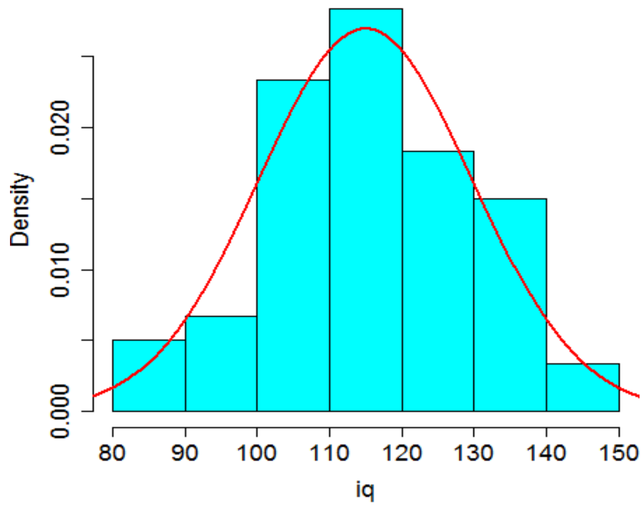
Review so far:

- Types of data (categorical, quantitative)
- Displaying sampled data (bar graph, histogram)
- Describing data with numbers (mean, median, variance, percentiles)
- Densities: what they are
- Normal density: properties, usage



What's ahead:

- Statistical methods are more than descriptive

- ***The goal of statistics is the creation of models to better understand complex systems***

- All statistical models involve *chance events* and *probability*

(Section 4 from the textbook)

4.1: What are chance events, and what is probability?

All statistical models involve chance events and probability

- Everyone knows: dice rolls have certain chances
- But really, the outcome is determined by physics:

Throw angle

Air currents

Obstacles

- No randomness: so why do we still say **probability**?

4.1: What are chance events, and what is probability?

All statistical models involve chance events and probability

- Probability models are useful when ***the system is too complex to predict the outcome perfectly***

- Physical systems are *approximated* by prob. models

4.1: What are chance events, and what is probability?

All statistical models involve *chance events* *and* *probability*

Probability models:

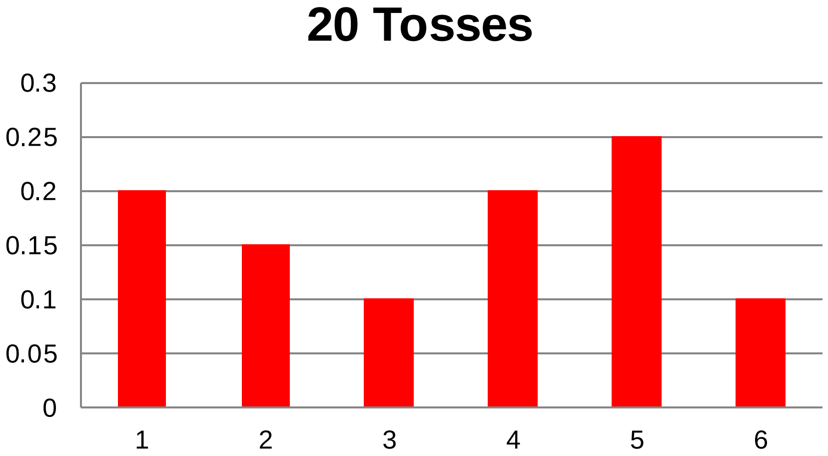
- Constructed and checked with current data
- Help predict future data
- Yields understanding of long-term behavior

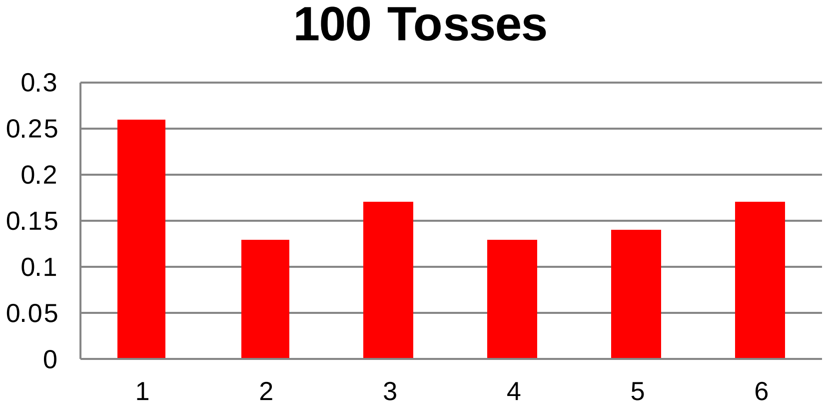
Every model has *events* and each event has a *probability***:**

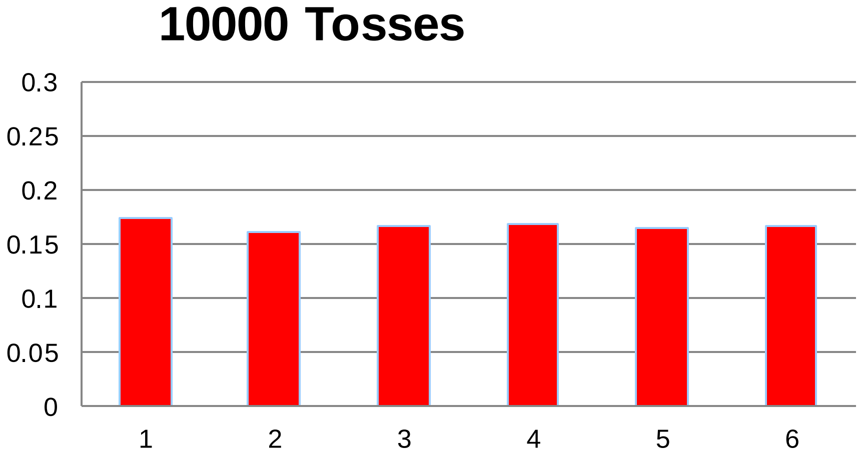
- **Event**: a possible outcome from a model
- **Probability** of an event: proportion of times an event is observed over the long-term

Example: Motivation of definition of probability

- Toss a **fair** 6-sided die a number of times.







4.1: Randomness (pgs. 233 – 235)

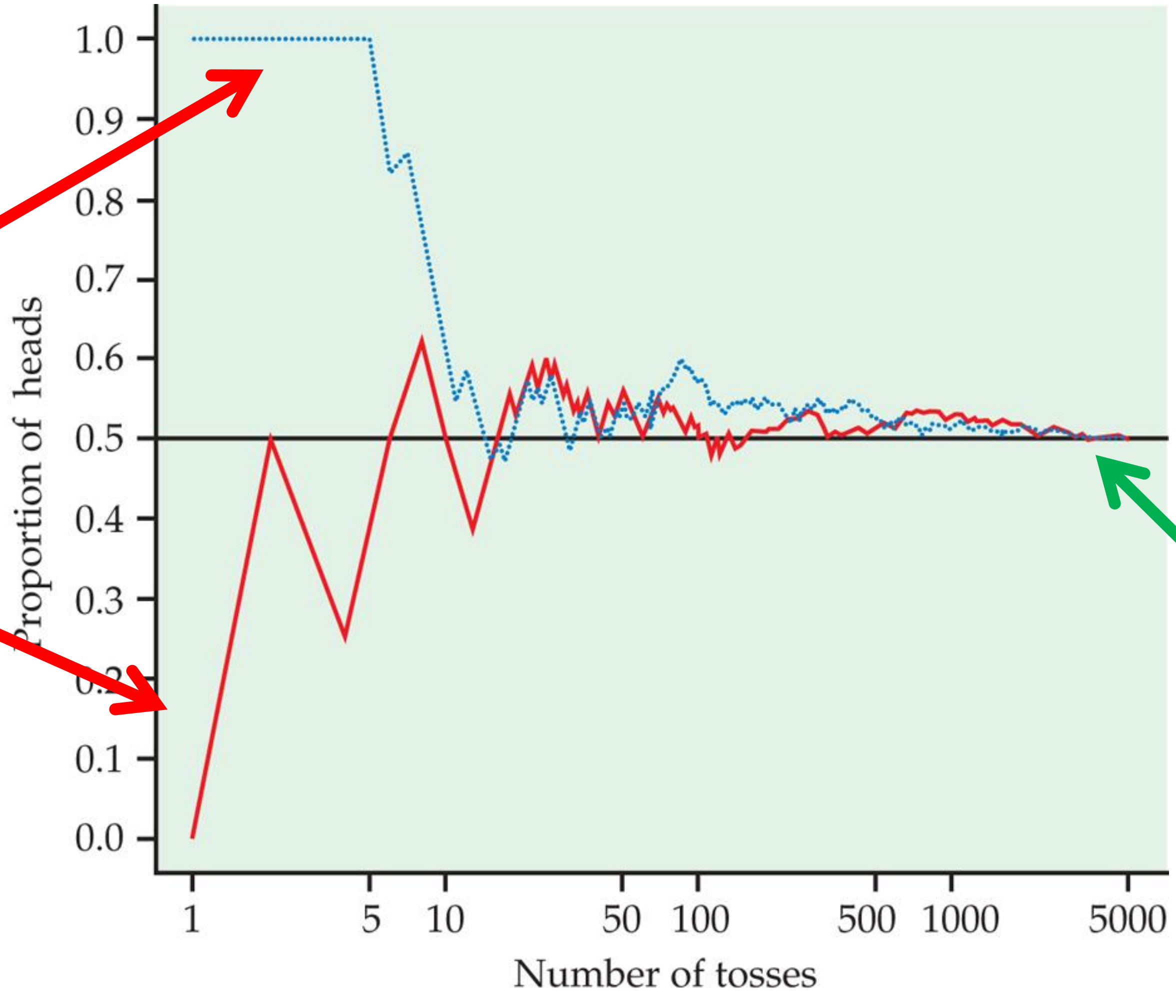
Common systems modelled with probability:

- toss a fair coin, note whether it comes up H or T
- take a SRS from a population, ask opinions on the Affordable Care Act
- deal two cards, see if they match
- The next Carolina-Duke basketball game equals win or loss?

In each case, *probability* = what we would expect for the distribution of outcomes after many repetitions.

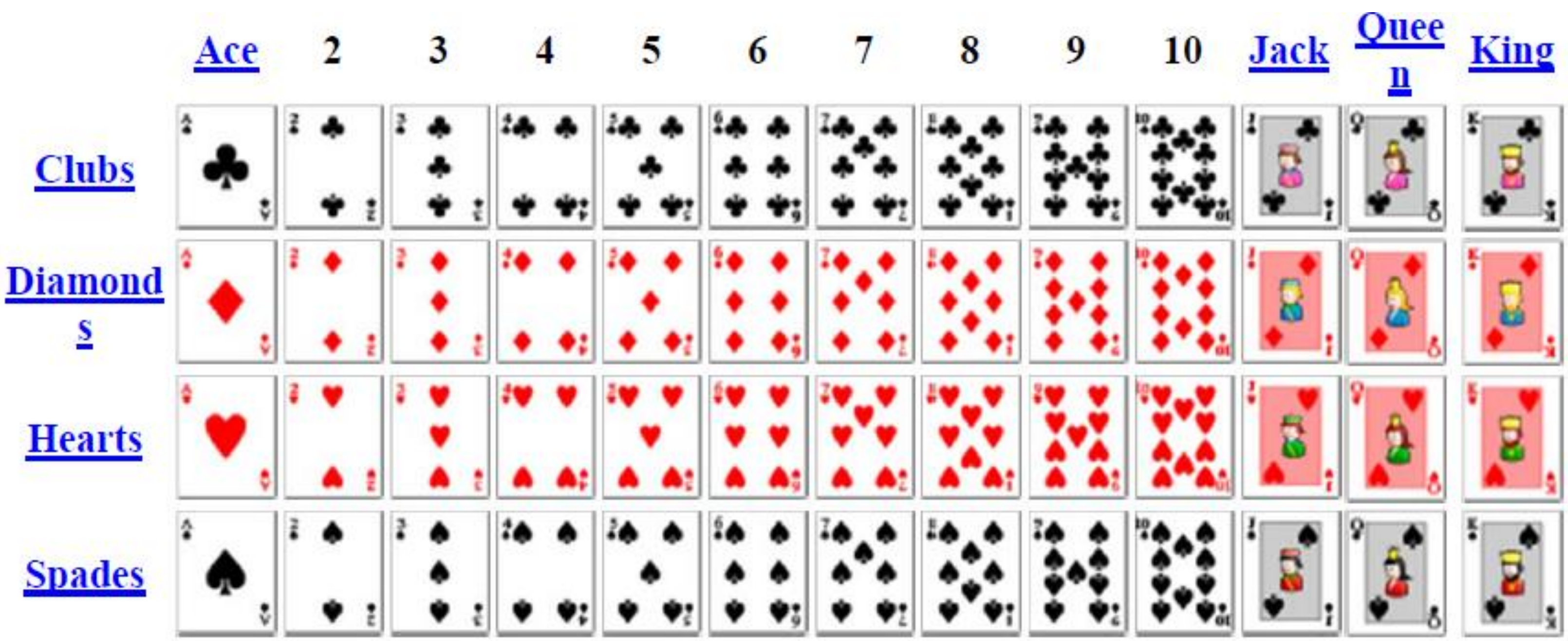
Example from page 232 (Section 4.1)

What actually happens



What our model says

Needed for an example:
Standard Deck of Cards



Wikipedia.org

4.1: Randomness

Example 1:

Phenomenon: Deal 2 cards at random from a shuffled deck.

Event of interest: The two cards have the same number.

What actually happens:

In 15,000 simulated repetitions, the outcome occurred on 963 repetitions. So we estimate the probability to be 963/16000 = 0.0602.

4.1: Randomness

Example 1 (continued):

What actually happens: Event occurs 6.02% of times

What our model says:

Given that we have drawn one card from the deck already, there are 51 cards in the deck. 3 of them match the one already drawn.

Assume the deck is thoroughly shuffled so that every remaining card has an equal chance of being dealt at any time. Then the probability of ***a match*** with the card drawn is

$$(3/51) = 0.05882 = 5.88\%$$

4.2 Probability models

- Key Ideas
  - Sample Space
  - Events
  - Probabilities
  - Independence and the Multiplication Rule

PROBABILITY MODEL (Section 4.2):

1. Describe a phenomenon
2. Describe all possible outcomes (the <u>sample space</u>)
3. Assign a probability to each possible outcome

Example 1: Toss a fair coin four times and note the outcomes (H or T), in order.

- *Possible outcomes*: This is the <u>sample space</u>

  TTTT,
  HTTT, THTT, TTHT, TTTH
  HHTT, THHT, TTHH, HTHT, HTTH, THTH
  HHHT, HHTH, HTHH, THHH
  HHHH

- *Probability of each outcome*:

  We assume that each outcome has probability 1/16.

Event: An outcome or set of outcomes (more formally: a subset of the sample space)

- *Example:* Find the probability of this event: "3 heads and 1 tail (in any order) in 4 tosses"

-This corresponds to the *subset of outcomes* {HHHT, HHTH, HTHH, THHH}.

- *Probability of this event*: Each of the 16 outcomes has probability $\frac{1}{16}$.

- Probability of 3 heads and 1 tail is $\frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = 1/4$.

The probability of an event is the sum of the probabilities for outcomes that make up the event.

Sec. 4.2. Probability models

Example 2: Choose a person at random from a group of 100 people that includes 43 Democrats, 39 Republicans, and 18 Independents and note their party preference.

-*Possible outcomes*: D, R, I

-*Probabilities:* $P(D) = 0.43$, $P(R) = 0.39$, $P(I) = 0.18$

- *Event:* "Person chosen is not a Democrat" corresponds to a *subset* of the set of outcomes:

"not a Democrat" $= \{R, I\}$.

- *Probability of an event*: The probability of an event is the sum of the probabilities of its outcomes.

Probability of "not a Democrat" is $0.39 + 0.18 = 0.57$.

Sec. 4.2. Probability models

Example 3 (Less Obvious): Choose a random member of a very large population whose scores on a certain test are normally distributed with mean $\mu = 100$ and standard deviation $\sigma = 10$.

A probability model for this:

- *Possible outcomes*:

  All real numbers

- *Event:* "score is less than 115" corresponds to the subset of all scores from $-\infty$ to 115

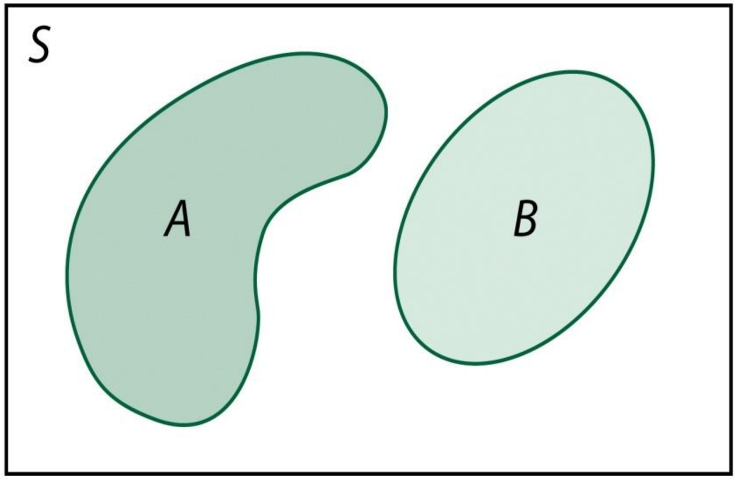
- *Probability of this event* = the proportion of area under the N(100, 10) curve, before x = 115!

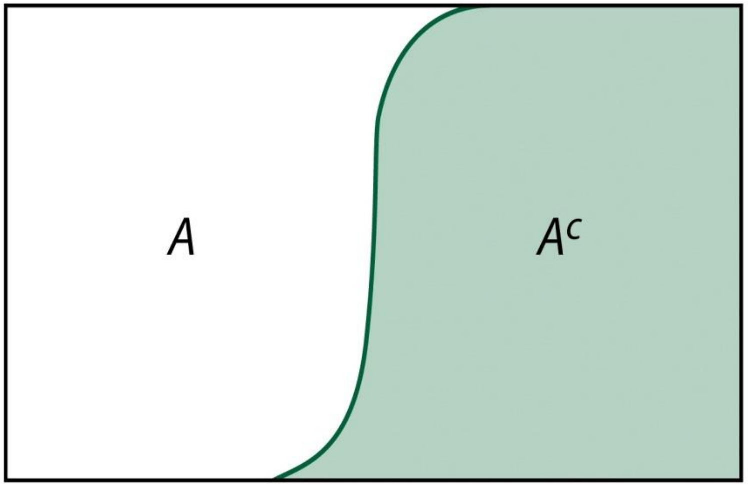
- (More about this example later on)

Sec. 4.2. Probability models

Describing events in a sample space:

Events A and B are *disjoint* Event A^c is *complement* of A





Example 4: The experiment: toss a coin 10 times and record the number of heads. *What is the sample space?*

Let $A = \{0, 2, 4, 6, 8, 10\}$, $B = \{1, 3, 5, 7, 9\}$, $C = \{0, 1, 2, 3, 4, 5\}$.

A or B = ? are A and B disjoint? are A and C disjoint? A^c = ?

Sec. 4.2. Probability models (pgs 239 – 240)

Rules for assignment of probabilities to events:

$P(A)$ denotes the probability of event A in sample space S.

Rule 1: For any event A, $0 \leq P(A) \leq 1$.

Rule 2: $P(S) = 1$.

Rule 3: If events A and B are ***disjoint*** (have no outcomes in common, i.e. cannot occur together), then

$$P(A \text{ or } B) = P(A) + P(B).$$

Rule 4: If A^c denotes the ***complement*** of A (the event that A does not occur), then

$$P(A^c) = 1 - P(A).$$

Sec. 4.2. Probability models (pgs 242 – 243)

Assigning probabilities when the sample space is finite:

Example:

For numbers in tables of financial, demographic, and other data, the ***first digits*** are not equally likely!

Their proportions typically are close to ***Benford's Law***:

digit	1	2	3	4	5	6	7	8	9
prob	.301	.176	.125	.097	.079	.067	.058	.051	.046

Sec. 4.2. Probability models

Populations of NC counties (2008 estimates)

County	Pop	County	Pop	County	Pop	County	Pop
MECKLENBURG	877,007	BRUNSWICK	102,857	PENDER	51,853	CURRITUCK	23,773
WAKE	864,429	CLEVELAND	97,936	EDGECOMBE	51,800	HERTFORD	23,764
GUILFORD	468,344	CRAVEN	97,757	RICHMOND	46,842	CASWELL	23,422
FORSYTH	343,704	NASH	93,981	STOKES	46,638	GREENE	21,205
CUMBERLAND	316,914	ROCKINGHAM	91,691	BEAUFORT	46,590	NORTHAMPTON	21,168
DURHAM	260,420	BURKE	89,259	WATAUGA	45,319	MADISON	20,810
BUNCOMBE	227,875	MOORE	85,280	MCDOWELL	44,562	BERTIE	20,074
GASTON	204,971	CALDWELL	80,020	HOKE	44,432	WARREN	19,918
NEW HANOVER	192,235	WILSON	78,917	VANCE	43,502	POLK	18,992
UNION	191,108	LINCOLN	74,538	PASQUOTANK	41,330	YANCEY	18,592
ONSLOW	176,004	SURRY	73,388	DAVIE	40,970	AVERY	18,428
CABARRUS	170,406	WILKES	67,297	YADKIN	38,162	MITCHELL	16,034
JOHNSTON	162,746	SAMPSON	65,396	PERSON	37,510	CHOWAN	14,687
DAVIDSON	158,866	RUTHERFORD	63,555	SCOTLAND	37,064	SWAIN	13,982
PITT	155,570	CARTERET	63,520	JACKSON	36,990	WASHINGTON	13,172
CATAWBA	154,941	CHATHAM	60,881	ALEXANDER	36,953	PERQUIMANS	12,962
IREDELL	154,135	STANLY	59,714	MACON	34,227	PAMLICO	12,892
ALAMANCE	145,995	FRANKLIN	57,923	DARE	33,955	GATES	11,836
RANDOLPH	140,980	LENOIR	57,521	BLADEN	32,153	ALLEGHANY	11,125
ROWAN	138,512	LEE	57,500	TRANSYLVANIA	30,991	CLAY	10,458
ROBESON	130,316	HAYWOOD	57,108	MONTGOMERY	27,651	JONES	10,292
ORANGE	129,296	GRANVILLE	56,250	CHEROKEE	27,128	CAMDEN	9,730
WAYNE	115,696	HALIFAX	55,217	ASHE	26,319	GRAHAM	8,087
HARNETT	109,637	COLUMBUS	54,758	ANSON	25,368	HYDE	5,516
HENDERSON	103,836	DUPLIN	53,431	MARTIN	23,870	TYRRELL	4,280

1st digit	1	2	3	4	5	6	7	8	9
prop.	.32	.13	.13	.10	.11	.05	.05	.06	.05
B. Law	.301	.176	.125	.097	.079	.067	.058	.051	.046

Sec. 4.2. Probability models

Assigning probabilities when the sample space is finite:

The probabilities predicted by Benford's Law:

Digit	1	2	3	4	5	6	7	8	9
Prob	.301	.176	.125	.097	.079	.067	.058	.051	.046

This is an assignment of probabilities to the individual outcomes. What about other events?

Pick a number at random from such a table and let X be its first digit.

1. Probability that X is 1 or 2:
2. Probability that X is greater than 2:
3. Probability that X is even:
4. Probability that X is either 1 or even:

Sec. 4.2. Probability models

Assigning probabilities when the sample space is finite:

Example:

Roll 3 fair dice, X = the number of 6's that come up.
Sample space is $S = \{0,1,2,3\}$.

We'll see later why the following is an appropriate assignment of probabilities to the four outcomes.

value of X	**0**	**1**	**2**	**3**
probability	.579	.347	.069	.005

1. Probability that X is 2 or 3:
2. Probability that X is not 0:

Sec. 4.2. Probability models

Finite sample spaces with equally likely outcomes:

Example 7: SRS of size 2 from a set of 5 people (1,2,3,4,5).

S has 10 ***equally likely*** members:

{1,2}{1,3}{1,4}{1,5}{2,3}{2,4}{2,5}{3,4}{3,5}{4,5}

So each is assigned a probability of _____.

Event *A* = "2 is chosen but 3 is not." Find *P(A)*

Event *B* = "either 2 or 3 is chosen." Find *P(B)*

P(A) = 3/10, *P(B)*= 7/10

Sec. 4.2. Probability models (pgs 244-245)

Independence of events:

Events A and B are *independent* if knowing that one occurs does not change the probability that the other occurs.

Rule: Events A and B are independent if and only if

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

2 ways independence can occur:

- **If the events are physically separate (e.g. coin flips)**
- **If the *information* given by one event is unrelated to the second**

Sec. 4.2. Probability models

Independence of events example:

You are a CIA agent trying to catch two criminals. One criminal is on a plane and one is on a boat. Two planes have just taken off, and two boats have just left port. Your operatives have an equal chance of stopping any one of the 4 vehicles.

Event A: You stop a boat.

Event B: You catch one of the criminals.

Independence of events example:

2 boats, 2 planes, 2 criminals, one with each *type* of vehicle.

Event A: You stop a boat.

Event B: You catch one of the criminals.

Intuitively: These events are independent because your chances of getting a criminal are a coin flip whether or not you know the vehicle type.

Sec. 4.2. Probability models

Independence of events example:

2 boats, 2 planes, 2 criminals, one with each *type* of vehicle.

Event A: You stop a boat.

Event B: You catch one of the criminals.

Theoretically:

P(A) = 1 / 2

P(B) = 1 / 2

P(A and B) = P(you catch the boat with the criminal) = ¼ = P(A)P(B)

The rule is satisfied, so the events are independent.

Physical *independence example (easier):*

Example 8: Roll one fair die twice. A = "6 on first roll" and B = "6 on second roll."

We assume the outcome of one roll can't affect the probabilities for another roll (the die has no memory).

So events A and B are independent.

1. Describe the event (A and B).
2. Find $P(A$ and $B)$.

$P(A)$ = ______ and $P(B)$ = _______.

So, ***because A and B are independent***,

$P(A$ and $B) = 1/36$.

Sec. 4.2. Probability models

Independence and the multiplication rule:

Example: 5 people labeled 1,2,3,4,5. Take a SRS of size 2.

S is {12, 13, 14, 15, 23, 24, 25, 34, 35, 45}.

Let A = "1 is chosen" and B = "2 is chosen." We want to find $P(A \text{ and } B)$.

$P(A) = 4/10$ and $P(B) = 4/10$, but are they independent?

No – if A occurs, B's probability decreases a little.

So we **can't** find P(A and B) using Rule 5.

If we could, it would be $P(A)P(B)$ = ______.

But what is $P(A \text{ and } B)$?

$P(A \text{ and } B)=0.1$, so $P(A \text{ and } B) \neq P(A)P(B)$

Sec. 4 More examples like ones from this lecture:

Probability models: Examples 4.4, 4.5, 4.6.

Probability calculations: Examples 4.8 – 4.11

More probability calculations: Examples 4.14 – 4.15

Independence: Examples 4.16 – 4.18