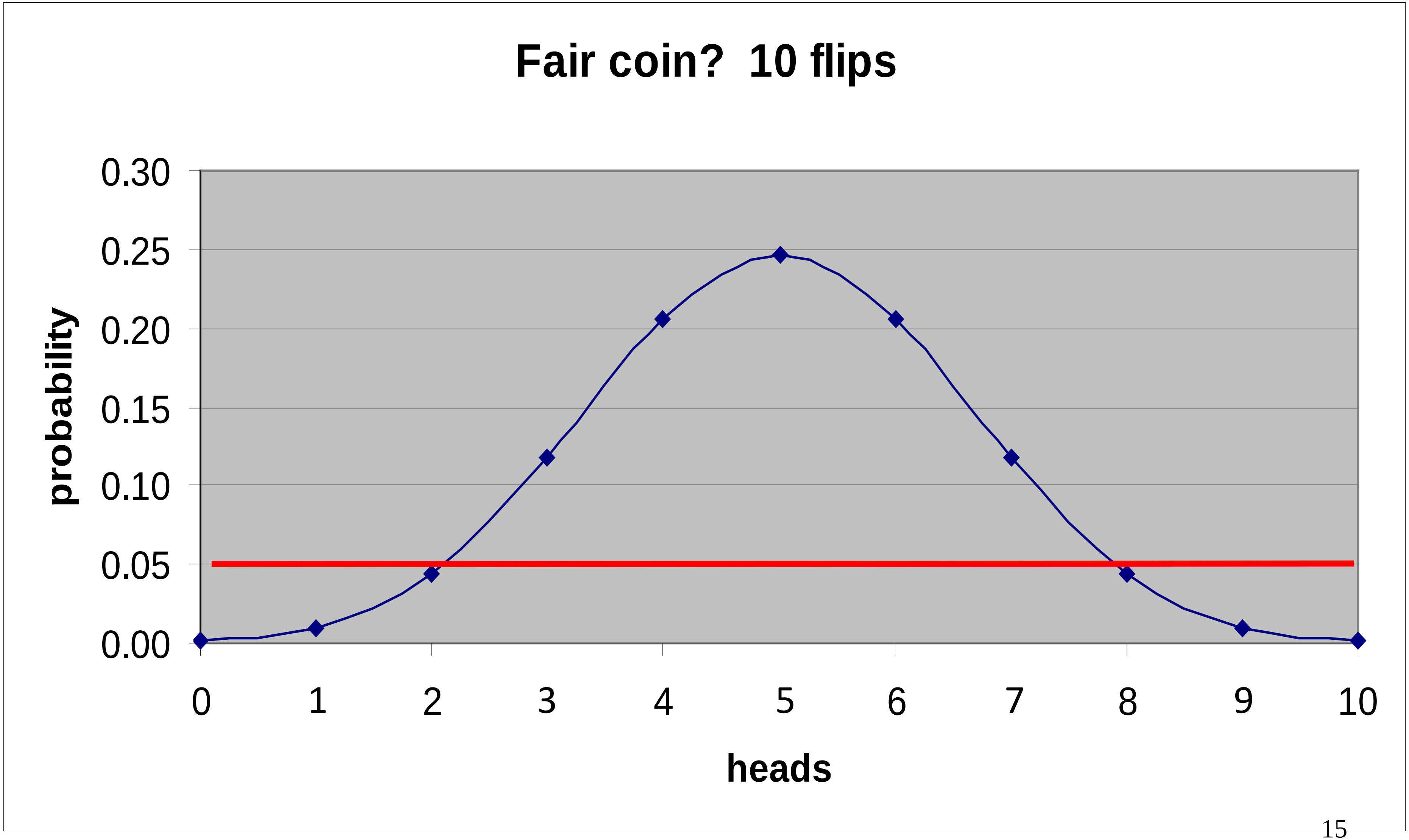


Random Sampling in R: a coin

```
# flip a coin only 10 times

# get the mean proportion of heads
mean(sample(0:1, size=10, prob=p, replace=T))
[1] 0.7
mean(sample(0:1, size=10, prob=p, replace=T))
[1] 0.6
mean(sample(0:1, size=10, prob=p, replace=T))
[1] 0.3
mean(sample(0:1, size=10, prob=p, replace=T))
[1] 0.4
```

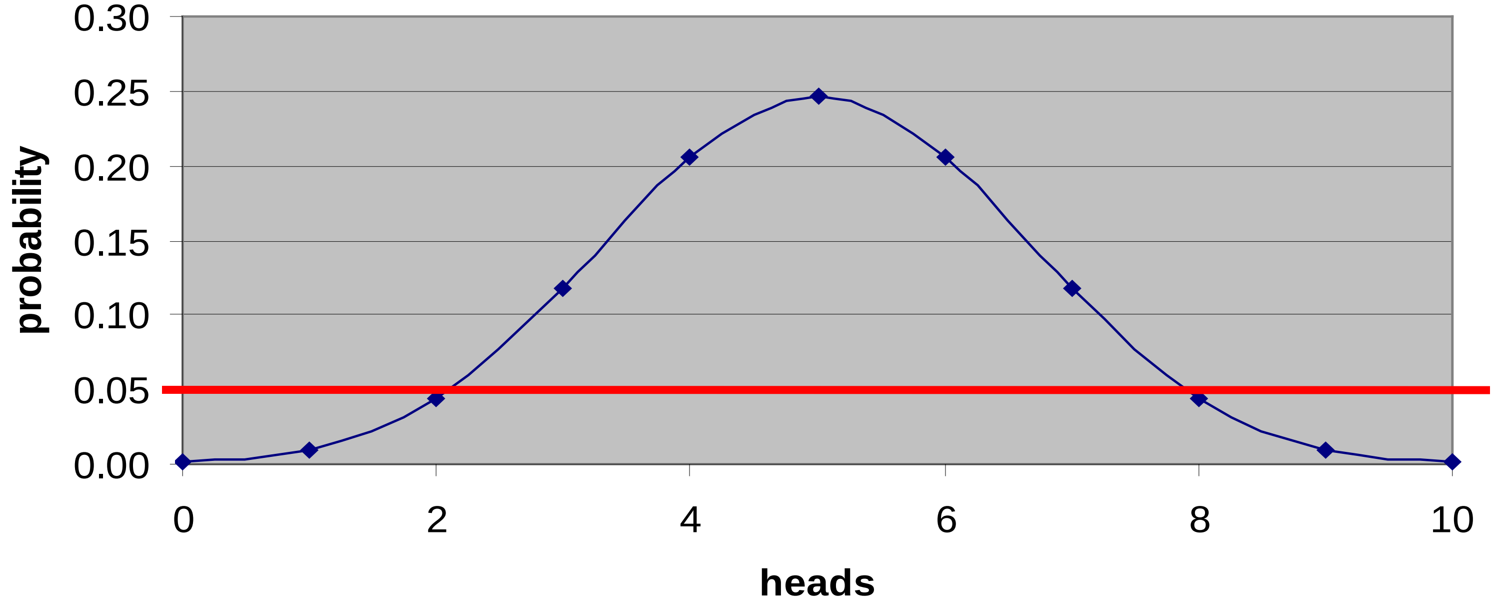
How are these proportions different from before?

Coin flips and discrete probability (binomial distribution)

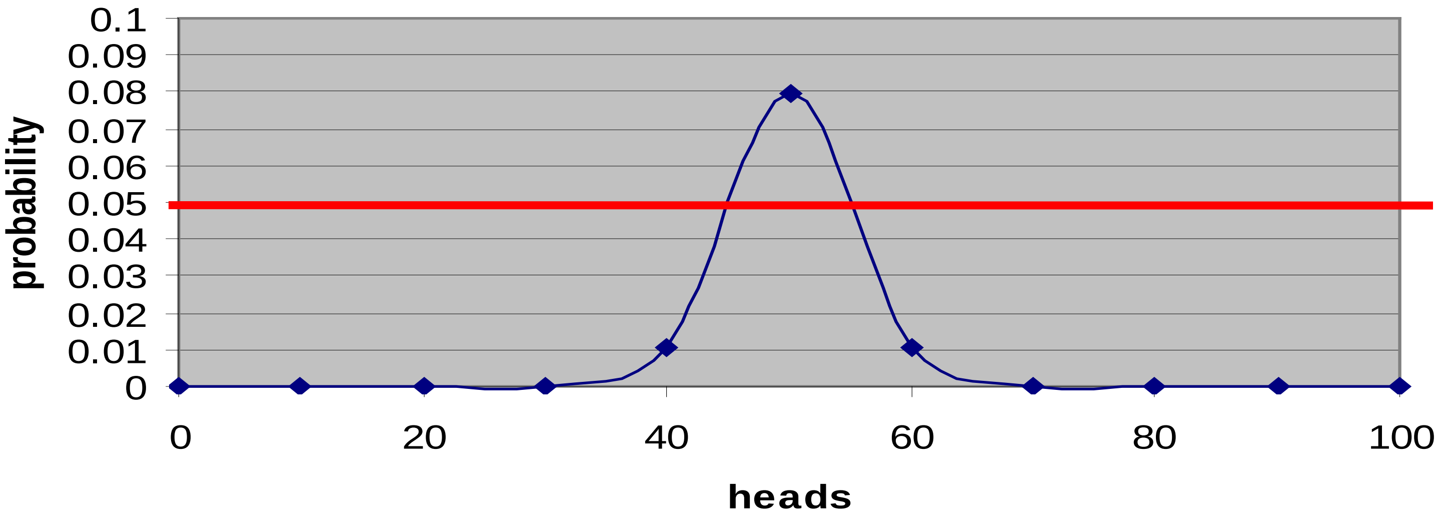


online calculator: http://vassarstats.net/binomialX.html

A larger sample size will provide a better estimate of actual proportions (probabilities)



N = 10



N = 100

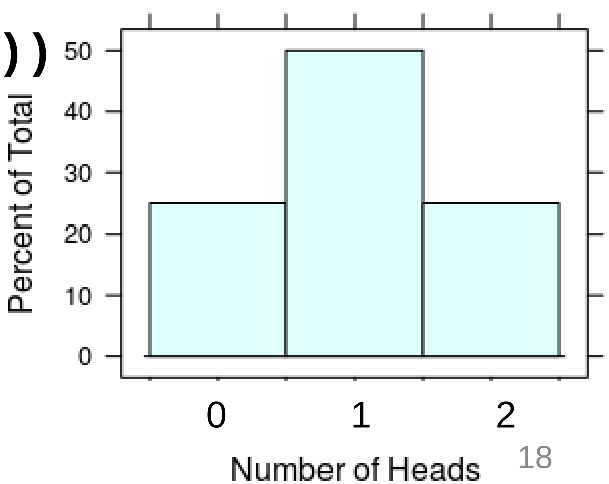
Random Sampling in R

Flip TWO coins

```
# TWO coin tosses: 0 = tails, 1 = heads, will sum them
# dv = 0:2
# coin tosses probabilities from before of 0.25 0.50  0.25
# T-T              H-T or T-H                HH
#  0                    1                     2
  (one way)          (two ways)           (one way)
p <- c(0.25,0.50,0.25); # proportions must sum to 1
# OR you can also do this, use proportional counts:
# p <- c(1,2,1); or # p <- c(25,50,25);
# flip two coins and get number of heads
sample(0:2, size=1, prob=p)
```

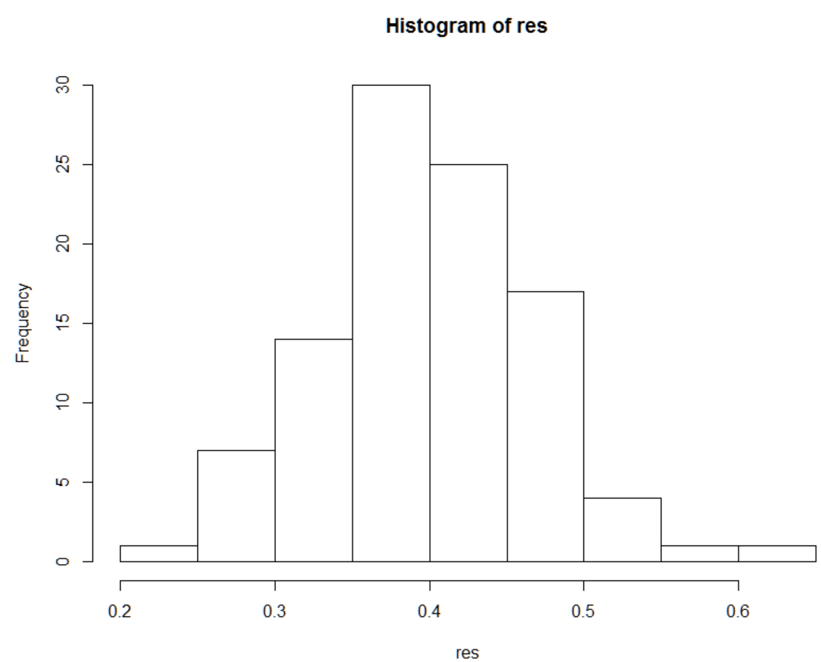
Random Sampling in R: TWO coins

```
# flip two coins 100 times, replace = T
table(sample(0:2, size=100, prob=p, replace=T))
sample(0:2, size=100, prob=p, replace=T)
 [1] 1 2 1 1 2 1 1 1 0 0 1 1 1 0 1 1 2 0 0 1 2 1 1 1 2 1 0 1 2 1 1 0 1 0
 [35] 1 2 1 2 1 1 1 0 2 0 0 2 2 2 2 2 1 1 2 1 1 1 0 0 2 1 0 2 1 2 1 1 0 1
 [69] 0 1 1 0 2 1 1 2 2 2 0 1 2 0 1 1 0 2 2 2 0 1 2 0 1 2 1 2 1 0 0 2
mean(sample(0:2, size=100, prob=p, replace=T))
[1] 1.02
mean(sample(0:2, size=10, prob=p, replace=T))
[1] 1
mean(sample(0:2, size=10, prob=p, replace=T))
[1] 1.2
mean(sample(0:2, size=10, prob=p, replace=T))
[1] 0.6
```



Random Sampling in R: uneven probabilities

```
urndraw <- sample(c("red","green"), size=100, prob=c(0.59,0.41), replace = T)
table(urndraw)
# red = 0, green = 1;
mean(sample(c(0:1), size=40, prob=c(0.59,0.41), replace = T));

# we can run this 100 times and see what the distribution looks like;
# create blank vector res for results
res <- c()
for (i in c(1:100)) { res[i] <- mean(sample(c(0:1), size=39, prob=c(0.59,0.41), replace = T)) };
hist(res);
```



What is the probability of getting a 1 on a die roll?

What is the probability of getting a 1 OR a 6 on a die roll?

Distribution of values of a die

One die: Uniform distribution (random outcome)

Random = 1/6 probability for each

Mean = 1 + 2 + 3 + 4 + 5 + 6 = 21/6 = 3.5

Random Sampling in R: a die

```
# roll a die, function assumes equal probs
sample(1:6, size=1)
[1] 5
mean(sample(1:6, size=100, replace = T))
[1] 3.39
mean(sample(1:6, size=100, replace = T))
[1] 3.47
mean(sample(1:6, size=100, replace = T))
[1] 3.53
mean(sample(1:6, size=100, replace = T))
[1] 3.62
mean(sample(1:6, size=100, replace = T))
[1] 3.55
mean(sample(1:6, size=10, replace = T))
[1] 3.8
mean(sample(1:6, size=10, replace = T))
[1] 3.9
```

What is the probability of getting 1 AND THEN 6 on two rolls of a die?

What is the probability of getting 1 AND 6 (any order) on a DICE roll?

What is the probability of getting a 7 TOTAL on a dice roll?

Sums of random values

One die: Uniform distribution

Probability = 1/6

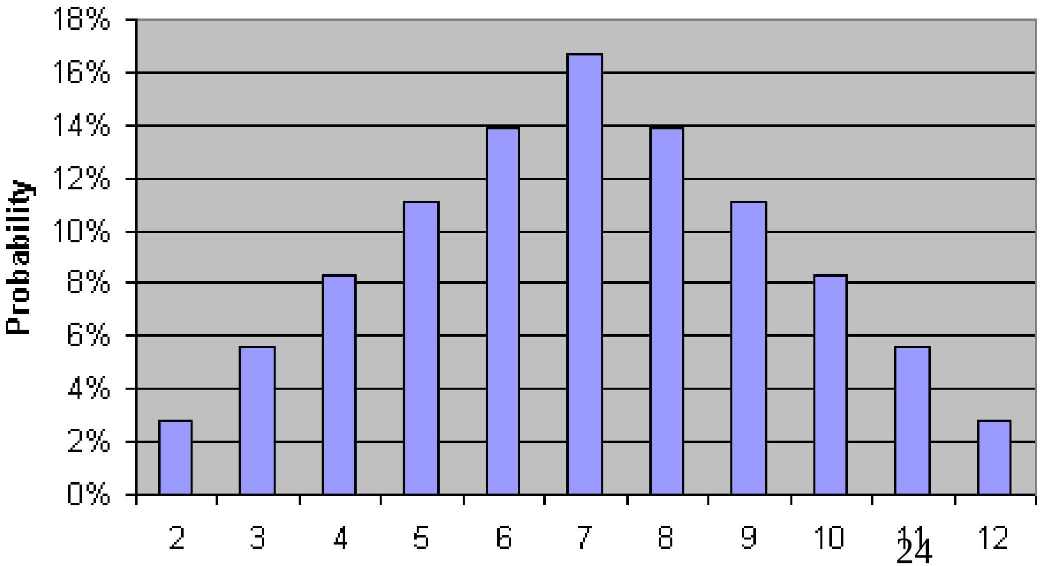
Mean = 3.5

Sum of two dice:

Looks close to normal

- same for summing coin flips

Mean = 7
$p(7) = 1/6$



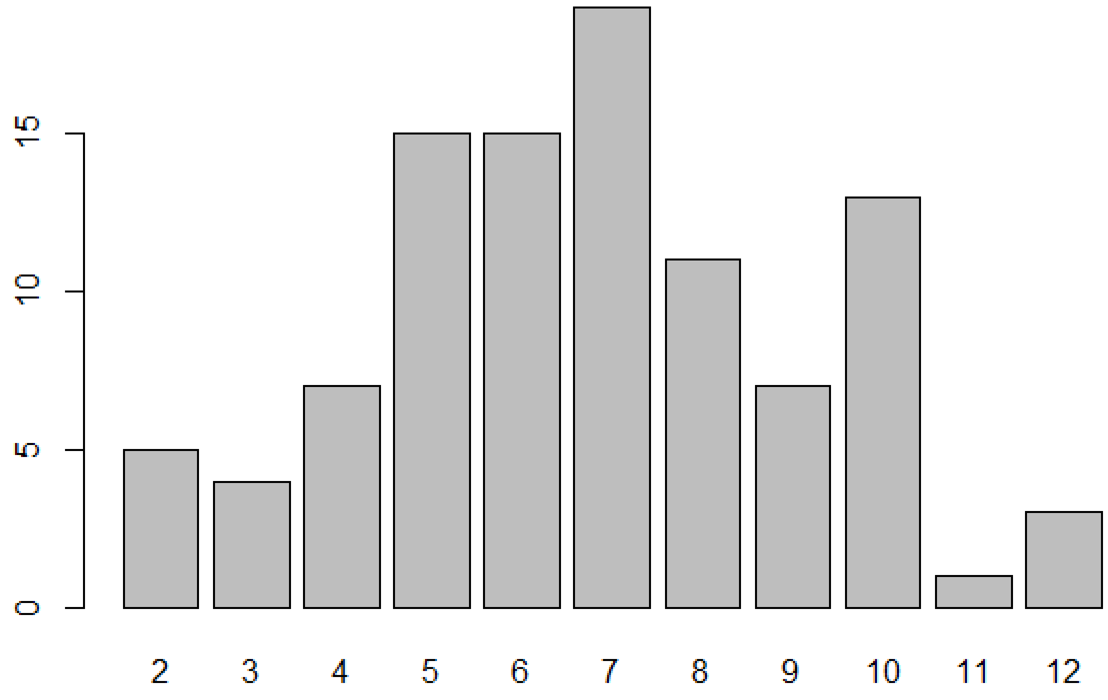
Sum of two dice

Random Sampling in R: dice

```
# roll two dice, independent events, calculate sums

sample(1:6, size=1) + sample(1:6, size=1)

mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 7.05
mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 7.44
mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 6.81
mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 7.07
mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 6.9
mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 7.15
mean(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
[1] 7.08

mean(sample(1:6, size=10, replace = T) + sample(1:6, size=100, replace = T))
[1] 6.7
```

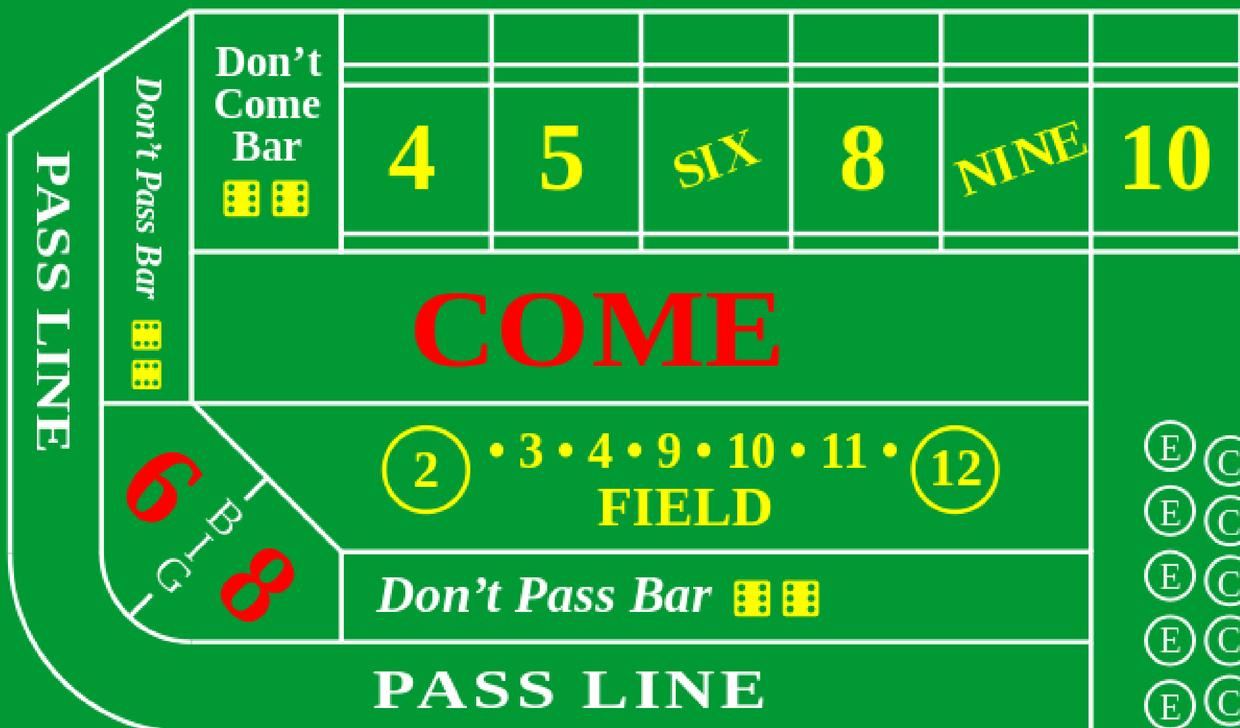
Graphing simulated discrete probabilities: dice

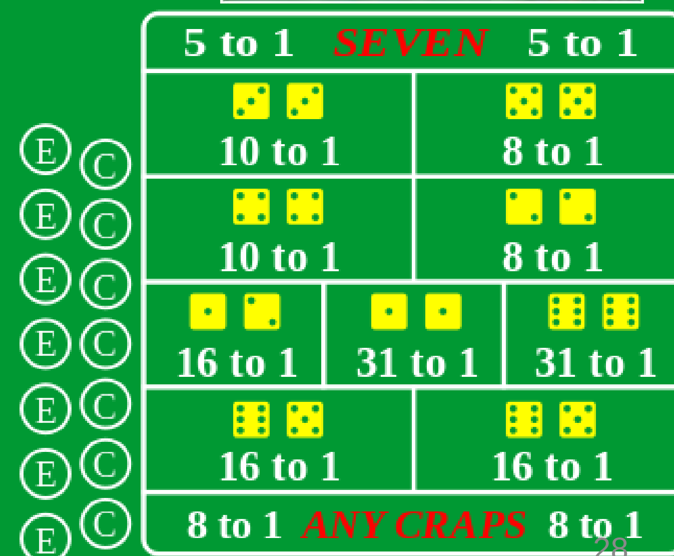
```
bt <- table(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T))
bt
 2  3  4  5  6  7  8  9 10 11 12
 5  4  7 15 15 19 11  7 13  1  3

barplot(bt)

barplot(table(sample(1:6, size=100, replace = T) + sample(1:6, size=100, replace = T)))

barplot(table(sample(1:6, size=10, replace = T) + sample(1:6, size=100, replace = T)))
```



How to make money gambling: Craps

Pass Line Bet - **You win if the first roll is a natural (7, 11)** and lose if it is craps (2, 3, 12). If a point is rolled (4, 5, 6, 8, 9, 10) it must be repeated before a 7 is rolled in order to win. If 7 is rolled before the point comes up again you lose.





How to make money gambling: Craps

Pass Line Bet - **You win if the first roll is a natural (7, 11)** and lose if it is craps (2, 3, 12). If a point is rolled (4, 5, 6, 8, 9, 10) it must be repeated before a 7 is rolled in order to win. If 7 is rolled before the point comes up again you lose.

House advantage

2 - 17%

House Advantage for specific bets

Pass/Come 1.41%

Don't Pass/Don't Come 1.40%

Pass + 1 x Odds 0.85%

Pass + 2 x Odds 0.61%

Pass + 5 x Odds 0.32%

Pass + 10 x Odds 0.18%

Pass + 100 x Odds 0.02%

Field (2:1 on 12) 5.56%

Field (3:1 on 12) 2.78%

Any Craps 11.11%

Big 6,8 9.09%

Hard 4,10 11.11%

Hard 6,8 9.09%

Place (to win) 6,8 1.52%

Place (to win) 5,9 4.00%

Place (to Win) 4,10 6.67%

Place (to lose) 6,8 1.82%

Place (to lose) 5,9 2.50%

Place (to lose) 4,10 3.03%

Proposition 2,12 13.89%

Proposition 3,11 11.11%

Proposition 7 16.67%

How to make money gambling: Craps

Proposition Bets

These bets can be made at any time and, except for the hardways, they are all one roll bets:

- **Any Craps**: Wins if a 2, 3 or 12 is thrown. **Payoff 8:1** **(p = 4/36 = 1/9)**
- **Seven**: Wins if a 7 is rolled. **Payoff 4:1** **(odds: 5:1)**
- **Eleven**: Wins if a 11 is thrown. **Payoff 15:1** **(odds: 17:1)**
- **Ace Deuce**: Wins if a 3 is rolled. **15:1** **(odds: 17:1)**
- **Aces or Boxcars**: Wins if a 2 or 12 is thrown. **Payoff 30:1**
- **Horn Bet**: it acts as the bets on 2, 3, 11 and 12 all at once. Wins if one of these numbers is rolled. Payoff is determined according to the number rolled. The other three bets are lost.
- **Hardways**: The bet on a hardway number wins if it's thrown hard (sum of pairs: 1-1, 3-3, 4-4...) before it's rolled easy and a 7 is thrown. Payoffs: Hard 4 and 10, 8:1; Hard 6 and 8, **Payoff 10:1**

How to make money gambling: Roulette

(0 to 36)

Payouts

A bet on one number only, called a straight-up bet, pays 35 to 1.

You collect 36 on a bet of 1 (a gain of 35).

p = 1/37 in Europe; p = 1/38 in the US (US has 00 (double-zero) also)

A two-number bet, called a split bet, pays 17 to 1.
A three-number bet, called a street bet, pays 11 to 1.
A four-number bet, called a corner bet, pays 8 to 1.
A six-number bet pays 5 to 1.
A bet on the outside dozen or column, pays 2 to 1.
A bet on the outside even money bets, pays 1 to 1.

House advantage

On a single zero roulette table the House advantage is 5.4% (1 - (1/37 * 35)). On a US double-zero roulette table it is 9.2% (1 - (1/38 * 35)). The House advantage is gained by paying the winners a chip or two (or a proportion of it) less than what it should have been if there was no advantage.

How to make money gambling

There are many many many different games.

The house has the advantage with every game,

so to make money gambling...

Homework 5

The National Basketball Association lottery to award the first pick in the draft is held by putting 1,000 balls into a hopper and selecting one. The teams with the worst records the previous year have a greater proportion of the balls. The data set nba_draft.csv in the usual location contains the ball allocation for the year 2002.

Use **sample** to generate the probability of each **Team** using **Balls** to simulate the draft.

Calculate the probability of being picked for each team. (consult slides)

Repeat sampling until Golden State is chosen.

How many tries did it take?

Due before class Feb. 13