Great Theoretical Ideas In Computer Science

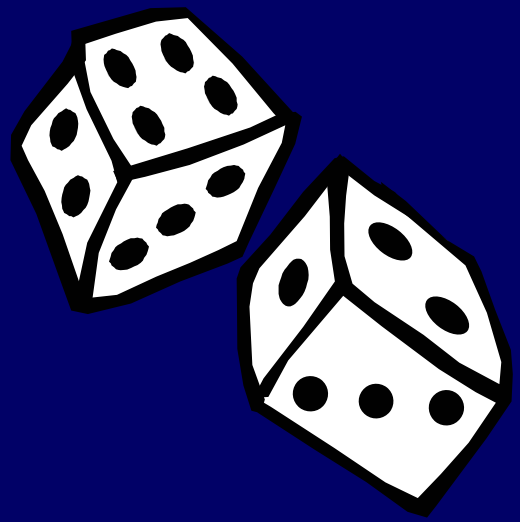
Steven Rudich, Anupam Gupta

CS 15-251 Spring 2004

Lecture 19 March 23, 2004

Carnegie Mellon University

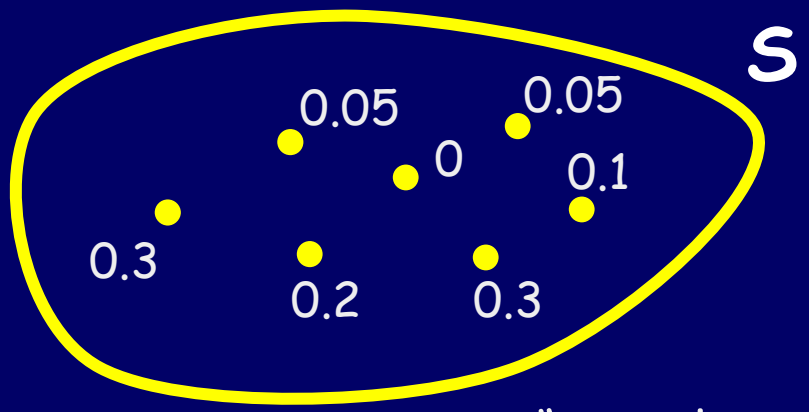
Probability Theory: Paradoxes and Pitfalls



Probability Distribution

A (finite) probability distribution D

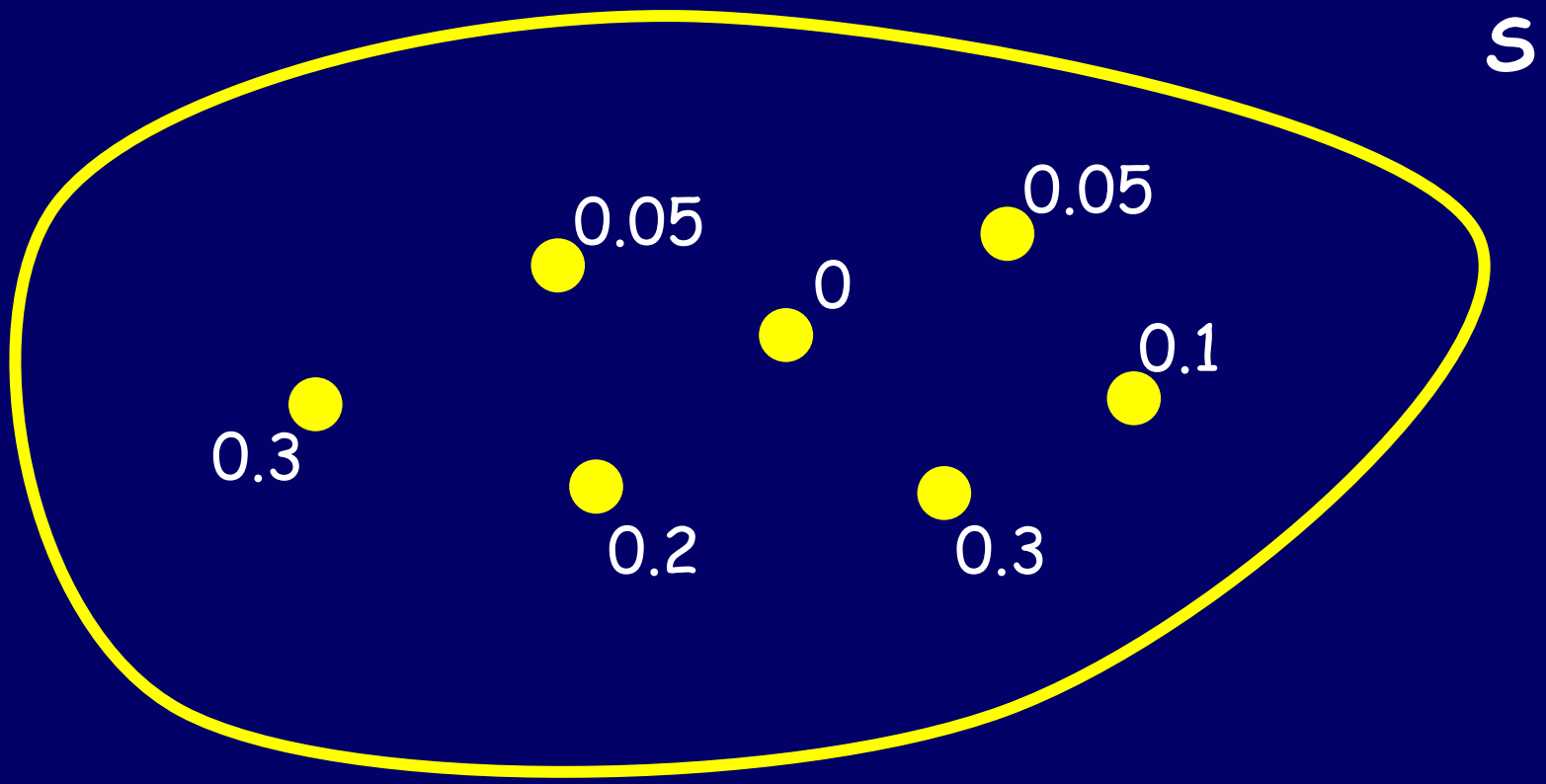
- a finite set S of elements (samples)
- each $x \in S$ has probability $p(x) \in [0,1]$



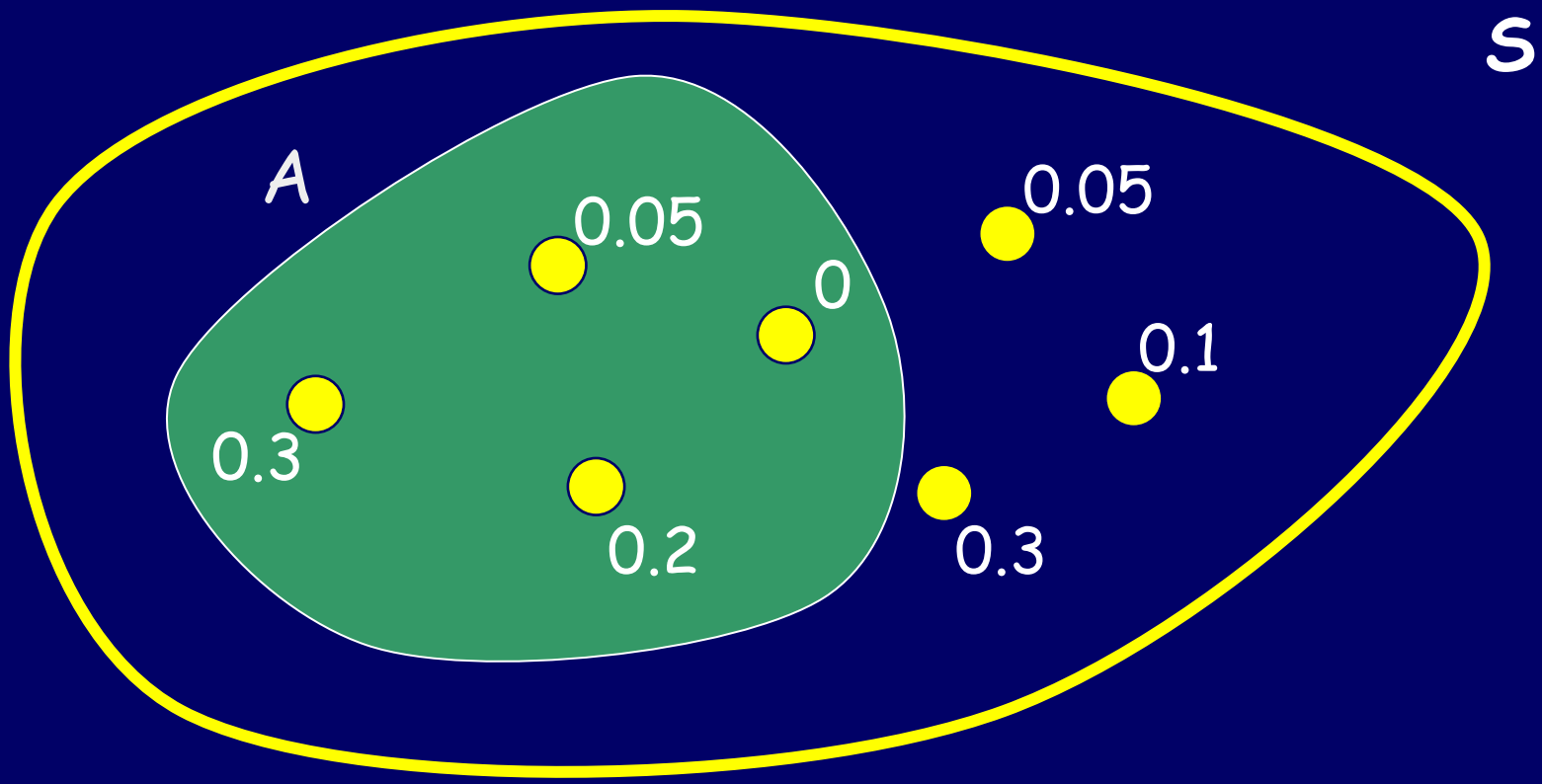
weights must sum to 1

"Sample space"

Probability Distribution

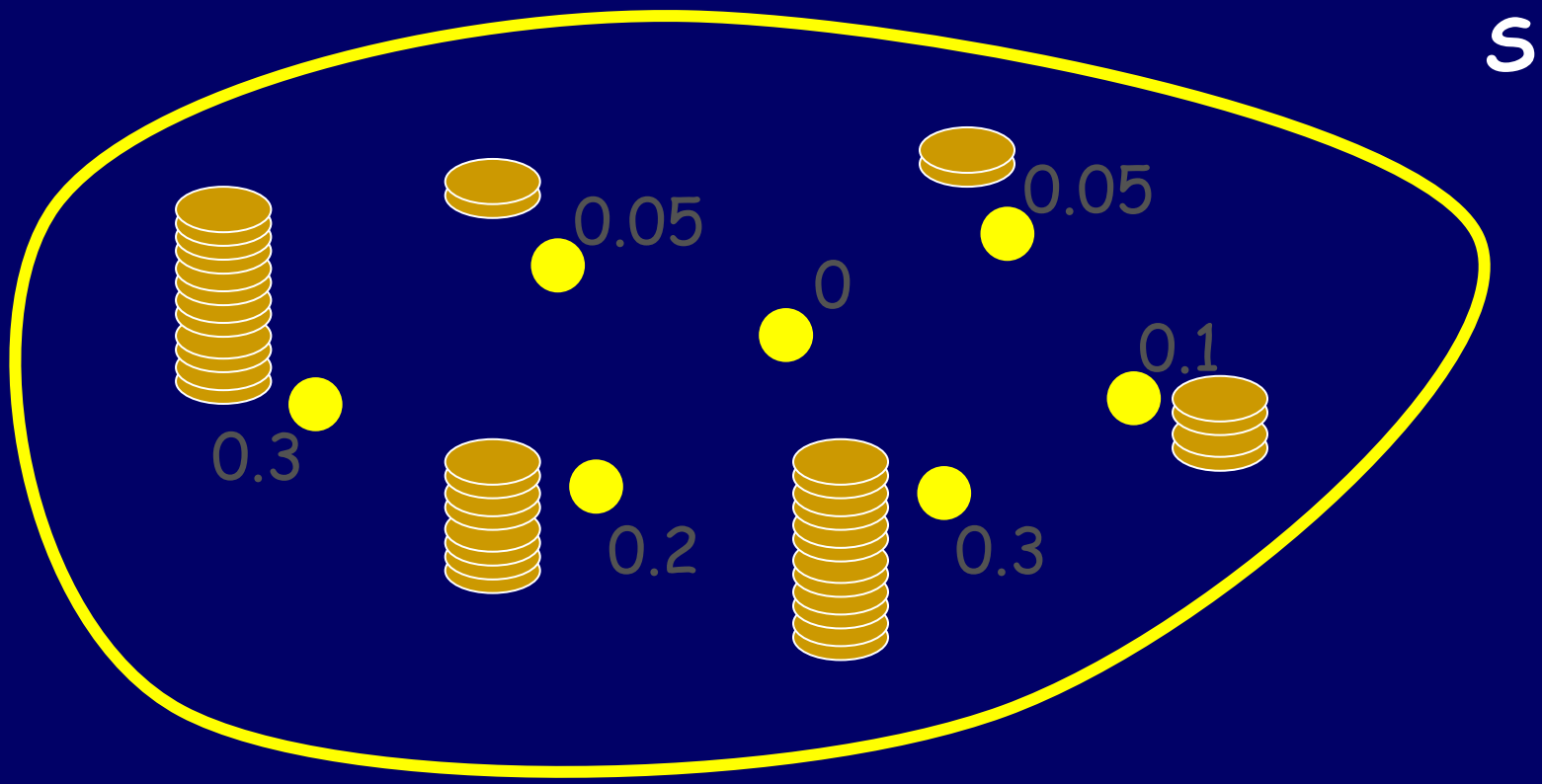


An "Event" is a subset



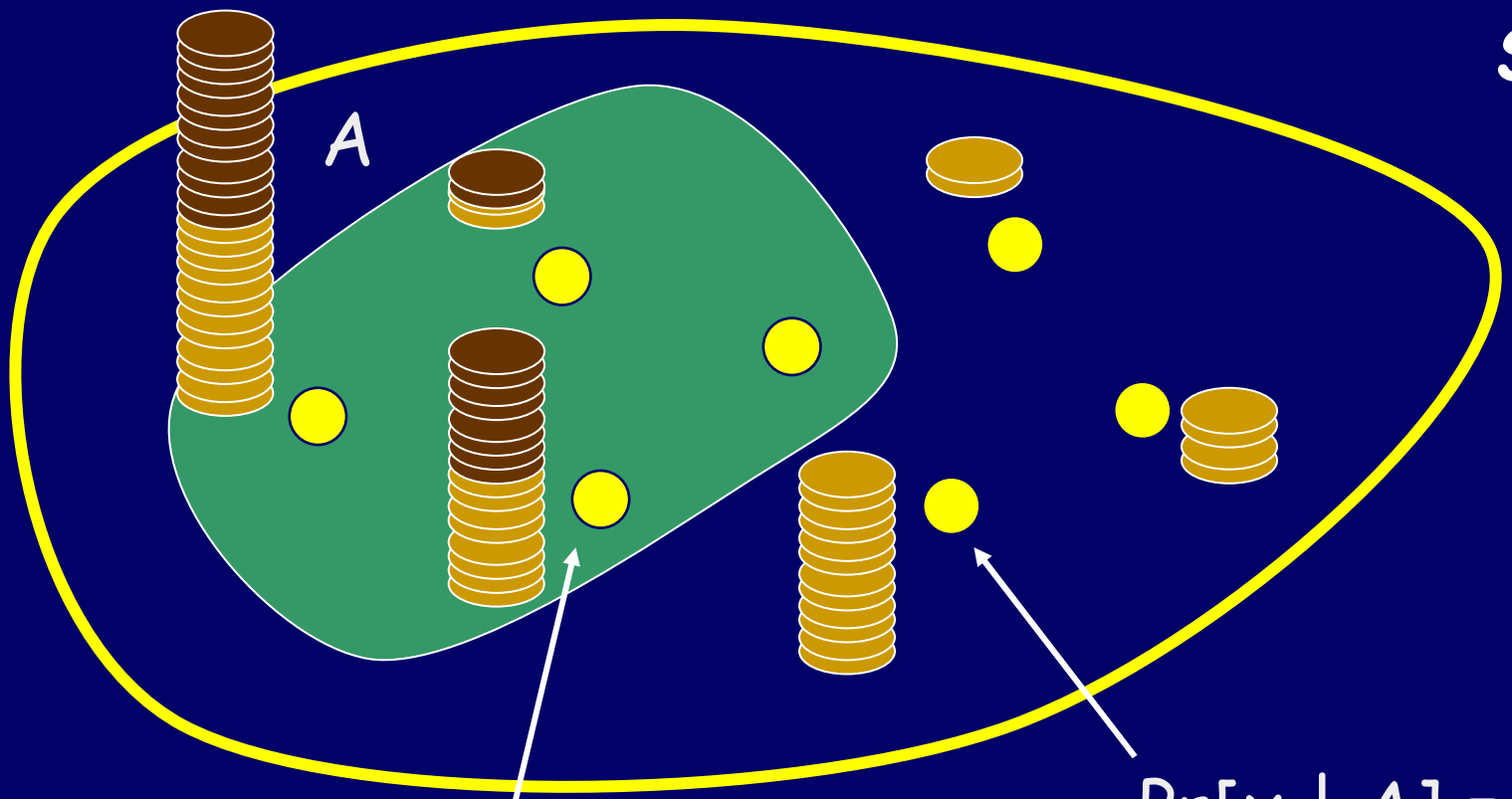
Pr[A] = 0.55

Probability Distribution



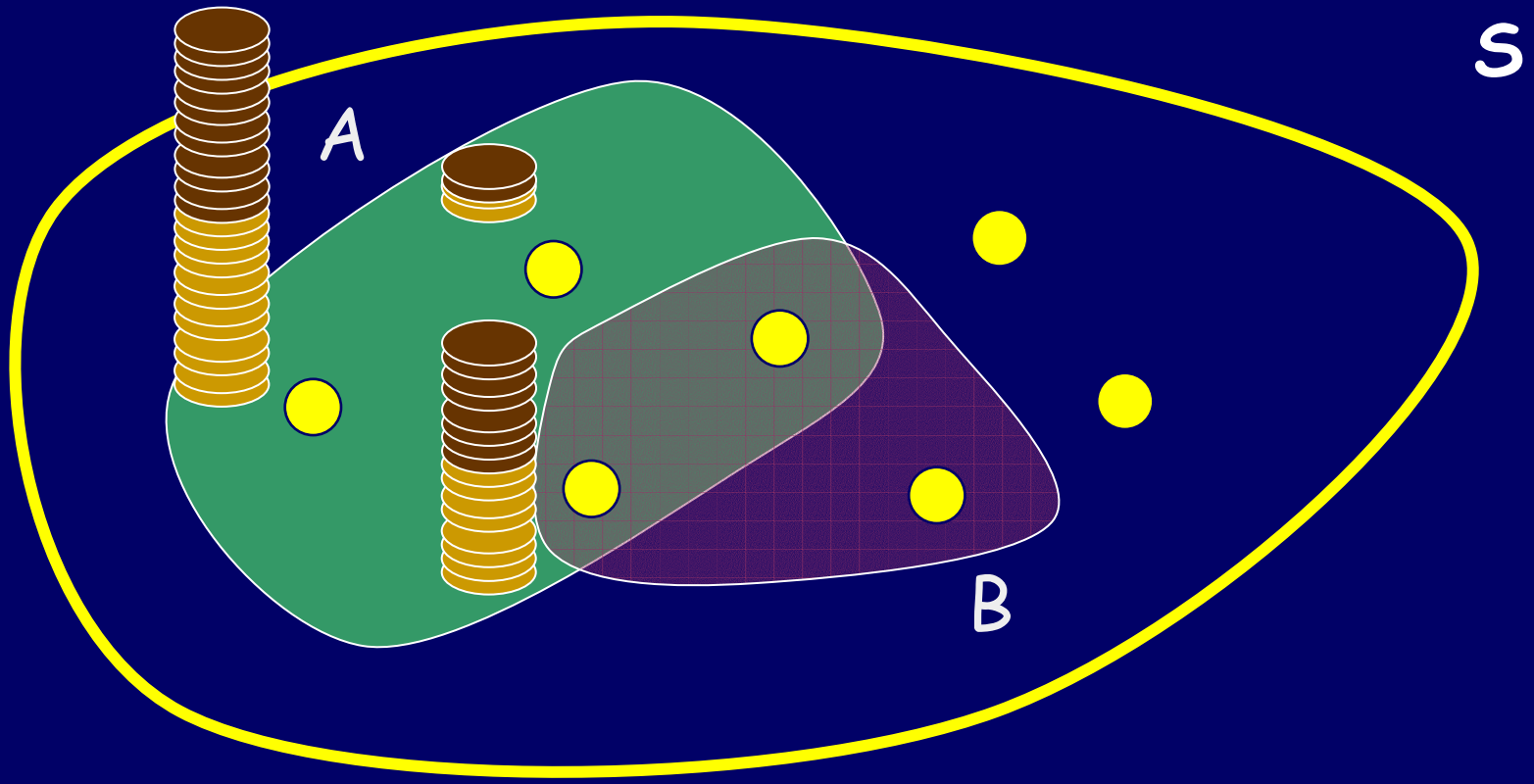
Total money = 1

Conditional probabilities



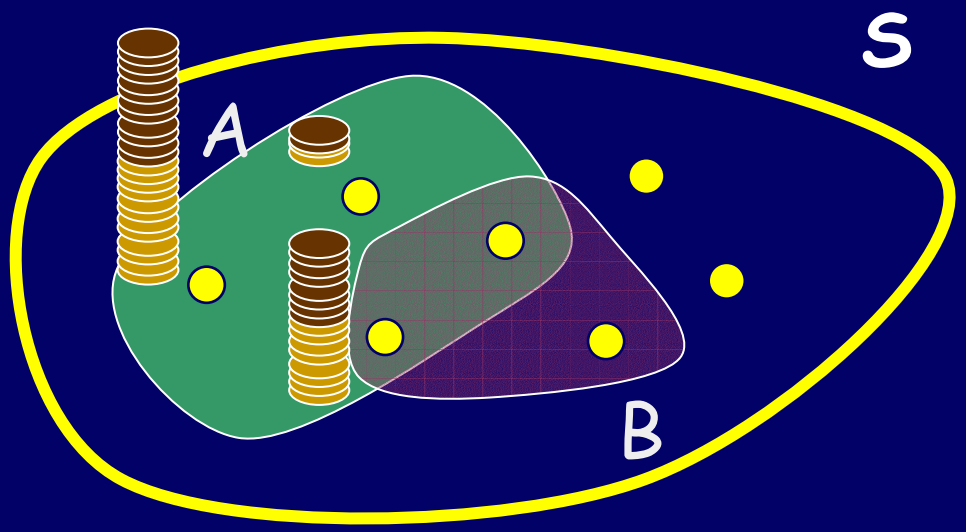
Pr[y | A] = Pr[y] / Pr[A]

Conditional probabilities



$$\Pr[B \mid A] = \sum_{x \in B} \Pr[x \mid A]$$

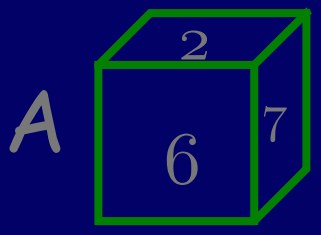
Conditional probabilities

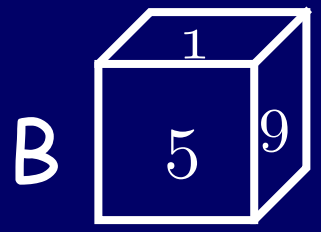


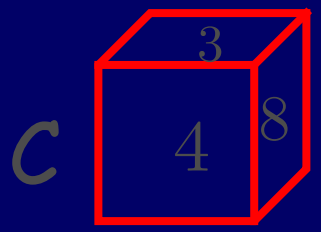
$$\Pr[B \mid A] = \sum_{x \in B} \Pr[x \mid A]$$
$$= \sum_{x \in A \cap B} \Pr[x \mid A]$$
$$= \sum_{x \in A \cap B} \Pr[x] / \Pr[A]$$
$$= \Pr[A \cap B] / \Pr[A]$$

Now, on to some fun puzzles!

You have 3 dice



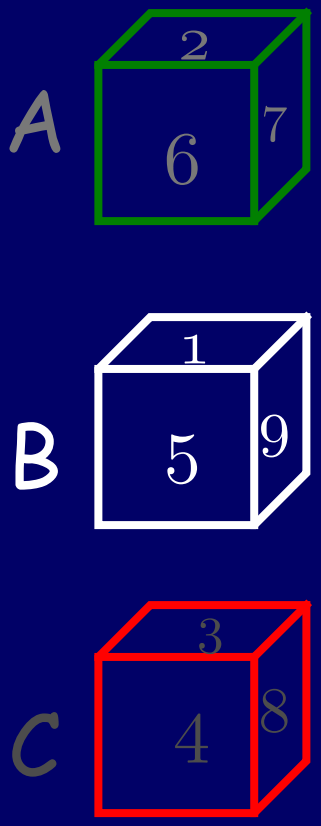




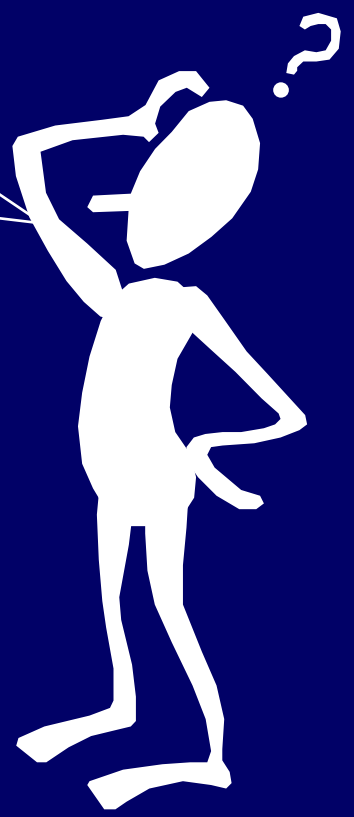
2 Players each rolls a die.

The player with the higher number wins

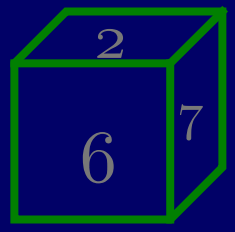
You have 3 dice

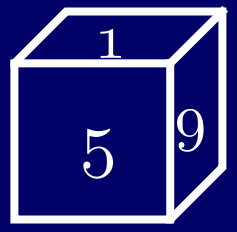


Which die is best to have – A, B, or C ?



A is better than B



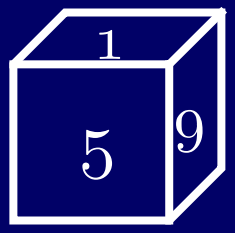


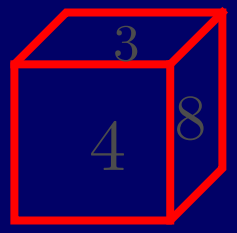
When rolled, 9 equally likely outcomes

2 1	2 5	2 9
6 1	6 5	6 9
7 1	7 5	7 9

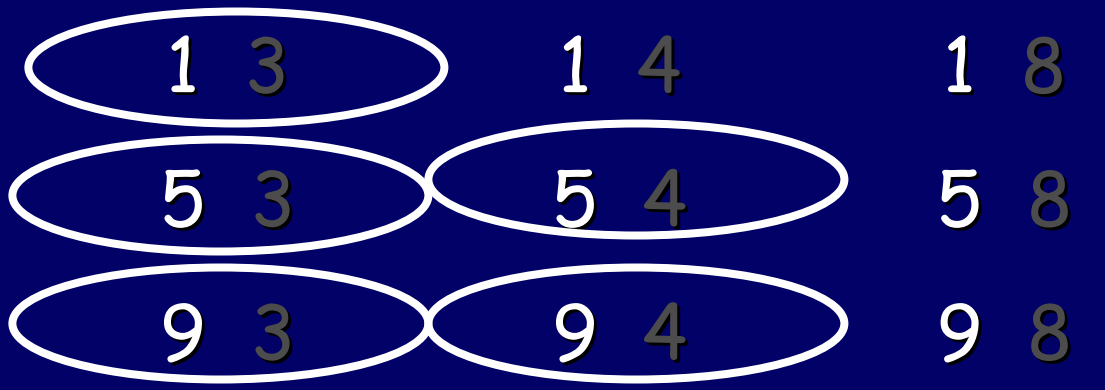
A beats B 5/9 of the time

B is better than C





Again, 9 equally likely outcomes



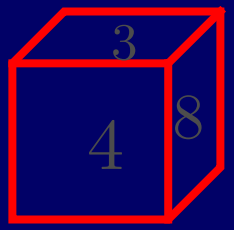
B beats C 5/9 of the time

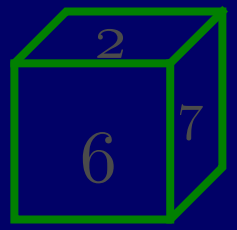
A beats B with Prob. 5/9
B beats C with Prob. 5/9

Q) If you chose first, which die would you take?

Q) If you chose second, which die would you take?

C is better than A!

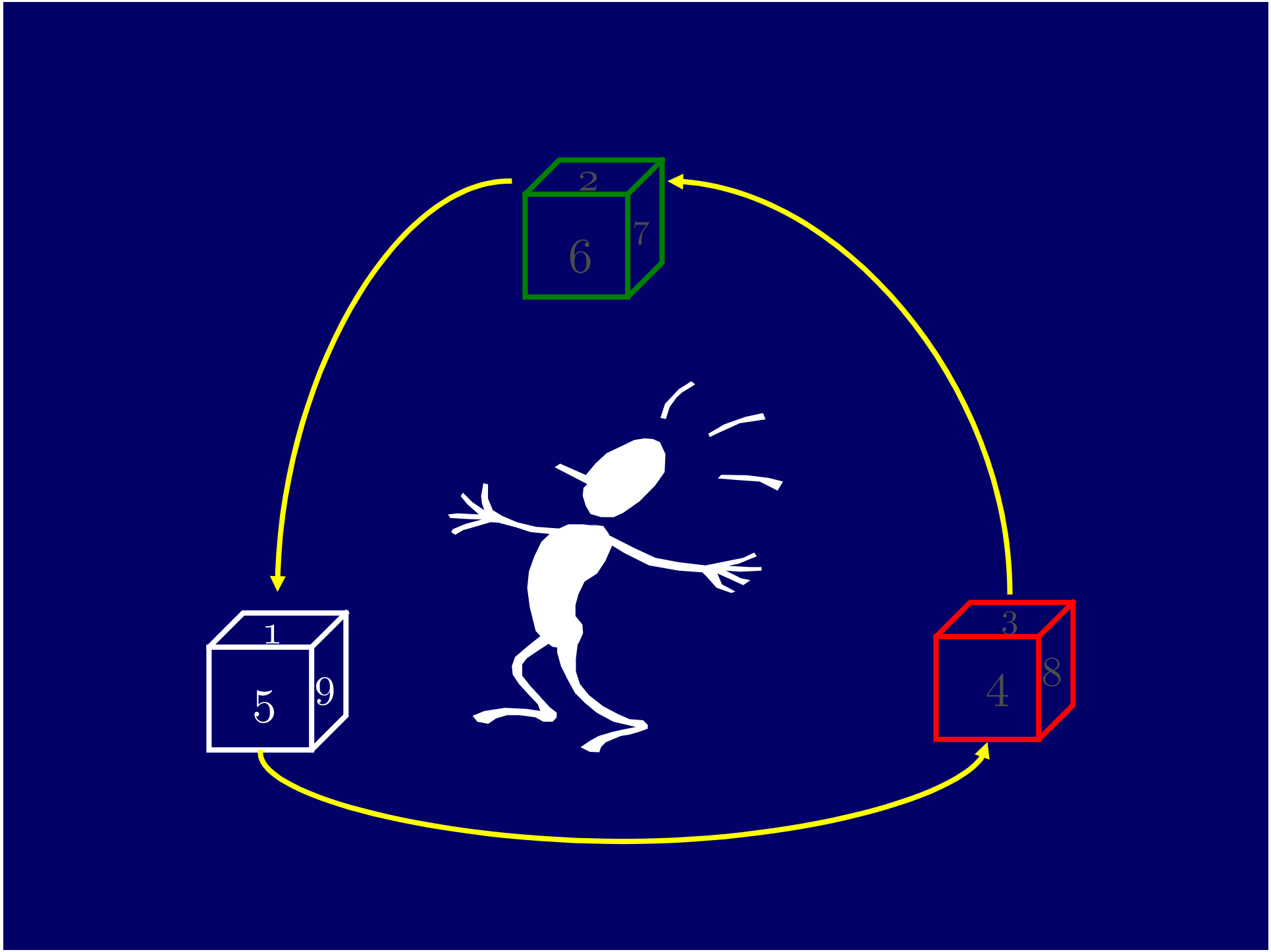




Alas, the same story!

3 2	3 6	3 7
4 2	4 6	4 7
8 2	8 6	8 7

C beats A 5/9 of the time!


2
6
7
1
5
9
3
4
8

First Moral

"Obvious" properties, such as transitivity, associativity, commutativity, etc…
need to be rigorously argued.

Because sometimes they are
FALSE.

Second Moral

When reasoning about probabilities….

Stay on your toes!

Third Moral

To make money from a sucker in a bar, offer him the first choice of die.

(Allow him to change to your "lucky" die any time he wants.)

Coming up next…

More of the pitfalls of probability.

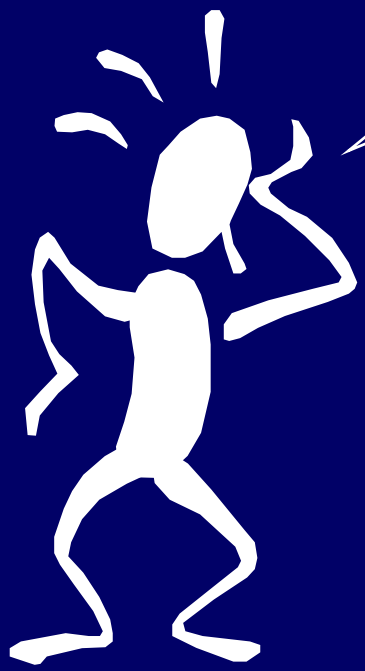
A Puzzle…

Name a body part that almost everyone on earth had an above average number of.

FINGERS !!

- Almost everyone has 10
- More people are missing some than have extras (# fingers missing > # of extras)
- Average: 9.99 …

Almost everyone
can be
above average!



Is a simple average a good statistic?

Several years ago Berkeley faced a law suit …

1. % of male applicants admitted to graduate school was 10%
2. % of female applicants admitted to graduate school was 5%

Grounds for discrimination?

SUIT

Berkeley did a survey of its departments to find out which ones were at fault

The result was

SHOCKING…

Every department was more likely to admit a female than a male

$$\frac{\#\text{of females accepted to department X}}{\#\text{of female applicants to department X}} > \frac{\#\text{of males accepted to department X}}{\#\text{of male applicants to department X}}$$

How can this be ?

Answer

Women tend to apply to departments that admit a smaller percentage of their applicants

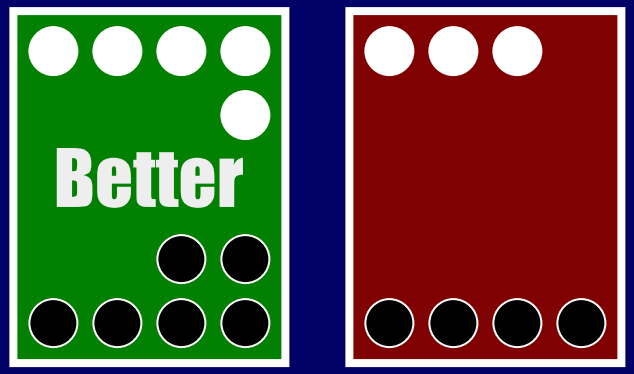
	Women		Men	
Dept	Applied	Accepted	Applied	Accepted
A	99	4	1	0
B	1	1	99	10
total	100	5	100	10

Newspapers would publish these data…

Meaningless junk!

A single summary statistic
(such as an average, or a median)
may not summarize the data well !

Try to get a white ball

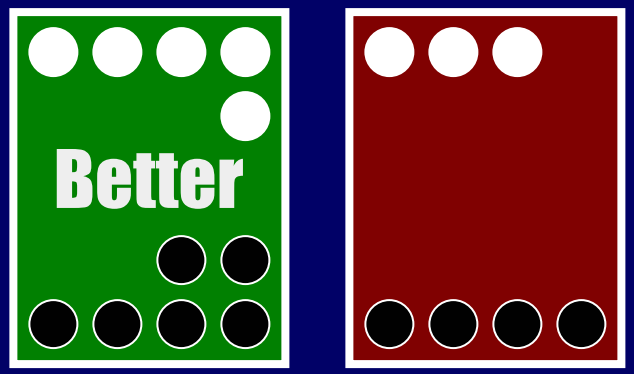


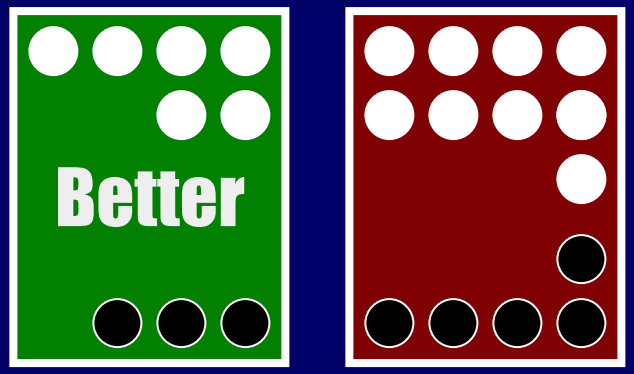
Choose one box and pick a random ball from it.

Max the chance of getting a white ball…

5/11 > 3/7

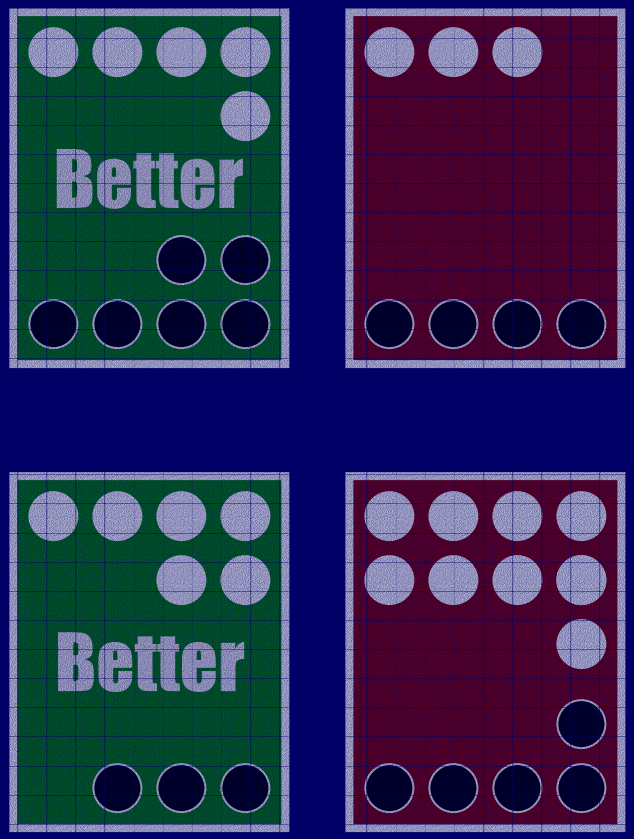
Try to get a white ball

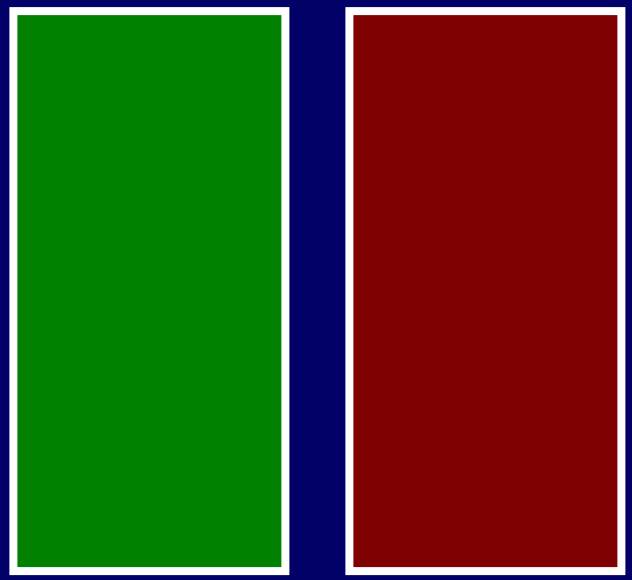




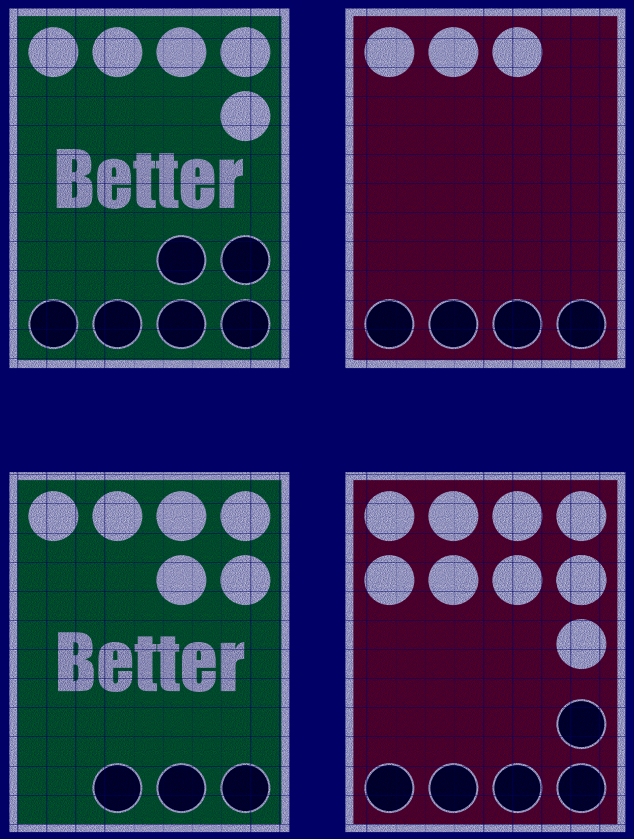
6/9 > 9/14

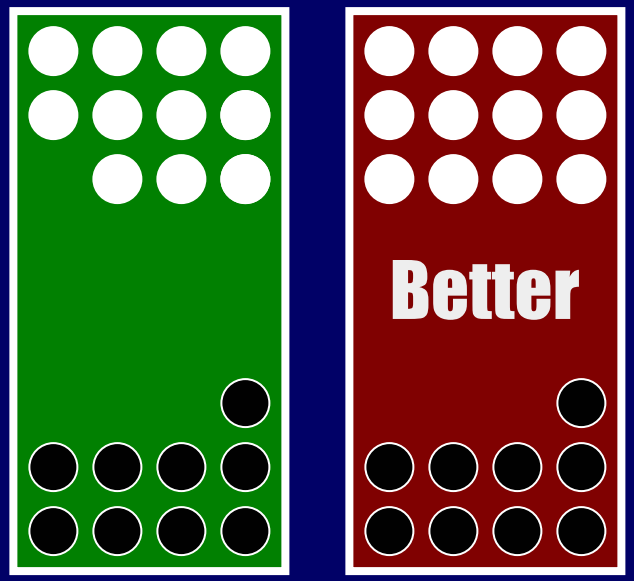
Try to get a white ball





Try to get a white ball





11/20 < 12/21 !!!

Simpson's Paradox

Arises all the time…

Be careful when you interpret numbers

Department of Transportation requires that each month all airlines report their "on-time record"

$$\frac{\text{\# of on-time flights landing at nation's 30 busiest airports}}{\text{\# of total flights into those airports}}$$

http://www.bts.gov/programs/oai/

Different airlines serve different airports with different frequency

An airline sending most of its planes into fair weather airports will crush an airline flying mostly into foggy airports

It can even happen that an airline has a better record at each airport, but gets a worse overall rating by this method.

	Alaska airlines		America West	
	% on time	# flights	% on time	# flights
LA	88.9	559	85.6	811
Phoenix	94.8	233	92.1	5255
San Diego	91.7	232	85.5	448
SF	83.1	605	71.3	449
Seattle	85.8	2146	76.7	262
OVERALL	86.7	3775	89.1	7225

Alaska Air beats America West at <u>each</u> airport but America West has a better overall rating!

An average may have several different possible explanations…

US News and World Report ('83)

	# Doctors	Average salary (1982)
1970	334,000	$103,900
1982	480,000	$99,950

"Physicians are growing in number, but not in pay"

Thrust of article:

Market forces are at work

Here's another possibility

Doctors earn more than ever.

But many old doctors have retired and been replaced with younger ones.

Rare diseases

Rare Disease

A person is selected at random and
given test for rare disease "painanosufulitis".

Only 1/10,000 people have it.

The test is 99% accurate: it gives the wrong answer (positive/negative) only 1% of the time.

The person tests POSITIVE!!!

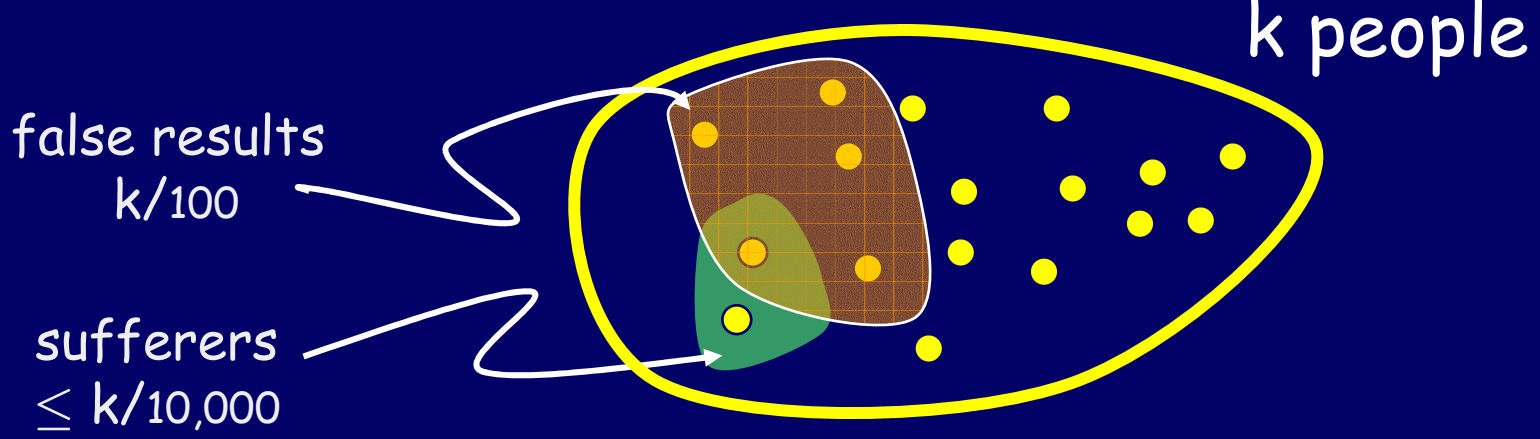
Does he have the disease?

What is the probability that he has the disease?

Disease Probability

- Suppose there are k people in the population
- At most k/10,000 have the disease
- But k/100 have false test results

So ≥ k/100 – k/10,000 have false test results but have no disease!



It's about 100 times more likely that he got a false positive!!

And we thought 99% accuracy was pretty good.

Conditional Probabilities

You walk into a pet shop…

Shop A: there are two parrots in a cage
The owner says "At least one parrot is male."

What is the chance that you get two males?

Shop B: again two parrots in a cage
The owner says "The darker one is male."

Pet Shop Quiz

Shop owner A says "At least one of the two is male"

What is the chance they are both male?

~~FF~~
FM
MF
MM

1/3 chance they are both male

Shop owner B says "The dark one is male"

~~FF~~
~~FM~~
MF
MM

1/2 chance they are both male

Intuition in probability

Playing Alice and Bob

you beat Alice with probabilty 1/3
you beat Bob with probability 5/6

You need to win two consecutive games out of 3.

Should you play

Bob Alice Bob or Alice Bob Alice?

Look closely

To win, we need
win middle game
win one of {first, last} game.

$\Rightarrow$ must beat second player (for sure)
must beat first player once in two tries.

Should you play
Bob Alice Bob or Alice Bob Alice?

Playing Alice and Bob

Bob Alice Bob:

Pr[{WWW, WWL, LWW}]

$= {}^{1}/_{3}\,(1 - {}^{1}/_{6} * {}^{1}/_{6}) = 35/108.$

Alice Bob Alice:

Pr[{WWW, WWL, LWW}]

$= {}^{5}/_{6}\,(1 - {}^{2}/_{3} * {}^{2}/_{3}) = 50/108$

Bridge Hands have 13 cards

What distribution of the 4 suits is most likely?

5 3 3 2 ?

4 4 3 2 ?

4 3 3 3 ?

4 3 3 3

$$\binom{13}{4} \bullet \binom{13}{3}^3 \bullet 4$$

4 4 3 2

$$\binom{13}{4}^2 \bullet \binom{13}{3} \bullet \binom{13}{2} \bullet 4 \bullet 3 = \#(4333)\frac{10}{4}\frac{3}{11}3$$

5 3 3 2

$$\binom{13}{5} \bullet \binom{13}{3}^2 \bullet \binom{13}{2} \bullet 4 \bullet 3 = \#(4432)\frac{4}{105}\frac{9}{}$$

Intuition could be wrong

Work out the math to be 100% sure

"Law of Averages"

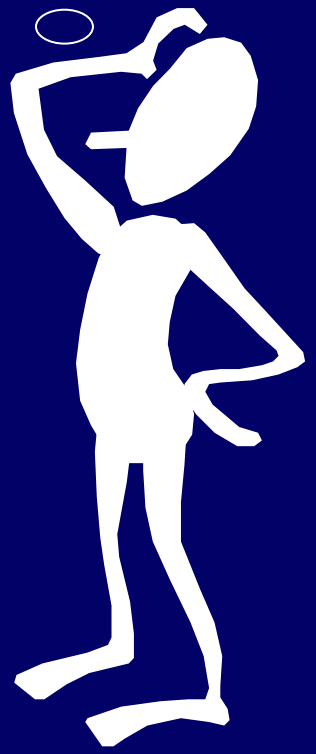
I flip a coin 10 times.
It comes up heads each time!

What are the chances that
my next coin flip is also heads?

"Law of Averages"?

"The number of heads and tails have to even out…"

Be Careful



Though the sample average gets closer to $\frac{1}{2}$,
the deviation from the average may grow!

After 100: 52 heads, sample average 0.52
deviation = 2
After 1000: 511 heads, sample average 0.511
deviation = 11
After 10000: 5096 heads, sample average 0.5096
deviation = 96

A voting puzzle

N (odd) people, each of whom has a random bit (50/50) on his/her forehead.

No communication allowed. Each person goes to a private voting booth and casts a vote for 1 or 0.

If the outcome of the election coincided with the parity of the N bits, the voters "win" the election

A voting puzzle

Example:

N = 5, with bits 1 0 1 1 0

Parity = 1

If they vote 1 0 0 1 1, then majority = 1, they win.
If they vote 0 0 1 1 0, then majority = 0, they lose.

A voting puzzle

N (odd) people, each of whom has a random bit on his/her forehead.

No communication allowed. Each person goes to a private voting booth and casts a vote for 1 or 0.

If the outcome of the election coincided with the parity of the N bits, the voters "win" the election.

How do voters maximize the probability of winning?

Note that each individual has no information about the parity

Since each individual is wrong half the time, the outcome of the election is wrong half the time

Beware of the Fallacy!

Solution

Note: to know parity is equivalent to knowing the bit on your forehead

STRATEGY:

Each person assumes the bit on his/her head is the same as the majority of bits he/she sees.

Vote accordingly

(in the case of even split, vote 0).

Analysis

STRATEGY: Each person assumes the bit on his/her head is the same as the majority of bits he/she sees. Vote accordingly (in the case of even split, vote 0).

Two cases:

- difference of (# of 1's) and (# of 0's) > 1
- difference = 1

Analysis

STRATEGY: Each person assumes the bit on his/her head is the same as the majority of bits he/she sees. Vote accordingly (in the case of even split, vote 0).

ANALYSIS: The strategy works so long as the difference in the number of 1's and the number of 0's is at least two.

Probability of winning = $1-\binom{N}{N/2}\Big/2^N = 1-\frac{1}{O(\sqrt{N})}$

A Final Game

Greater or Smaller?

Alice and Bob play a game

Alice picks two distinct random numbers x and y between 0 and 1

Bob chooses to know any one of them, say x

Now, Bob has to tell whether $x < y$ or $x > y$

If Bob guesses at random,
chances of winning are 50%

Can Bob improve his chances of winning?

Bob picks a number between 0 and 1 at random, say z.

If $x > z$, he says x is greater

If $x < z$, he says x is smaller

Analysis

If z lies between x and y, Bob's answer is correct

Analysis

If z lies between x and y, Bob's answer is correct

If z does not lie between x and y, Bob's answer is wrong 50% of the times.

Since x and y are distinct, there is a non-zero probability for z to lie between x and y

Hence, Bob's probability of winning is more than 50%

Final Lesson for today…

Keep your mind open towards new possibilities !