MTH 2215
Applied Discrete Mathematics

Chapter 6.1
Discrete Probability

These class notes are based on material from our textbook, **Discrete Mathematics and Its Applications**, 6^{th} ed., by Kenneth H. Rosen, published by McGraw Hill, Boston, MA, 2006. They are intended for classroom use only and are **not** a substitute for reading the textbook.

Why Probability?

- In the real world, we often do not know whether a given proposition is true or false.
- Probability theory gives us a way to reason about propositions whose truth is *uncertain*.
- It is useful in weighting evidence, diagnosing problems, and analyzing situations whose exact details are unknown.

Terminology

- Experiment
  - A repeatable procedure that yields one of a given set of outcomes
  - Rolling a die, for example
- Sample space
  - The range of outcomes possible
  - For a die, that would be values 1 to 6
- Event
  - One of the sample outcomes that occurred
  - If you rolled a 4 on the die, the event is the 4

Finite Probability

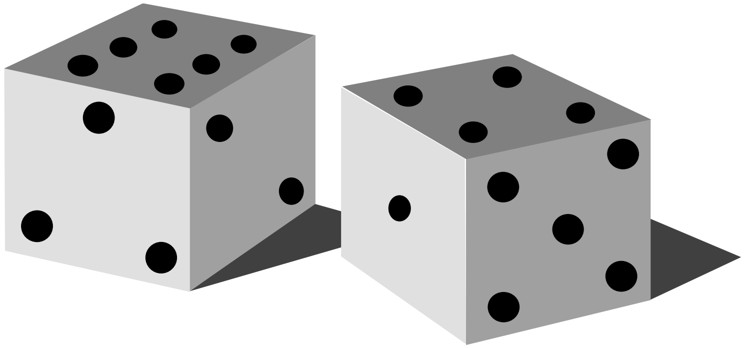
The *probability* of an event E is

$$p(E) = \frac{|E|}{|S|}$$

- Where E is the set of desired events (outcomes)
- Where S is the set of all possible events (outcomes)
- Note that $0 \leq |E| \leq |S|$
  - Thus, the probability will always between 0 and 1
  - An event that will never happen has probability 0
  - An event that will always happen has probability 1

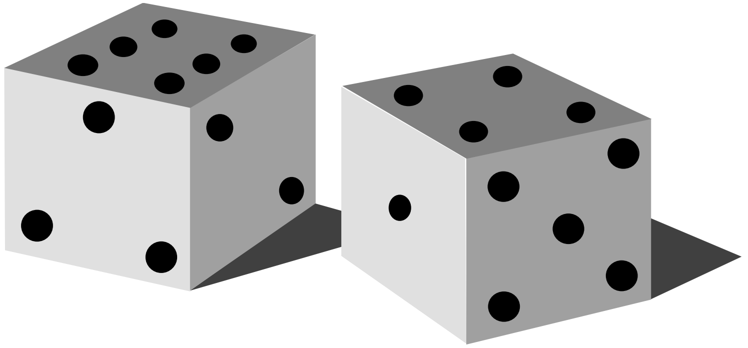
Example

Suppose two dice are rolled. The sample space would be



	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

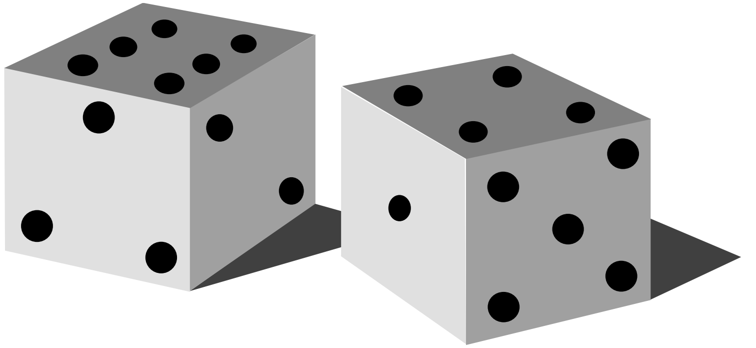
p(sum is 11)



$|S| = 36$

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

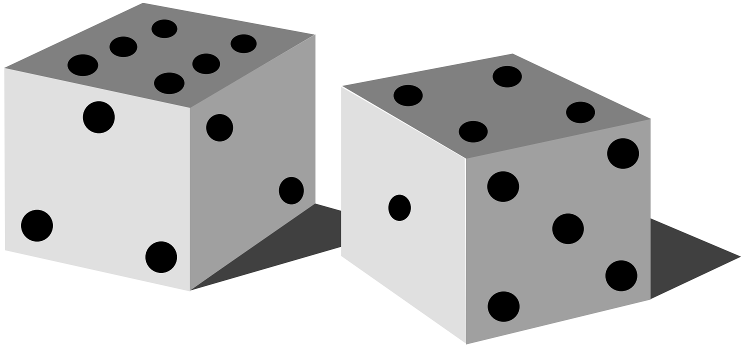
p(sum is 11)



$|S| = 36$

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

p(sum is 11)



$$|S| = 36$$

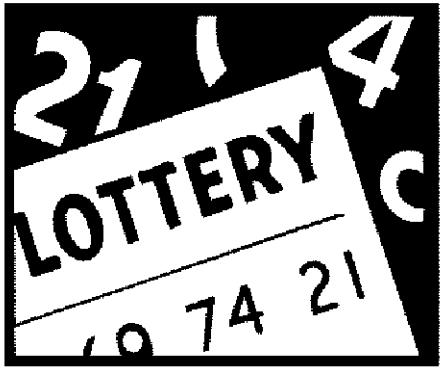
$$|E| = 2$$

$$p(E) = \frac{|E|}{|S|} = \frac{2}{36}$$

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

Dice probability

- What is the probability of getting “snake-eyes” (two 1’s) on two six-sided dice?
  - Probability of getting a 1 on a 6-sided die is 1/6
  - Via product rule, probability of getting two 1’s is the probability of getting a 1 AND the probability of getting a second 1
  - Thus, it’s 1/6 * 1/6 = 1/36
- What is the probability of getting a 7 by rolling two dice?
  - There are six combinations that can yield 7: (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)
  - Thus, $|E| = 6$, $|S| = 36$, $P(E) = 6/36 = 1/6$



12 8 2 39 7 4

Suppose a lottery randomly selects 6 numbers from 40. What is the probability that you selected the correct six numbers? Order is not important.

$|E| = 1$

$|S| = C(40,6)$

$$p(E) = \frac{1}{C(40,6)}$$

$$= \frac{1}{3{,}838{,}380}$$

Combinations of Events

Let E be an event in a sample space S. The probability of the event $\overline{E}$, the complementary event of E, is given by

$$p(\overline{E}) = 1 - p(E)$$

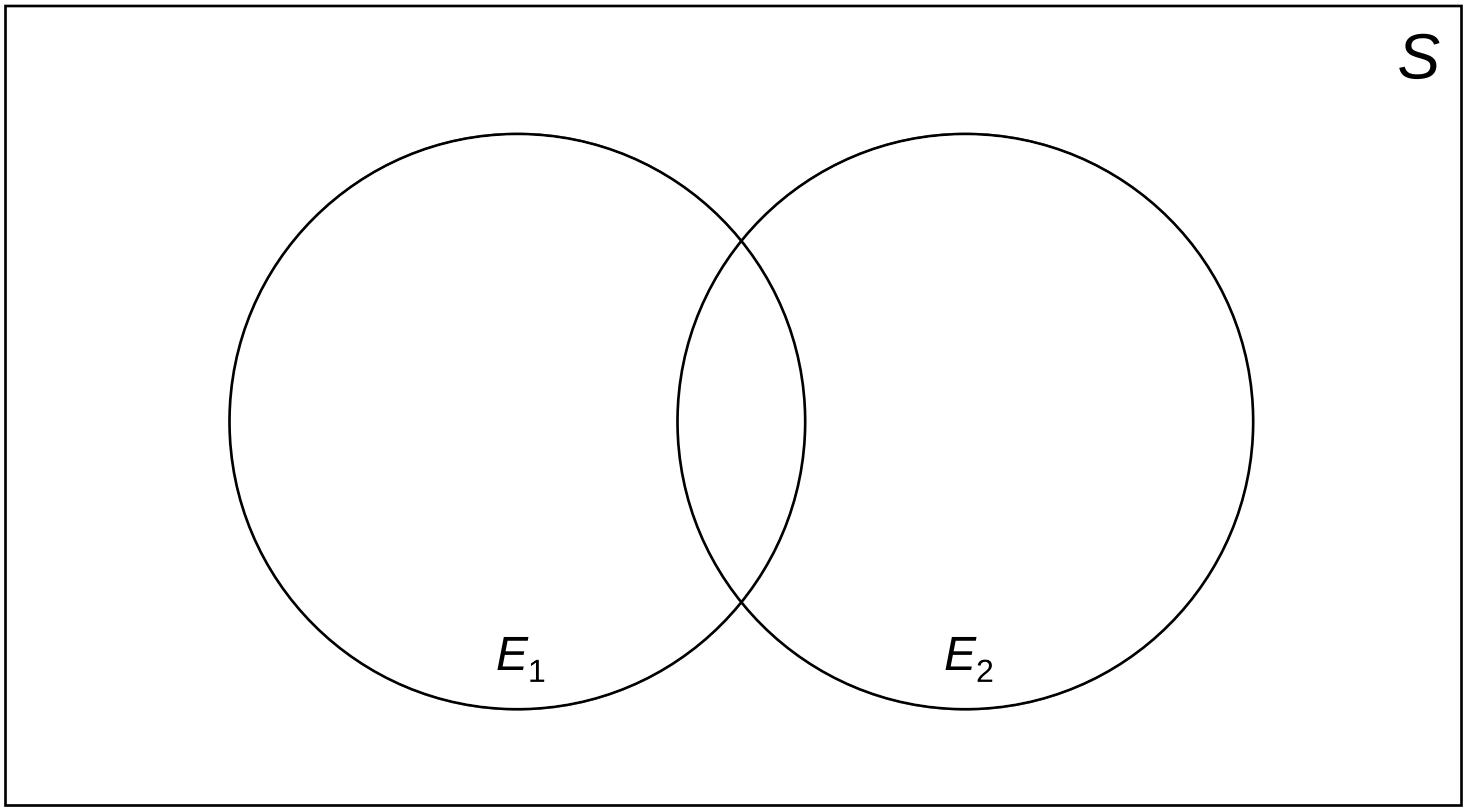
Combinations of Events

Let E_1 and E_2 be events in the sample space S. Then

$$p(E_1 \cup E_2) = p(E_1) + p(E_2) - p(E_1 \cap E_2)$$

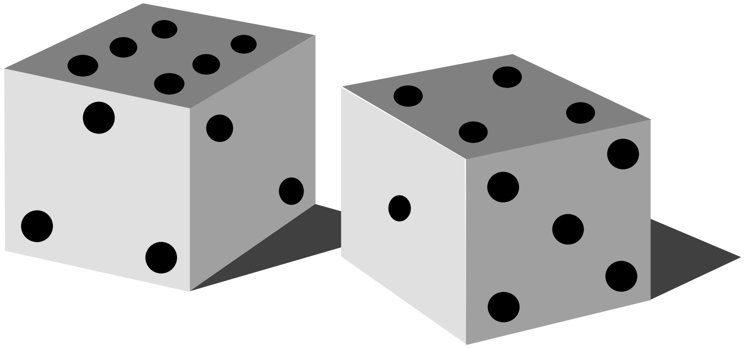
Probability of the union of two events

$p(E1 \cup E2)$



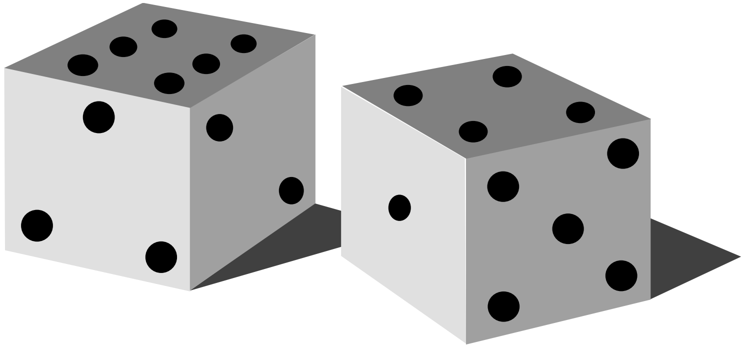
Example

Suppose a red die and a blue die are rolled. The sample space would be



	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

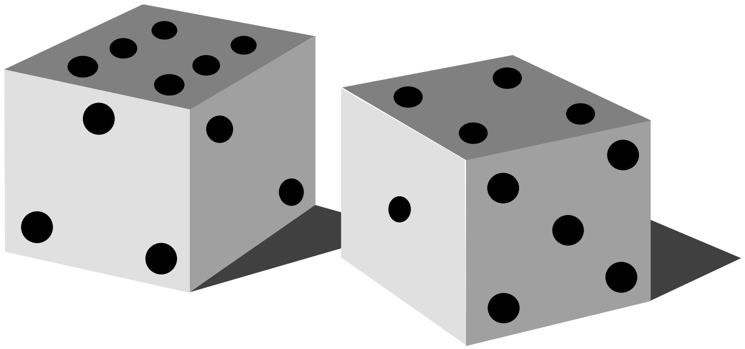
p(sum is 7 or
blue die is 3)



|S| = 36

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

p(sum is 7 or
blue die is 3)



|S| = 36

|sum is 7| = 6

|blue die is 3| = 6

| in intersection | = 1

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

p(sum is 7 or blue die is 3) =

6/36 + 6/36 - 1/36 = 11/36

Probability of the union of two events

- If you choose a number between 1 and 100, what is the probability that it is divisible by 2 or 5 or both?
- Let *n* be the number chosen
  - $p(2|n) = 50/100$ (all the even numbers)
  - $p(5|n) = 20/100$
  - $p(2|n)$ and $p(5|n) = p(10|n) = 10/100$
  - $p(2|n)$ or $p(5|n) = p(2|n) + p(5|n) - p(10|n)$

    $= 50/100 + 20/100 – 10/100$

    $= 3/5$

When is gambling worth it?

- This is a *statistical* analysis, not a moral/ethical discussion

- What if you gamble $1, and have a ½ probability to win $10?
  - If you play 100 times, you will win (on average) 50 of those times
    - Each play costs $1, each win yields $10
    - For $100 spent, you win (on average) $500
  - Average win is $5 (or $10 * ½) per play for every $1 spent
- What if you gamble $1 and have a 1/100 probability to win $10?
  - If you play 100 times, you will win (on average) 1 of those times
    - Each play costs $1, each win yields $10
    - For $100 spent, you win (on average) $10
  - Average win is $0.10 (or $10 * 1/100) for every $1 spent
- One way to determine if gambling is worth it:
  - probability of winning * payout ≥ amount spent
  - Or p(winning) * payout ≥ investment
  - Of course, this is a ***statistical*** measure

When is lotto worth it?

- Many lotto games you have to choose 6 numbers from 1 to 48
  - Total possible choices is C(48,6) = 12,271,512
  - Total possible winning numbers is C(6,6) = 1
  - Probability of winning is 0.0000000814
    - Or 1 in 12.3 million

- If you invest $1 per ticket, it is only statistically worth it if the payout is > $12.3 million
  - As, on the "average" you will only make money that way
  - Of course, "average" will require trillions of lotto plays…

An aside: probability of multiple events

- Assume you have a 5/6 chance for an event to happen
  - Rolling a 1-5 on a die, for example
- What's the chance of that event happening twice in a row?
- Cases:
  - Event happening neither time: 1/6 * 1/6 = 1/36
  - Event happening first time: 1/6 * 5/6 = 5/36
  - Event happening second time: 5/6 * 1/6 = 5/36
  - Event happening both times: 5/6 * 5/6 = 25/36
- For an event to happen twice, the probability is the product of the individual probabilities

An aside: probability of multiple events

- Assume you have a 5/6 chance for an event to happen
  - Rolling a 1-5 on a die, for example
- What's the chance of that event happening at least once?
- Cases:
  - Event happening neither time: 1/6 * 1/6 = 1/36
  - Event happening first time: 1/6 * 5/6 = 5/36
  - Event happening second time: 5/6 * 1/6 = 5/36
  - Event happening both times: 5/6 * 5/6 = 25/36
- It's 35/36!
- For an event to happen at least once, it's 1 minus the probability of it never happening
- Or 1 minus the compliment of it never happening

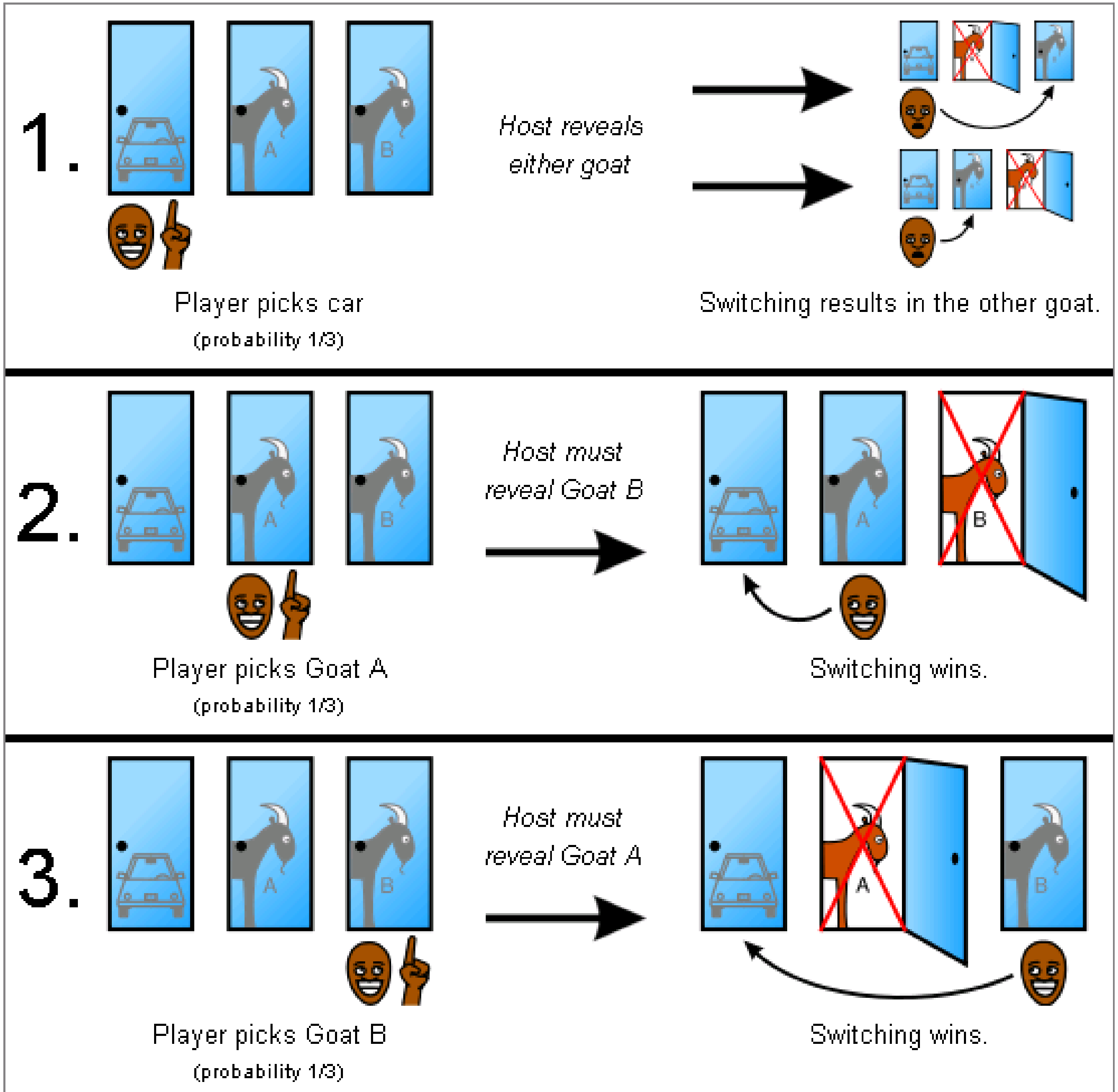
Probability vs. odds

- Consider an event that has a 1 in 3 chance of happening
- Probability is 0.333
- Which is a 1 in 3 chance
- Or 2:1 odds
  - Meaning if you play it 3 (2+1) times, you will lose 2 times for every 1 time you win
- This, if you have x:y odds, you probability is $y/(x+y)$
  - The y is usually 1, and the x is scaled appropriately
  - For example 2.2:1
    - That probability is 1/(1+2.2) = 1/3.2 = 0.313
- 1:1 odds means that you will lose as many times as you win

Monty Hall Paradox

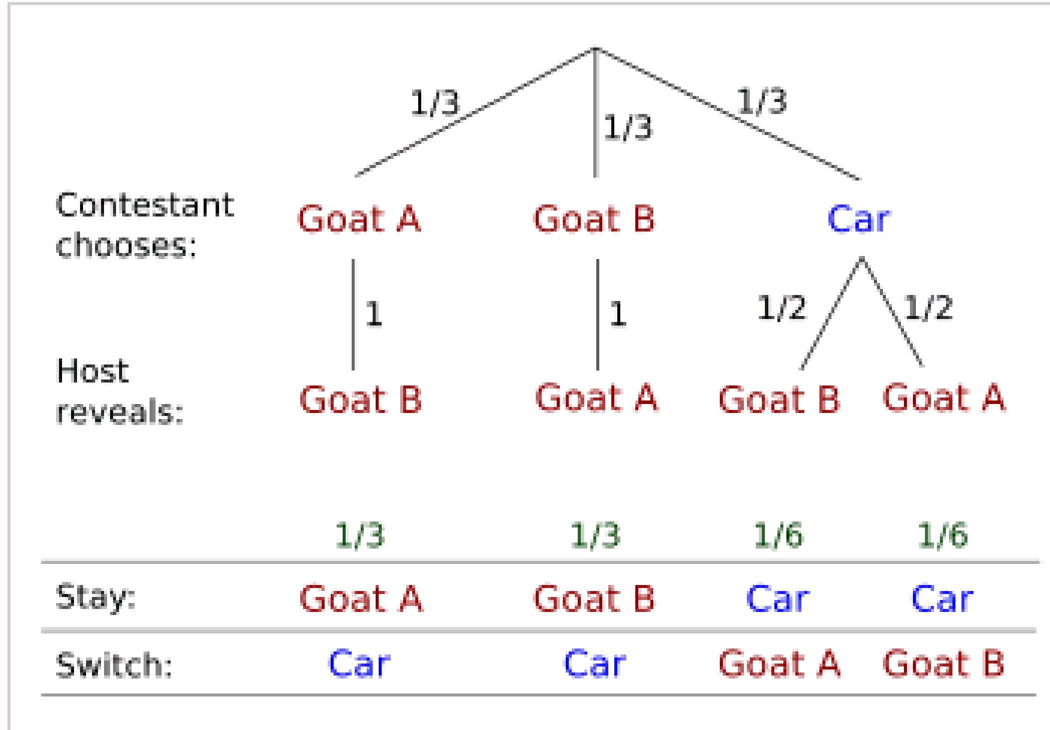
What's behind door number three?

- The Monty Hall problem paradox
  - Consider a game show where a prize (a car) is behind one of three doors
  - The other two doors do not have prizes (goats instead)
  - After picking one of the doors, the host (Monty Hall) opens a different door to show you that the door he opened is not the prize
  - Do you change your decision?
- Your initial probability to win (i.e. pick the right door) is 1/3
- What is your chance of winning if you change your choice after Monty opens a wrong door?
- After Monty opens a wrong door, if you change your choice, your chance of ***winning*** is 2/3
  - Thus, your chance of winning ***doubles*** if you change
  - Huh?



The player has an equal chance of initially selecting the car, Goat A, or Goat B. Switching results in a win 2/3 of the time.

Decision Tree



Tree listing the probability of every possible outcome

In the first two cases, wherein the player has first chosen a goat, switching will yield the car.

In the third and fourth cases, since the player has chosen the car initially, a switch will lead to a goat.

The probability that switching wins is equal to the sum of the first two events:

$$1/3 + 1/3 = 2/3.$$

Likewise, the probability that staying wins is

$$1/6 + 1/6 = 1/3.$$

What's behind door number one hundred?

- Consider 100 doors
  - You choose one
  - Monty opens 98 wrong doors
  - Do you switch?
- Your initial chance of being right is 1/100
- Right before your switch, your chance of being right is still 1/100
  - Just because you know more info about the other doors doesn't change your chances
    - You didn't know this info beforehand!
- Your final chance of being right is 99/100 if you switch
  - You have two choices: your original door and the new door
  - The original door still has 1/100 chance of being right
  - Thus, the new door has 99/100 chance of being right
  - The 98 doors that were opened were not chosen at random!
    - Monty Hall knows which door the car is behind

- Reference: http://en.wikipedia.org/wiki/Monty_Hall_problem

Blackjack

Blackjack

- You are initially dealt two cards
  - 10, J, Q and K all count as 10
  - Ace is EITHER 1 or 11 (player's choice)
- You can opt to receive more cards (a "hit")
- You want to get as close to 21 as you can
  - If you go over, you lose (a "bust")
- You play against the house
  - If the house has a higher score than you, then you lose

Blackjack table

Blackjack probabilities

- Getting 21 on the first two cards is called a blackjack
  - Or a "natural 21"
- Assume there is only 1 deck of cards
- Possible blackjack blackjack hands:
  - First card is an A, second card is a 10, J, Q, or K
    - 4/52 for Ace, 16/51 for the ten card
    - = (4*16)/(52*51) = 0.0241 (or about 1 in 41)
  - First card is a 10, J, Q, or K; second card is an A
    - 16/52 for the ten card, 4/51 for Ace
    - = (16*4)/(52*51) = 0.0241 (or about 1 in 41)
- Total chance of getting a blackjack is the sum of the two:
  - p = 0.0483, or about 1 in 21
  - How appropriate!
  - More specifically, it's 1 in 20.72

Blackjack probabilities

- Another way to get 20.72
- There are C(52,2) = 1,326 possible initial blackjack hands
- Possible blackjack blackjack hands:
  - Pick your Ace: C(4,1)
  - Pick your 10 card: C(16,1)
  - Total possibilities is the product of the two (64)
- Probability is 64/1,326 = 20.72

Blackjack probabilities

- Getting 21 on the first two cards is called a blackjack
- Assume there is an infinite deck of cards
  - So many that the probably of getting a given card is not affected by any cards on the table
- Possible blackjack blackjack hands:
  - First card is an A, second card is a 10, J, Q, or K
    - 4/52 for Ace, 16/52 for second part
    - = (4*16)/(52*52) = 0.0236 (or about 1 in 42)
  - First card is a 10, J, Q, or K; second card is an A
    - 16/52 for first part, 4/52 for Ace
    - = (16*4)/(52*52) = 0.0236 (or about 1 in 42)
- Total chance of getting a blackjack is the sum:
  - p = 0.0473, or about 1 in 21
  - More specifically, it's 1 in 21.13 (vs. 20.72)
- In reality, most casinos use "shoes" of 6-8 decks for this reason
  - It slightly lowers the player's chances of getting a blackjack
  - And prevents people from counting the cards…

So always use a single deck, right?

- Most people think that a single-deck blackjack table is better, as the player's odds increase
  - And you can try to count the cards
- But it's usually not the case!
- Normal rules have a 3:2 payout for a blackjack
  - If you bet $100, you get your $100 back plus 3/2 * $100, or $150 additional
- Most single-deck tables have a 6:5 payout
  - You get your $100 back plus 6/5 * $100 or $120 additional
  - This lowered benefit of being able to count the cards OUTWEIGHS the benefit of the single deck!
    - And thus the benefit of counting the cards
  - You cannot win money on a 6:5 blackjack table that uses 1 deck
  - Remember, the house always wins

Blackjack probabilities: when to hold

- House usually holds on a 17
  - What is the chance of a bust if you draw on a 17? 16? 15?
- Assume all cards have equal probability

- Bust on a draw on a 18
  - 4 or above will bust: that's 10 (of 13) cards that will bust
  - 10/13 = 0.769 probability to bust
- Bust on a draw on a 17
  - 5 or above will bust: 9/13 = 0.692 probability to bust
- Bust on a draw on a 16
  - 6 or above will bust: 8/13 = 0.615 probability to bust
- Bust on a draw on a 15
  - 7 or above will bust: 7/13 = 0.538 probability to bust
- Bust on a draw on a 14
  - 8 or above will bust: 6/13 = 0.462 probability to bust

Buying (blackjack) insurance

- If the dealer's visible card is an Ace, the player can buy insurance against the dealer having a blackjack
  - There are then two bets going: the original bet and the insurance bet
  - If the dealer has blackjack, you lose your original bet, but your insurance bet pays 2-to-1
    - So you get twice what you paid in insurance back
    - Note that if the player also has a blackjack, it's a "push"
  - If the dealer does not have blackjack, you lose your insurance bet, but your original bet proceeds normal

- Is this insurance worth it?

Buying (blackjack) insurance

- If the dealer shows an Ace, there is a 4/13 = 0.308 probability that they have a blackjack
  - Assuming an infinite deck of cards
  - Any one of the "10" cards will cause a blackjack
- If you bought insurance 1,000 times, it would be used 308 (on average) of those times
  - Let's say you paid $1 each time for the insurance
- The payout on each is 2-to-1, thus you get $2 back when you use your insurance
  - Thus, you get 2*308 = $616 back for your $1,000 spent
- Or, using the formula p(winning) * payout ≥ investment
  - 0.308 * $2 ≥ $1
  - 0.616 ≥ $1
  - Thus, it's not worth it
- Buying insurance is considered a very poor option for the player
  - Hence, almost every casino offers it

Blackjack strategy

- These tables tell you the best move to do on each hand
- The odds are still (slightly) in the house's favor
- The house always wins…

	2	3	4	5	6	7	8	9	10	A
17-21	S	S	S	S	S	S	S	S	S	S
16	S	S	S	S	S	H	H	SR	SR	SR
15	S	S	S	S	S	H	H	H	SR	H
14	S	S	S	S	S	H	H	H	H	H
13	S	S	S	S	S	H	H	H	H	H
12	H	H	S	S	S	H	H	H	H	H
11	D	D	D	D	D	D	D	D	D	H
10	D	D	D	D	D	D	D	D	H	H
9	H	D	D	D	D	H	H	H	H	H
5-8	H	H	H	H	H	H	H	H	H	H

	2	3	4	5	6	7	8	9	10	A
A,9	S	S	S	S	S	S	S	S	S	S
A,8	S	S	S	S	S	S	S	S	S	S
A,7	S	D/S	D/S	D/S	D/S	S	S	H	H	H
A,6	H	D/H	D/H	D/H	D/H	H	H	H	H	H
A,5	H	H	D/H	D/H	D/H	H	H	H	H	H
A,4	H	H	D/H	D/H	D/H	H	H	H	H	H
A,3	H	H	H	D/H	D/H	H	H	H	H	H
A,2	H	H	H	D/H	D/H	H	H	H	H	H

	2	3	4	5	6	7	8	9	10	A
A/A	SP	SP	SP	SP	SP	SP	SP	SP	SP	SP
10/10	S	S	S	S	S	S	S	S	S	S
9/9	SP	SP	SP	SP	SP	S	SP	SP	S	S
8/8	SP	SP	SP	SP	SP	SP	SP	SP	SP	SP
7/7	SP	SP	SP	SP	SP	SP	H	H	H	H
6/6	SP/D	SP	SP	SP	SP	H	H	H	H	H
5/5	D	D	D	D	D	D	D	D	H	H
4/4	H	H	H	SP/D	SP/D	H	H	H	H	H
3/3	SP/D	SP/D	SP	SP	SP	SP	H	H	H	H
2/2	SP/D	SP/D	SP	SP	SP	SP	H	H	H	H

VegasReference.com

Why counting cards doesn't work well…

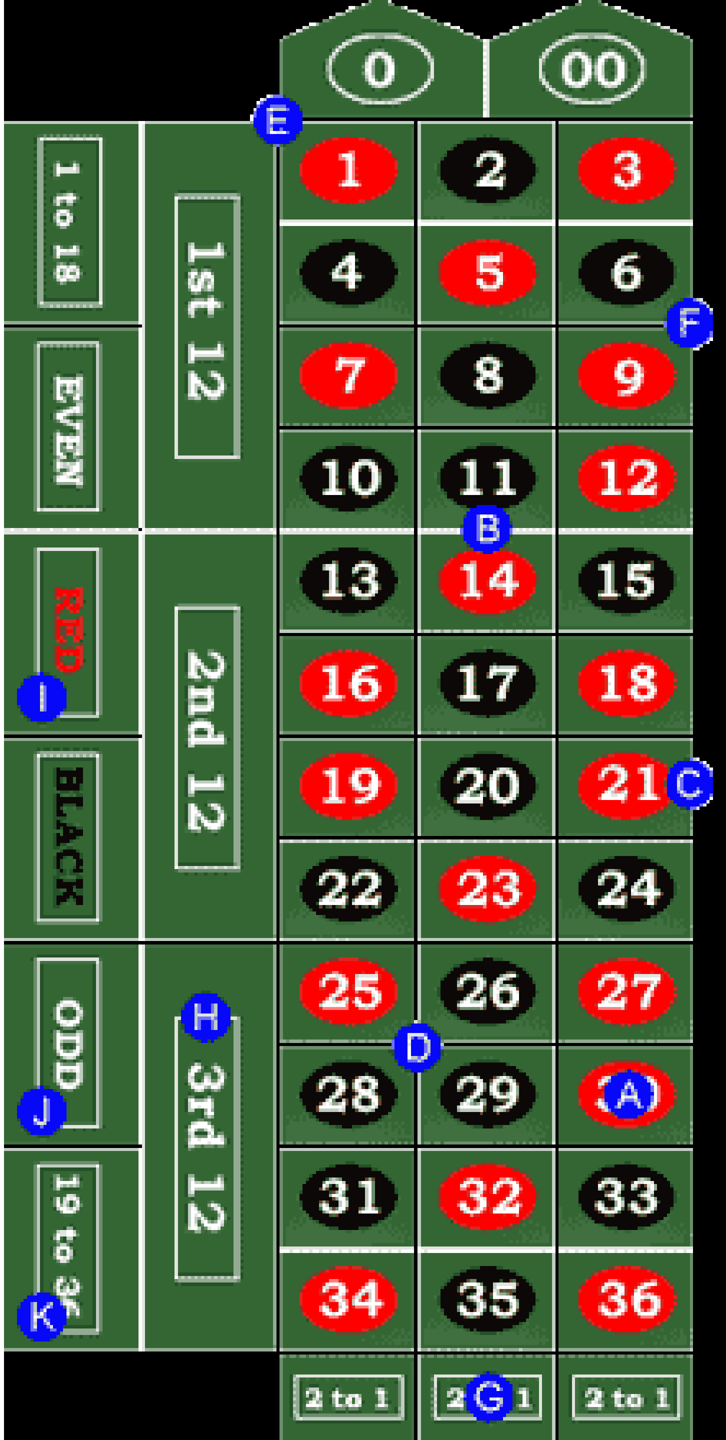
- If you make two or three mistakes an hour, you lose any advantage
  - And, in fact, cause a disadvantage!
- You lose lots of money learning to count cards
- Then, once you can do so, you are banned from the casinos

Roulette

Roulette

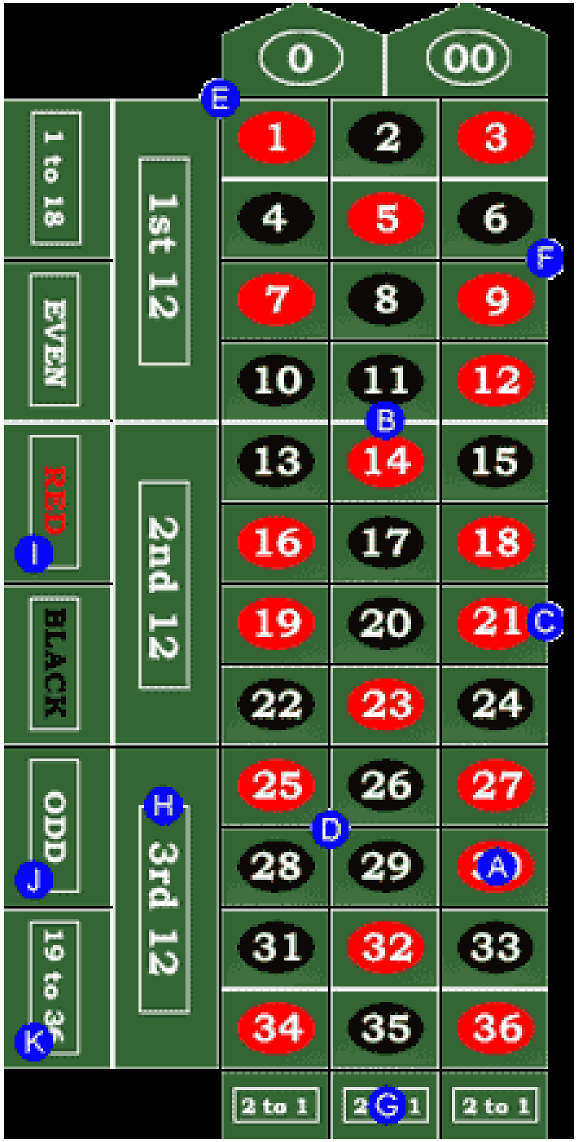
- A wheel with 38 spots is spun
  - Spots are numbered 1-36, 0, and 00
  - European casinos don't have the 00
- A ball drops into one of the 38 spots
- A bet is placed as to which spot or spots the ball will fall into
  - Money is then paid out if the ball lands in the spot(s) you bet upon

The Roulette table



The Roulette table

- Bets can be placed on: Probability:
  - A single number 1/38
  - Two numbers 2/38
  - Four numbers 4/38
  - All even numbers 18/38
  - All odd numbers 18/38
  - The first 18 nums 18/38
  - Red numbers 18/38



The Roulette table

Bets can be placed on:	Probability:	Payout:
A single number	1/38	36x
Two numbers	2/38	18x
Four numbers	4/38	9x
All even numbers	18/38	2x
All odd numbers	18/38	2x
The first 18 nums	18/38	2x
Red numbers	18/38	2x

Roulette

- It has been proven that no advantageous strategies exist
- Including:
  - Learning the wheel's biases
    - Casino's regularly balance their Roulette wheels
  - Martingale betting strategy
    - Where you double your bet each time (thus making up for all previous losses)
    - It still won't work!
    - You can't double your money forever
      - It could easily take 50 times to achieve finally win
      - If you start with $1, then you must put in $\$1*2^{50}$ = $1,125,899,906,842,624 to win this way!
      - That's 1 quadrillion
    - See http://en.wikipedia.org/wiki/Martingale_(roulette_system) for more info

Quick survey

- I felt I understood Roulette probability…

a) Very well

b) With some review, I'll be good

c) Not really

d) Not at all

Quick survey

- If I was going to spend money gambling, would I choose Roulette?

a) Definitely – a way to make money

b) Perhaps

c) Probably not

d) Definitely not – it's a way to lose money

MTH 2215
Applied Discrete Mathematics

Chapter 6.2
Probability Theory

These class notes are based on material from our textbook, **Discrete Mathematics and Its Applications**, 6th ed., by Kenneth H. Rosen, published by McGraw Hill, Boston, MA, 2006. They are intended for classroom use only and are **not** a substitute for reading the textbook.

Terminology

- A (stochastic) *experiment* is a procedure that yields one of a given set of possible outcomes
- The *sample space S* of the experiment is the set of possible outcomes.
- An event is a subset of sample space.
- A random variable is a function that assigns a real value to each outcome of an experiment

Normally, a probability is related to an experiment or a trial.

Let's take flipping a coin for example, what are the possible outcomes?

Heads or tails (front or back side) of the coin will be shown upwards.

After a sufficient number of tossing, we can "statistically" conclude that the probability of head is 0.5.

In rolling a dice, there are 6 outcomes. Suppose we want to calculate the prob. of the event of odd numbers of a dice. What is that probability?

Random Variables

- A *"random variable"* V is any variable whose value is unknown, or whose value depends on the precise situation.
  - *E.g.*, the number of students in class today
  - Whether it will rain tonight (Boolean variable)
- The proposition $V=v_i$ may have an uncertain truth value, and may be assigned a *probability*.

Example 10

- A fair coin is flipped 3 times. Let S be the sample space of 8 possible outcomes, and let X be a random variable that assignees to an outcome the number of heads in this outcome.

- **Random variable *X* is a function** $X{:}S \rightarrow X(S)$,
where $X(S)=\{0, 1, 2, 3\}$ is the range of X, which is the number of heads, and
$S=\{$ (TTT), (TTH), (THH), (HTT), (HHT), (HHH), (THT), (HTH) $\}$

- X(TTT) = 0
X(TTH) = X(HTT) = X(THT) = 1
X(HHT) = X(THH) = X(HTH) = 2
X(HHH) = 3

- The **probability distribution (pdf) of random variable** X is given by
P(X=3) = 1/8, P(X=2) = 3/8, P(X=1) = 3/8, P(X=0) = 1/8.

Experiments & Sample Spaces

- A (stochastic) *experiment* is any process by which a given random variable V gets assigned some *particular* value, and where this value is not necessarily known in advance.
  - We call it the "actual" value of the variable, as determined by that particular experiment.
- The *sample space* S of the experiment is just the domain of the random variable, $S = \mathrm{dom}[V]$.
- The *outcome* of the experiment is the specific value v_i of the random variable that is selected.

Events

- An *event* E is any set of possible outcomes in S…
  - That is, $E \subseteq S = \mathbf{dom}[V]$.
    - E.g., the event that "less than 50 people show up for our next class" is represented as the set $\{1, 2, \ldots, 49\}$ of values of the variable V = (# of people here next class).
- We say that event E *occurs* when the actual value of V is in E, which may be written $V \in E$.
  - Note that $V \in E$ denotes the proposition (of uncertain truth) asserting that the actual outcome (value of V) will be one of the outcomes in the set E.

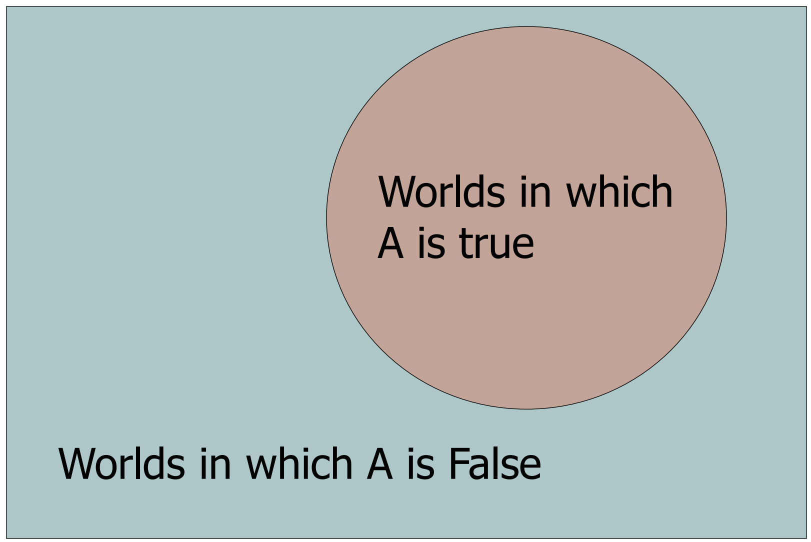
Probabilities

- We write P(A) as "the fraction of possible worlds in which A is true"
- We could at this point spend 2 hours on the philosophy of this.
- But we won't.

Visualizing A

Event space of all possible worlds

Its area is 1



P(A) = Area of reddish oval

Probability

- The *probability* $p = \Pr[E] \in [0,1]$ of an event E is a real number representing our degree of certainty that E will occur.
  - If $\Pr[E] = 1$, then E is absolutely certain to occur,
    - thus $V \in E$ has the truth value **True**.
  - If $\Pr[E] = 0$, then E is absolutely certain *not* to occur,
    - thus $V \in E$ has the truth value **False**.
  - If $\Pr[E] = ½$, then we are *maximally uncertain* about whether E will occur; that is,
    - $V \in E$ and $V \notin E$ are considered *equally likely*.
  - How do we interpret other values of p?

Note: We could also define probabilities for more general propositions, as well as events.

Four Definitions of Probability

- Several alternative definitions of probability are commonly encountered:
  - Frequentist, Bayesian, Laplacian, Axiomatic
- They have different strengths & weaknesses, philosophically speaking.
  - But fortunately, they coincide with each other and work well together, in the majority of cases that are typically encountered.

Probability: Frequentist Definition

- The probability of an event E is the limit, as $n \to \infty$, of the fraction of times that we find $V \in E$ over the course of n independent repetitions of (different instances of) the same experiment.

$$\Pr[E] :\equiv \lim_{n \to \infty} \frac{n_{V \in E}}{n}$$

- Some problems with this definition:
  - It is only well-defined for experiments that can be independently repeated, infinitely many times!
    - or at least, if the experiment can be repeated in principle, *e.g.*, over some hypothetical ensemble of (say) alternate universes.
  - It can never be measured exactly in finite time!
- **Advantage:** It's an objective, mathematical definition.

Probability: Bayesian Definition

- Suppose a rational, profit-maximizing entity R is offered a choice between two rewards:
  - Winning \$1 if and only if the event E actually occurs.
  - Receiving p dollars (where $p \in [0,1]$) unconditionally.
- If R can honestly state that he is completely indifferent between these two rewards, then we say that R's probability for E is p, that is, $\Pr_R[E] :\equiv p$.
- **Problem:** It's a subjective definition; depends on the reasoner R, and his knowledge, beliefs, & rationality.
  - The version above additionally assumes that the utility of money is linear.
    - This assumption can be avoided by using "utils" (utility units) instead of dollars.

Probability: Laplacian Definition

- First, assume that all individual outcomes in the sample space are *equally likely* to each other…
  - Note that this term still needs an operational definition!
- Then, the probability of any event E is given by,

  $\Pr[E] = |E|/|S|$. Very simple!
- **Problems:** Still needs a definition for *equally likely*, and depends on the existence of *some* finite sample space S in which all outcomes in S are, in fact, equally likely.

Probability: Axiomatic Definition

- Let p be any total function p:$S \rightarrow [0,1]$ such that

$$\sum_s p(s) = 1.$$

- Such a p is called a *probability distribution*.
- Then, the *probability under p* of any event $E \subseteq S$ is just:

$$p[E] :\equiv \sum_{s \in E} p(s)$$

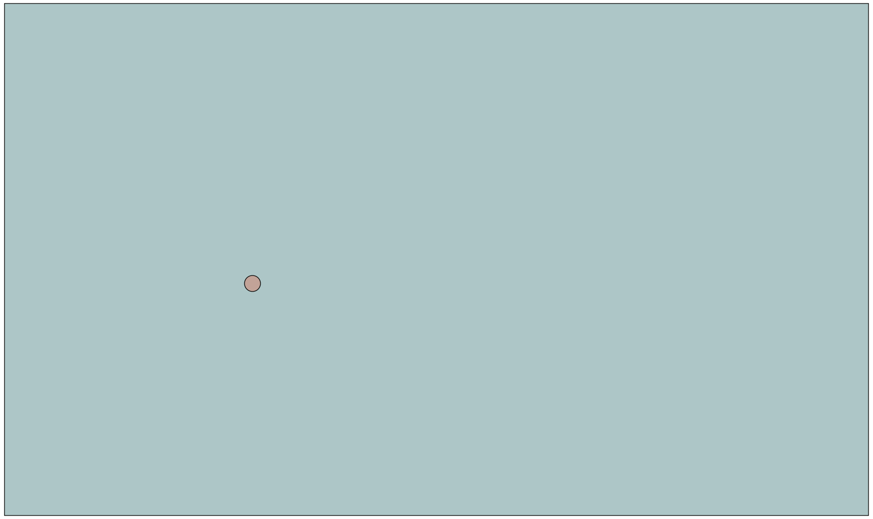
- **Advantage:** Totally mathematically well-defined!
  - This definition can even be extended to apply to infinite sample spaces, by changing $\sum \rightarrow \int$, and calling p a *probability density function* or a probability *measure*.
- **Problem:** Leaves operational meaning unspecified.

The Axioms of Probability

- $0 <= P(A) <= 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Interpreting the axioms

- $0 <= P(A) <= 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

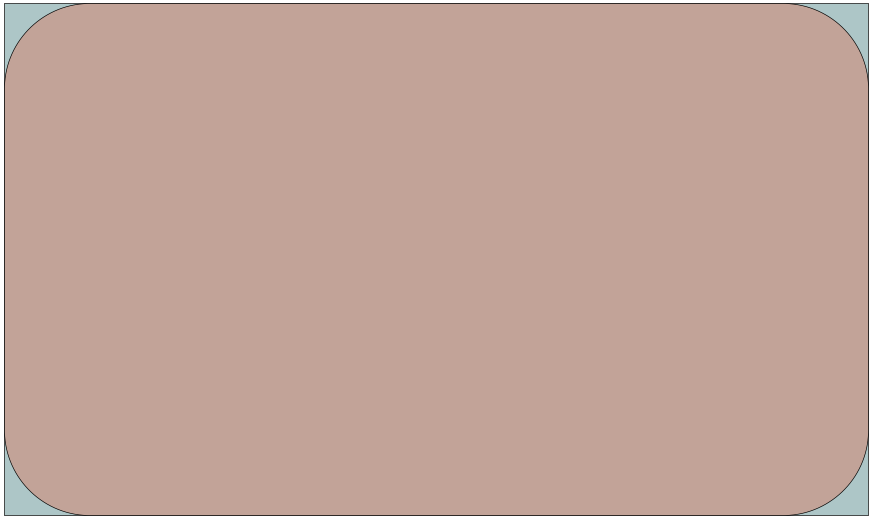


The area of A can't get any smaller than 0

And a zero area would mean no world could ever have A true

Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)

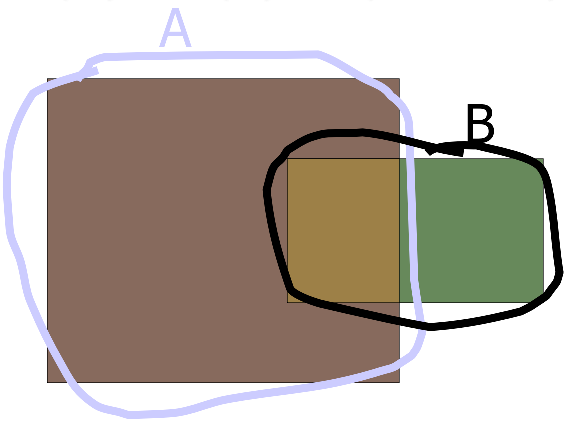


The area of A can't get any bigger than 1

And an area of 1 would mean all worlds will have A true

Interpreting the axioms

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) - P(A and B)



These Axioms are Not to be Trifled With

- There have been attempts to do different methodologies for uncertainty
  - Fuzzy Logic
  - Three-valued logic
  - Dempster-Shafer
  - Non-monotonic reasoning
- But the axioms of probability are the only system with this property:

If you gamble using them you can't be unfairly exploited by an opponent using some other system [di Finetti 1931]

Theorems from the Axioms

- $0 <= P(A) <= 1$, $P(True) = 1$, $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

From these we can prove:

P(not A) = P(~A) = 1-P(A)

- How?

Another important theorem

- $0 <= P(A) <= 1$, $P(True) = 1$, $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

From these we can prove:

$$P(A) = P(A \wedge B) + P(A \wedge \sim B)$$

- How?

Probability of an event E

The probability of an event E is the sum of the probabilities of the outcomes in E. That is

$$p(E) = \sum_{s \in E} p(s)$$

Note that, if there are n outcomes in the event E, that is, if $E = \{a_1, a_2, \ldots, a_n\}$ then

$$p(E) = \sum_{i=1}^{n} p(a_i)$$

Example

- What is the probability that, if we flip a coin three times, that we get an odd number of tails?

(TTT), (TTH), (THH), (HTT), (HHT), (HHH), (THT), (HTH)

Each outcome has probability 1/8,

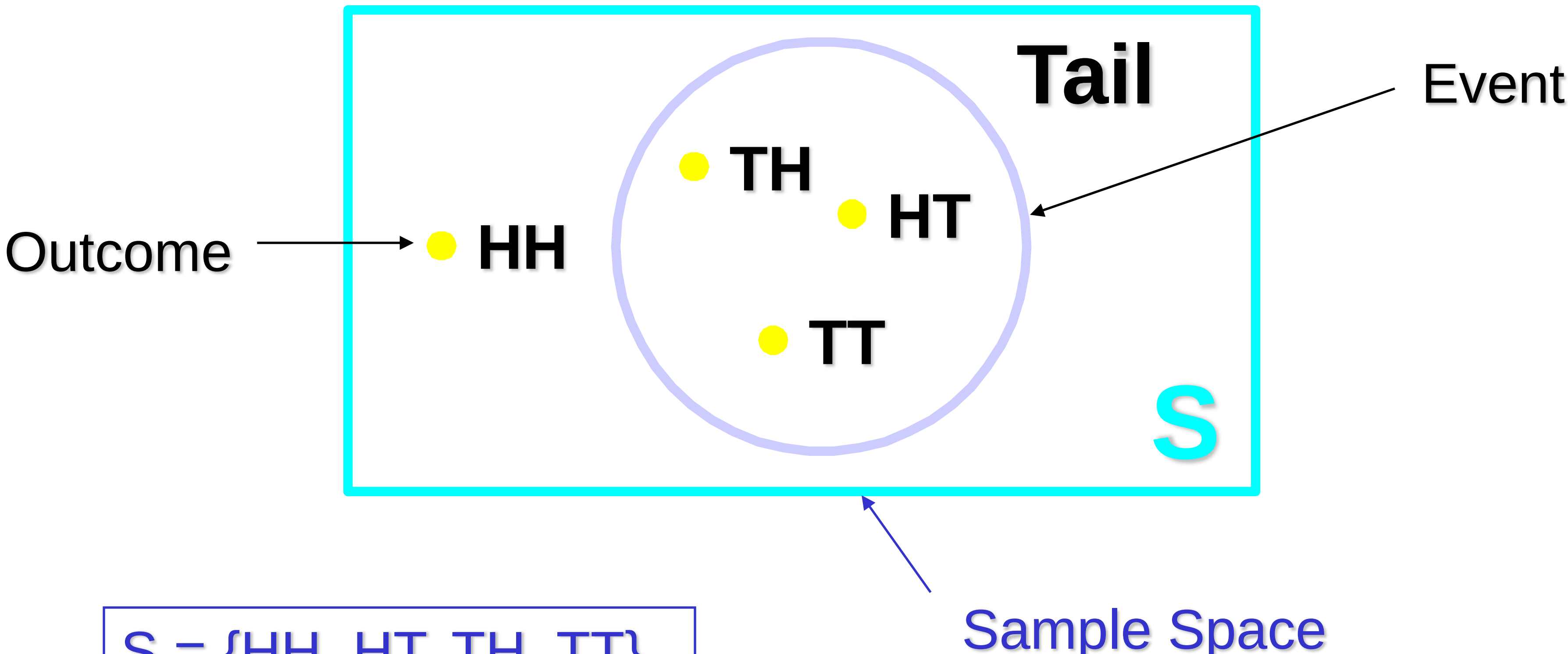
p(odd number of tails) = 1/8+1/8+1/8+1/8 = ½

Visualizing Sample Space

- 1. Listing
  - S = {Head, Tail}
- 2. Venn Diagram
- 3. Contingency Table
- 4. Decision Tree Diagram

Venn Diagram

Experiment: Toss 2 Coins. Note Faces.



S = {HH, HT, TH, TT}

Contingency Table

Experiment: Toss 2 Coins. Note Faces.

1^{st} Coin	2^{nd} Coin Head	2^{nd} Coin Tail	Total
Head	HH	HT	HH, HT
Tail	TH	TT	TH, TT
Total	HH, TH	HT, TT	S

Simple Event (Head on 1st Coin)

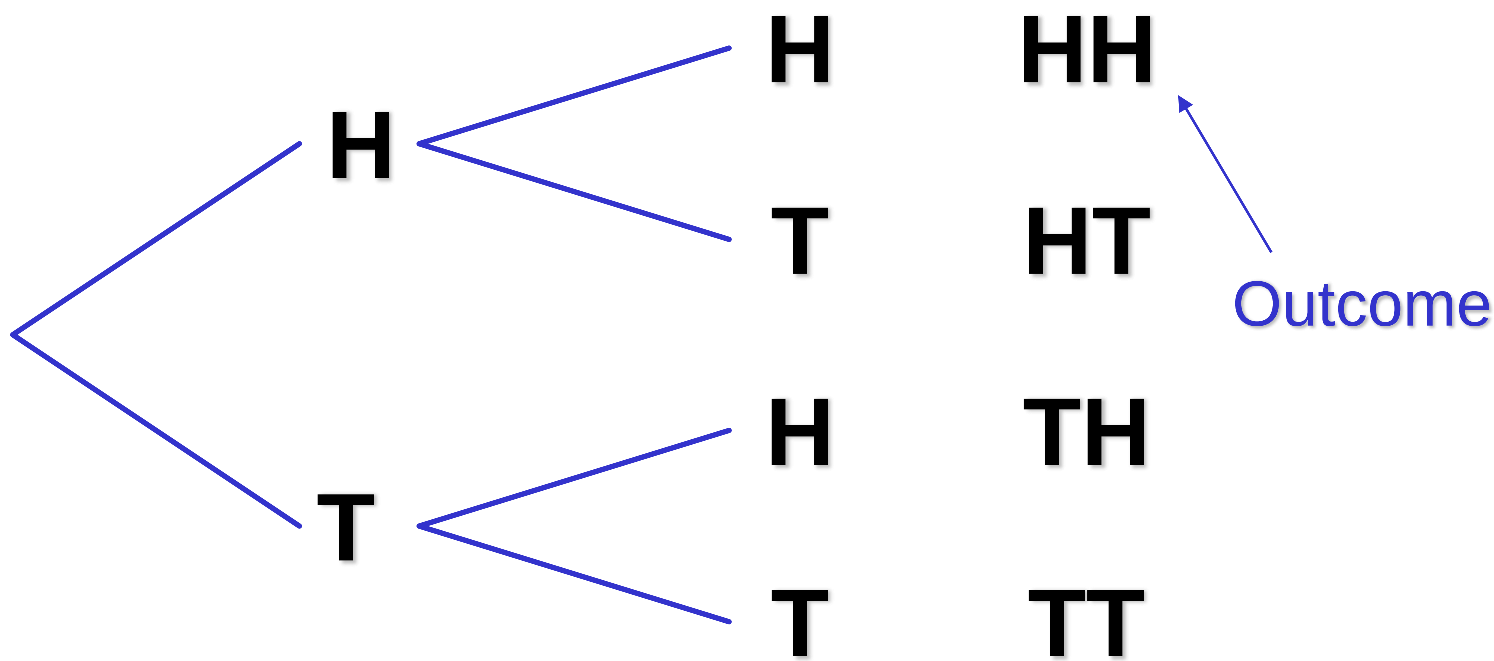
Outcome

Sample Space

S = {HH, HT, TH, TT}

Tree Diagram

Experiment: Toss 2 Coins. Note Faces.



S = {HH, HT, TH, TT}

Sample Space

Discrete Random Variable

- Possible values (outcomes) are discrete
  - E.g., natural number (0, 1, 2, 3 etc.)
- Obtained by Counting
- Usually Finite Number of Values
  - But could be infinite (must be “countable”)

Discrete Probability Distribution
(also called probability mass function (pmf))

1. List of All possible [x, $p(x)$] pairs
   - x = Value of Random Variable (Outcome)
   - $p(x)$ = Probability Associated with Value
2. Mutually Exclusive (No Overlap)
3. Collectively Exhaustive (Nothing Left Out)
4. $0 \leq p(x) \leq 1$
5. $\sum p(x) = 1$

Visualizing Discrete Probability Distributions

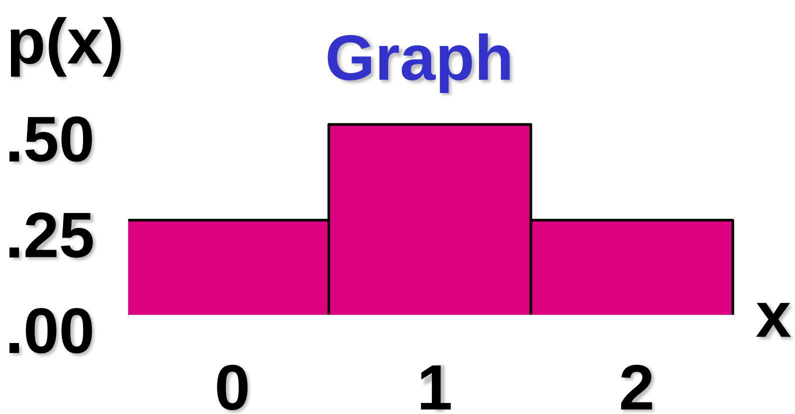
Listing

{ (0, .25), (1, .50), (2, .25) }

Table

# Tails	f(x) Count	p(x)
0	1	.25
1	2	.50
2	1	.25

Graph



Equation

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Arity of Random Variables

- Suppose A can take on more than 2 values
- A is a *random variable with arity k* if it can take on exactly one value out of $\{v_1, v_2, .. v_k\}$
- Thus…

$$P(A = v_i \wedge A = v_j) = 0 \text{ if } i \neq j$$

$$P(A = v_1 \vee A = v_2 \vee A = v_k) = 1$$

Mutually Exclusive Events

- Two events E_1, E_2 are called *mutually exclusive* if they are disjoint: $E_1 \cap E_2 = \varnothing$
  - Note that two mutually exclusive events *cannot both occur* in the same instance of a given experiment.
- For mutually exclusive events,

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2].$$

Exhaustive Sets of Events

- A set $\boldsymbol{E} = \{E_1, E_2, \ldots\}$ of events in the sample space S is called *exhaustive* iff $\bigcup E_i = S$.
- An exhaustive set $\boldsymbol{E}$ of events that are all mutually exclusive with each other has the property that

$$\sum \Pr[E_i] = 1.$$

An easy fact about Multivalued Random Variables:

- Using the axioms of probability…

  0 <= P(A) <= 1, P(True) = 1, P(False) = 0

  P(A or B) = P(A) + P(B) - P(A and B)

- And assuming that A obeys…

$$P(A=v_i \wedge A=v_j)=0 \text{ if } i \neq j$$

- It's easy to prove that $P(A=v_1 \vee A=v_2 \vee A=v_k)=1$

$$P(A=v_1 \vee A=v_2 \vee A=v_i)=\sum_{j=1}^{i} P(A=v_j)$$

- And thus we can prove

$$\sum_{j=1}^{k} P(A=v_j)=1$$

Another fact about Multivalued Random Variables:

- Using the axioms of probability…

  0 <= P(A) <= 1, P(True) = 1, P(False) = 0

  P(A or B) = P(A) + P(B) - P(A and B)

- And assuming that A obeys…

$$P(A = v_i \wedge A = v_j) = 0 \text{ if } i \neq j$$

- It's easy to prove that

$$P(A = v_1 \vee A = v_2 \vee A = v_k) = 1$$

$$P(B \wedge [A = v_1 \vee A = v_2 \vee A = v_i]) = \sum_{j=1}^{i} P(B \wedge A = v_j)$$

Elementary Probability Rules

- P(~A) + P(A) = 1
- P(B) = P(B ^ A) + P(B ^ ~A)

$$\sum_{j=1}^{k} P(A = v_j) = 1$$

$$P(B) = \sum_{j=1}^{k} P(B \wedge A = v_j)$$

Bernoulli Trials

- Each performance of an experiment with only two possible outcomes is called a **Bernoulli trial**.
- In general, a possible outcome of a Bernoulli trial is called a **success** or a **failure**.
- If p is the probability of a success and q is the probability of a failure, then p+q=1.

Example

A coin is biased so that the probability of heads is 2/3. What is the probability that exactly four heads come up when the coin is flipped seven times, assuming that the flips are independent?

The number of ways that we can get four heads is: $C(7,4) = 7!/4!3! = 7*5 = 35$

The probability of getting four heads and three tails is $(2/3)^4(1/3)^{3=}$ $16/3^7$

p(4 heads and 3 tails) is $C(7,4)$ $(2/3)^4(1/3)^3$ = $35*16/3^7 = 560/2187$

Probability of k successes in n independent Bernoulli trials.

The probability of k successes in n independent Bernoulli trials, with probability of success p and probability of failure $q = 1-p$ is $C(n,k)p^k q^{n-k}$

Find each of the following probabilities when n independent Bernoulli trials are carried out with probability of success, p.

- Probability of no successes.

$C(n,0)p^0q^{n-k} = 1(p^0)(1-p)^n = (1-p)^n$

- Probability of at least one success.

$1 - (1-p)^n$ (why?)

Find each of the following probabilities when n independent Bernoulli trials are carried out with probability of success, p.

- Probability of at most one success.

Means there can be no successes or one success

$C(n,0)p^0q^{n-0} + C(n,1)p^1q^{n-1}$

$(1-p)^n + np(1-p)^{n-1}$

- Probability of at least two successes.

$1 - (1-p)^n - np(1-p)^{n-1}$

A coin is flipped until it comes ups tails. The probability the coin comes up tails is p.

- What is the probability that the experiment ends after n flips, that is, the outcome consists of n-1 heads and a tail?

$(1-p)^{n-1}p$

Probability vs. Odds

Exercise: Express the probability p as a function of the odds in favor O.

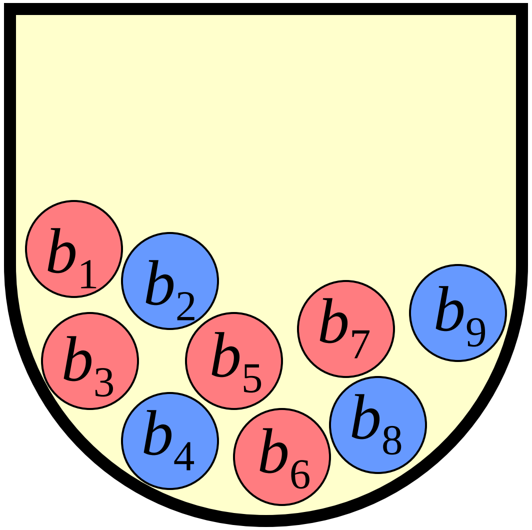
- You may have heard the term "odds."
  - It is widely used in the gambling community.
- This is not the same thing as probability!
  - But, it is very closely related.
- The *odds in favor* of an event E means the *relative* probability of E compared with its complement $\overline{E}$.

$$O(E) :\equiv \Pr(E)/\Pr(\overline{E}).$$

  - *E.g.*, if $p(E) = 0.6$ then $p(\overline{E}) = 0.4$ and $O(E) = 0.6/0.4 = 1.5$.
- Odds are conventionally written as a ratio of integers.
  - *E.g.*, 3/2 or 3:2 in above example. "Three to two in favor."
- The *odds against* E just means $1/O(E)$. "2 to 3 against"

Example 1: Balls-and-Urn

- Suppose an urn contains 4 blue balls and 5 red balls.
- An example **experiment:** Shake up the urn, reach in (without looking) and pull out a ball.
- A **random variable** V: Identity of the chosen ball.
- The **sample space** S: The set of all possible values of V:
  - In this case, $S = \{b_1,\ldots,b_9\}$
- An **event** E: "The ball chosen is blue": $E = \{$ ________________ $\}$
- What are the odds in favor of E?
- What is the probability of E?



Independent Events

- Two events E,F are called *independent* if

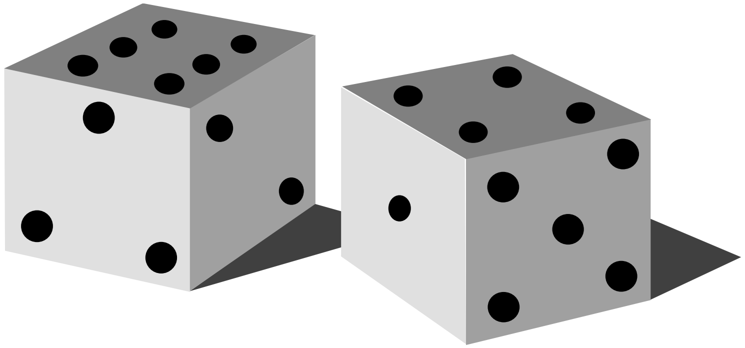
  $\Pr[E \cap F] = \Pr[E] \cdot \Pr[F]$.
- Relates to the product rule for the number of ways of doing two independent tasks.
- **Example:** Flip a coin, and roll a die.

  Pr[(coin shows heads) $\cap$ (die shows 1)] =

  Pr[coin is heads] × Pr[die is 1] = ½×1/6 =1/12.

Example

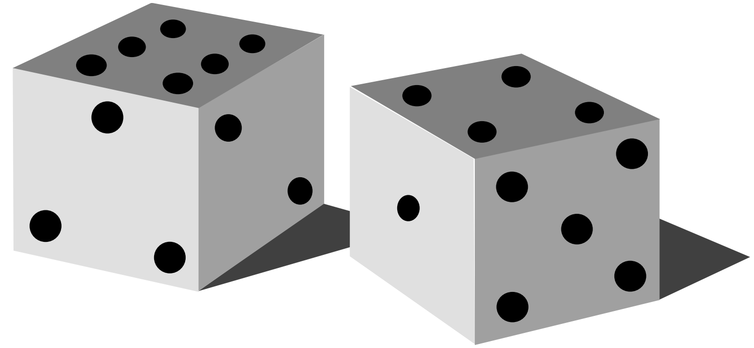
Suppose a red die and a blue die are rolled.
The sample space:



	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

Are the events sum is 7 and the blue die is 3 independent?

The events sum is 7 and
the blue die is 3 are independent:



$|S| = 36$

$|\text{sum is } 7| = 6$

$|\text{blue die is } 3| = 6$

$|\text{ in intersection }| = 1$

	1	2	3	4	5	6
1	x	x	x	x	x	x
2	x	x	x	x	x	x
3	x	x	x	x	x	x
4	x	x	x	x	x	x
5	x	x	x	x	x	x
6	x	x	x	x	x	x

p(sum is 7 and blue die is 3) =1/36

p(sum is 7) p(blue die is 3) =6/36*6/36=1/36

Thus, p((sum is 7) and (blue die is 3)) = p(sum is 7) p(blue die is 3)

Conditional Probability

- Let E,F be any events such that $\Pr[F]>0$.
- Then, the *conditional probability of E given F*, written $\Pr[E|F]$, is defined as

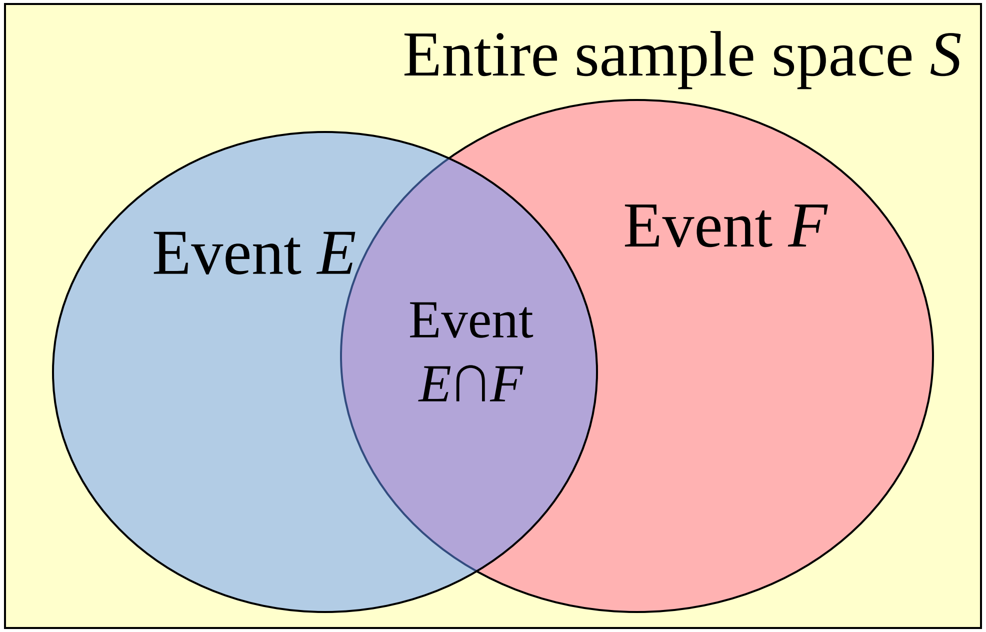
$$\Pr[E|F] :\equiv \Pr[E\cap F]/\Pr[F].$$

- This is what our probability that E would turn out to occur should be, if we are given *only* the information that F occurs.
- If E and F are independent then $\Pr[E|F] = \Pr[E]$.

$$\because\ \Pr[E|F] = \Pr[E\cap F]/\Pr[F] = \Pr[E]\times\Pr[F]/\Pr[F] = \Pr[E]$$

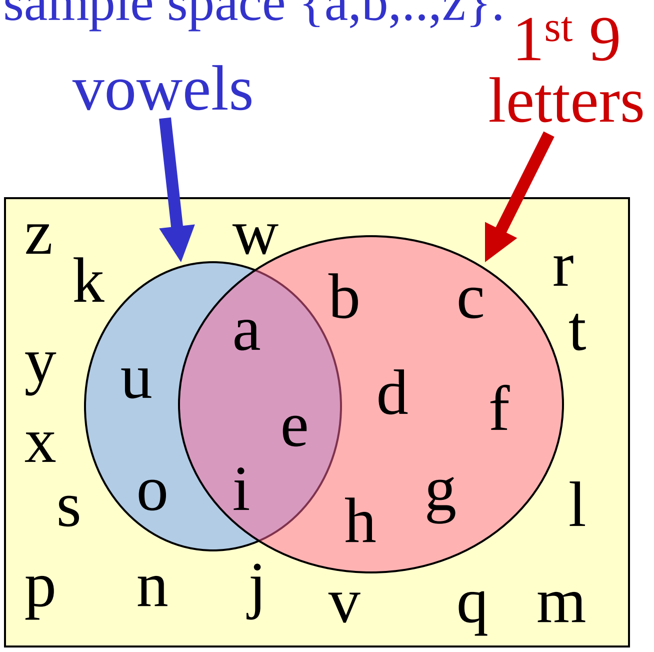
Visualizing Conditional Probability

- If we are given that event *F* occurs, then
  - Our attention gets restricted to the subspace *F*.
- Our *posterior* probability for *E* (after seeing *F*) corresponds to the *fraction* of *F* where *E* occurs also.
- Thus, $p'(E) = p(E \cap F)/p(F)$.



Conditional Probability Example

- Suppose I choose a single letter out of the 26-letter English alphabet, totally at random.
  - Use the Laplacian assumption on the sample space {a,b,..,z}.
  - What is the (prior) probability that the letter is a vowel?
    - Pr[Vowel] = __ / __ .
- Now, suppose I tell you that the letter chosen happened to be in the first 9 letters of the alphabet.
  - Now, what is the conditional (or posterior) probability that the letter is a vowel, given this information?
    - Pr[Vowel | First9] = ___ / ___ .



Sample Space *S*

Example

- What is the probability that, if we flip a coin three times, that we get an odd number of tails (=event *E*), if we know that the event *F*, the first flip comes up tails occurs?

(TTT), (TTH), (THH), (HTT),
(HHT), (HHH), (THT), (HTH)

Each outcome has probability 1/4,

p(*E* |*F*) = 1/4+1/4 = ½, where E=odd number of tails
or p(*E*|*F*) = p(*E*∩*F*)/p(*F*) = 2/4 = ½
For comparison p(*E*) = 4/8 = ½

E and *F* are independent, since p(*E* |*F*) = Pr(*E*).

Prior and Posterior Probability

- Suppose that, before you are given any information about the outcome of an experiment, your personal probability for an event E to occur is $p(E) = \Pr[E]$.
  - The probability of E in your original probability distribution p is called the *prior* probability of E.
    - This is its probability *prior* to obtaining any information about the outcome.
- Now, suppose someone tells you that some event F (which may overlap with E) actually occurred in the experiment.
  - Then, you should *update* your personal probability for event E to occur, to become $p'(E) = \Pr[E|F] = p(E \cap F)/p(F)$.
    - The conditional probability of E, *given* F.
  - The probability of E in your *new* probability distribution p' is called the *posterior* probability of E.
    - This is its probability *after* learning that event F occurred.
- After seeing F, the posterior distribution p' is defined by letting $p'(v) = p(\{v\} \cap F)/p(F)$ for each individual outcome $v \in S$.

MTH 2215
Applied Discrete Mathematics

Chapter 6.3
Bayes' Theorem

These class notes are based on material from our textbook, **Discrete Mathematics and Its Applications**, 6th ed., by Kenneth H. Rosen, published by McGraw Hill, Boston, MA, 2006. They are intended for classroom use only and are **not** a substitute for reading the textbook.

Bayes' Rule

Rev. Thomas Bayes
1702-1761

- One way to compute the probability that a hypothesis *H* is correct, given some data *D*:

$$\Pr[H \mid D] = \frac{\Pr[D \mid H] \cdot \Pr[H]}{\Pr[D]}$$

- This follows directly from the definition of conditional probability! (**Exercise:** Prove it.)
- This rule is the foundation of *Bayesian methods* for probabilistic reasoning, which are very powerful, and widely used in artificial intelligence applications:
  - For data mining, automated diagnosis, pattern recognition, statistical modeling, even evaluating scientific hypotheses!

Bayes' Theorem

- Allows one to compute the probability that a hypothesis *H* is correct, given data *D*:

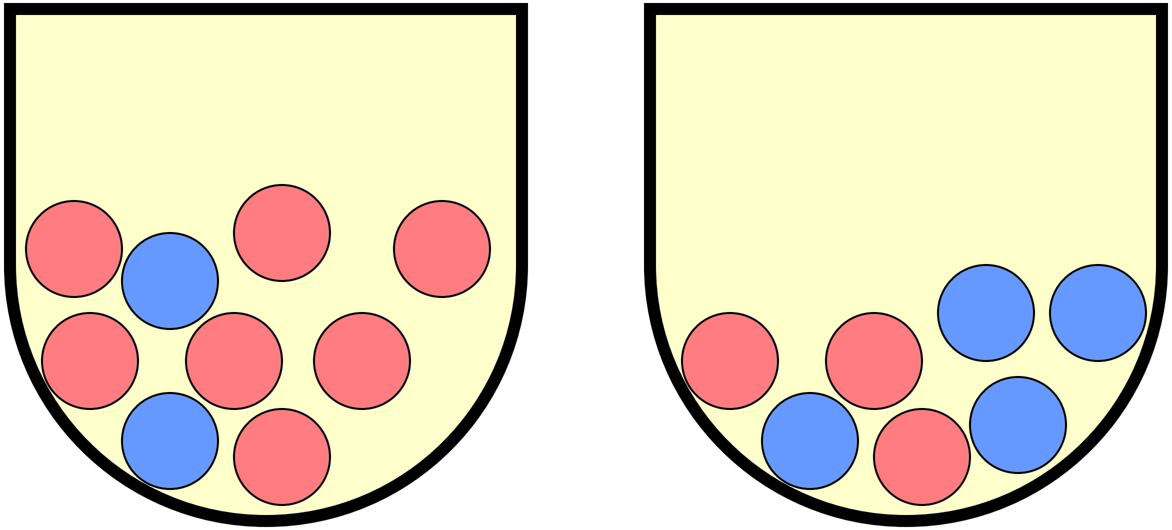
$$\Pr[H \mid D] = \frac{\Pr[D \mid H] \cdot \Pr[H]}{\Pr[D]}$$

$$\Pr[H_i \mid D] = \frac{\Pr[D \mid H_i] \cdot \Pr[H_i]}{\sum_j \Pr[D \mid H_j] \cdot \Pr[H_j]}$$

Set of H_j is exhaustive

Example 1: Two boxes with balls

- Two boxes: first: 2 blue and 7 red balls; second: 4 blue and 3 red balls
- Bob selects a ball by first choosing one of the two boxes, and then one ball from this box.
- If Bob has selected a red ball, what is the probability that he selected a ball from the first box.
- An **event** E: Bob has chosen a red ball.
- An **event** F: Bob has chosen a ball from the first box.
- We want to find $p(F \mid E)$



Example 2:

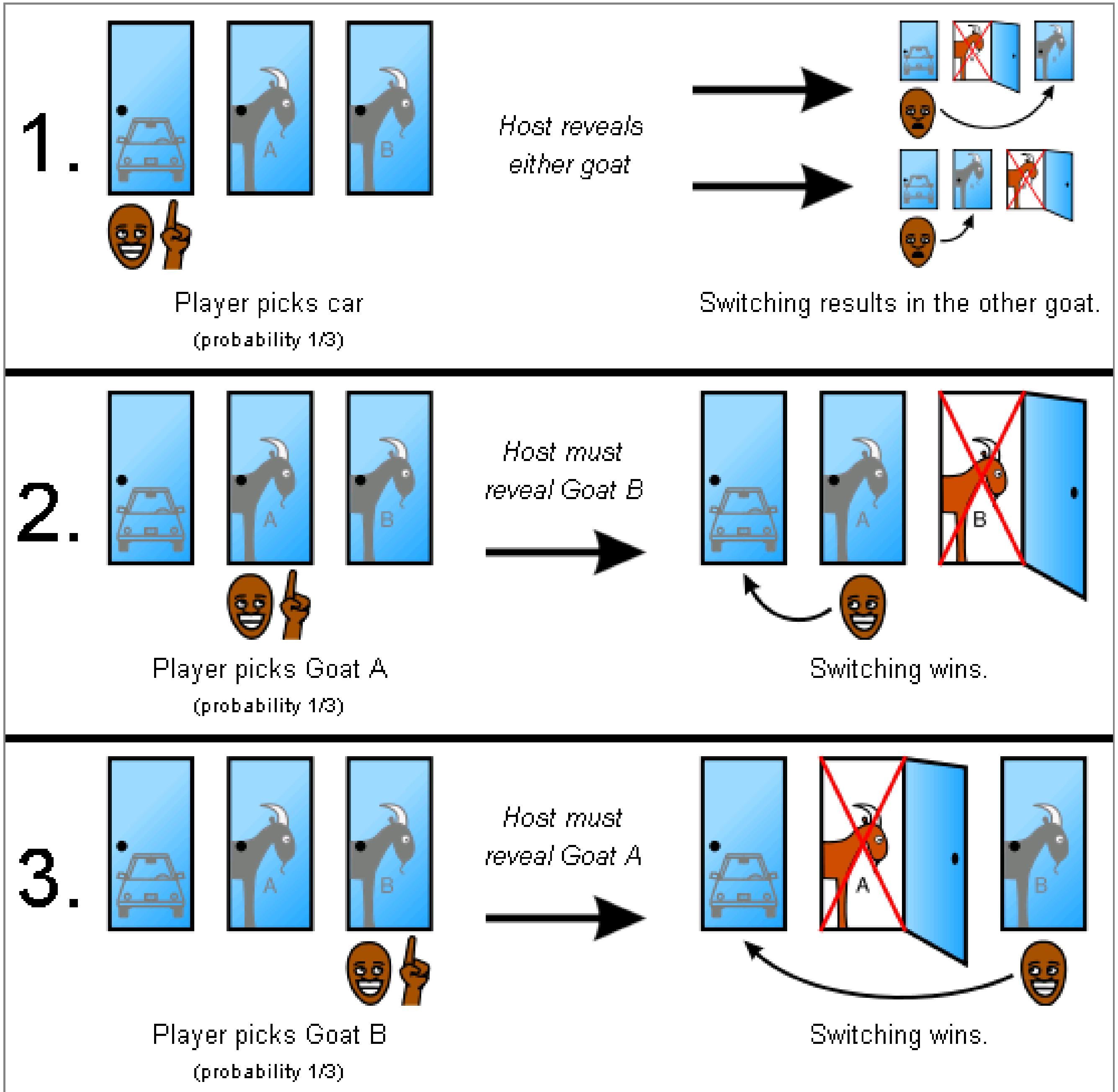
- Suppose 1% of population has AIDS
- Prob. that the positive result is right: 95%
- Prob. that the negative result is right: 90%
- What is the probability that someone who has the positive result is actually an AIDS patient?

- H: event that a person has AIDS
- D: event of positive result
- $P[D|H] = 0.95$ $P[D|\neg H] = 1 - 0.9$

- $P[D] = P[D|H]P[H]+P[D|\neg H]P[\neg H]$
  $= 0.95*0.01+0.1*0.99=0.1085$
- $P[H|D] = 0.95*0.01/0.1085=0.0876$

What's behind door number three?

- The Monty Hall problem paradox
  - Consider a game show where a prize (a car) is behind one of three doors
  - The other two doors do not have prizes (goats instead)
  - After picking one of the doors, the host (Monty Hall) opens a different door to show you that the door he opened is not the prize
  - Do you change your decision?
- Your initial probability to win (i.e. pick the right door) is 1/3
- What is your chance of winning if you change your choice after Monty opens a wrong door?
- After Monty opens a wrong door, if you change your choice, your chance of ***winning*** is 2/3
  - Thus, your chance of winning ***doubles*** if you change
  - Huh?



The player has an equal chance of initially selecting the car, Goat A, or Goat B. Switching results in a win 2/3 of the time.

Monty Hall Problem

C_i - The car is behind Door *i*, for *i* equal to 1, 2 or 3.

$$P(C_i) = \frac{1}{3}$$

H_{ij} - The host opens Door *j* after the player has picked Door *i*, for *i* and *j* equal to 1, 2 or 3.

Without loss of generality, assume, by re-numbering the doors if necessary, that the player picks Door 1,
and that the host then opens Door 3, revealing a goat.
In other words, the host *makes* proposition H_{13} true.

Then the posterior probability of winning by *not* switching doors is $P(C_1|H_{13})$.

$P(H_{13} \mid C_1) = 0.5$, since the host will always open a door that has no car behind it, chosen from among the two not picked by the player (which are 2 and 3 here)

$$P(C_1 \mid H_{13}) = \frac{P(H_{13} \mid C_1)P(C_1)}{P(H_{13})} = \frac{P(H_{13} \mid C_1)P(C_1)}{\sum_{i=1}^{3} P(H_{13} \mid C_i)P(C_i)}$$

$$= \frac{P(H_{13} \mid C_1)P(C_1)}{P(H_{13} \mid C_1)P(C_1) + P(H_{13} \mid C_2)P(C_2) + P(H_{13} \mid C_3)P(C_3)}$$

$$= \frac{\frac{1}{2}\cdot\frac{1}{3}}{\frac{1}{2}\cdot\frac{1}{3} + 1\cdot\frac{1}{3} + 0\cdot\frac{1}{3}} = \frac{\frac{1}{6}}{\frac{1}{2}} = \frac{1}{3}$$

The probability of winning by switching is P($C_2|H_{13}$), since under our assumption switching means switching the selection to Door 2, since P($C_3|H_{13}$) = 0 (the host will never open the door with the car)

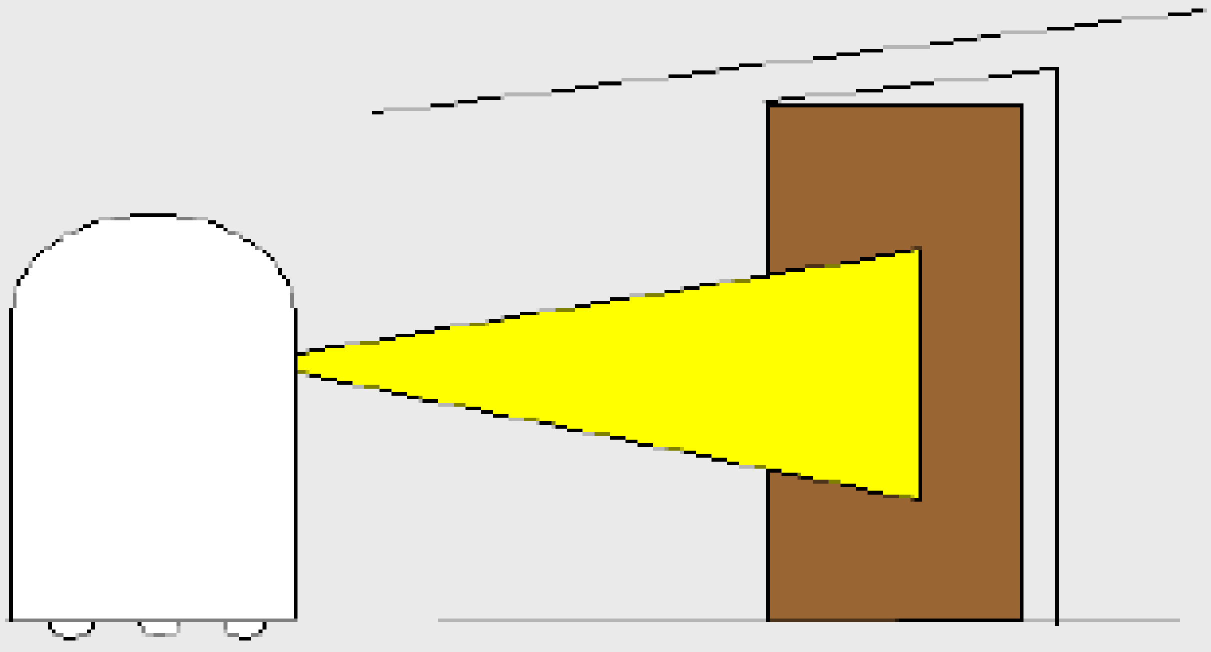
$$P(C_2 \mid H_{13}) = \frac{P(H_{13} \mid C_2)P(C_2)}{P(H_{13})} = \frac{P(H_{13} \mid C_2)P(C_2)}{\sum_{i=1}^{3} P(H_{13} \mid C_i)P(C_i)}$$

$$= \frac{1 \cdot \frac{1}{3}}{\frac{1}{2} \cdot \frac{1}{3} + 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3}} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

The posterior **probability of winning by *not* switching doors** is P($C_1|H_{13}$) = 1/3.

Simple Example of State Estimation

- Suppose a robot obtains measurement z
- What is *P(doorOpen|z)*?



Causal vs. Diagnostic Reasoning

- $P(open|z)$ is diagnostic.
- $P(z|open)$ is causal.
- Often causal knowledge is easier to obtain.
- Bayes rule allows us to use causal knowledge:

count frequencies!

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z)}$$

Example

- $P(z|open) = 0.6 \quad P(z|\neg open) = 0.3$
- $P(open) = P(\neg open) = 0.5$

$$P(open \mid z) = \frac{P(z \mid open)P(open)}{P(z \mid open)p(open) + P(z \mid \neg open)p(\neg open)}$$

$$P(open \mid z) = \frac{0.6 \cdot 0.5}{0.6 \cdot 0.5 + 0.3 \cdot 0.5} = \frac{2}{3} = 0.67$$

- z raises the probability that the door is open.

MTH 2215
Applied Discrete Mathematics

Chapter 6.4
Expected Value and Variance

These class notes are based on material from our textbook, **Discrete Mathematics and Its Applications**, 6^{th} ed., by Kenneth H. Rosen, published by McGraw Hill, Boston, MA, 2006. They are intended for classroom use only and are **not** a substitute for reading the textbook.

Expected Values

- For any random variable V having a numeric domain, its *expectation value* or *expected value* or *weighted average value* or *(arithmetic) mean value* **Ex**$[V]$, under the probability distribution $\Pr[v] = p(v)$, is defined as

$$\hat{V} :\equiv \mathbf{Ex}[V] :\equiv \mathbf{Ex}_p[V] :\equiv \sum_{v \in \mathbf{dom}[V]} v \cdot p(v).$$

- The term "expected value" is very widely used for this.
  - But this term is somewhat misleading, since the "expected" value might itself be totally unexpected, or even impossible!
    - *E.g.*, if $p(0)=0.5$ & $p(2)=0.5$, then $\mathrm{Ex}[V]=1$, even though $p(1)=0$ and so we know that $V\neq 1$!
    - Or, if $p(0)=0.5$ & $p(1)=0.5$, then $\mathrm{Ex}[V]=0.5$ even if V is an integer variable!

Derived Random Variables

- Let S be a sample space over values of a random variable V (representing possible outcomes).
- Then, any function f over S can also be considered to be a random variable (whose actual value $f(V)$ is derived from the actual value of V).
- If the range R = **range**[f] of f is numeric, then the mean value **Ex**[f] of f can still be defined, as

$$\hat{f} = \mathbf{Ex}[f] = \sum_{s \in S} p(s) \cdot f(s)$$

Recall that a random variable X is actually a function

$$f: S \to X(S),$$

where S is the sample space and $X(S)$ is the range of X. This fact implies that the **expected value** of X is

$$E(X) = \sum_{s \in S} p(s) X(s) = \sum_{r \in X(S)} p(X = r) r$$

Example 1. **Expected Value of a Die**.
Let X be the number that comes up when a die is rolled.

$$E(X) = \sum_{r \in \{1,\ldots,6\}} p(X = r) r = \sum_{r \in \{1,\ldots,6\}} \frac{1}{6} r = \frac{7}{2}$$

Example 2

- A fair coin is flipped 3 times. Let *S* be the sample space of 8 possible outcomes, and let *X* be a random variable that assignees to an outcome the number of heads in this outcome.

E(X) = 1/8[X(TTT) + X(TTH) + X(THH) + X(HTT) + X(HHT) +X(HHH) + X(THT) + X(HTH)]

= 1/8[0 + 1 + 2 + 1 + 2 + 3 + 1 + 2] = 12/8 =3/2

Linearity of Expectation Values

- Let X_1, X_2 be any two random variables derived from the *same* sample space S, and subject to the same underlying distribution.
- Then we have the following theorems:

  $\mathbf{Ex}[X_1+X_2] = \mathbf{Ex}[X_1] + \mathbf{Ex}[X_2]$

  $\mathbf{Ex}[aX_1 + b] = a\mathbf{Ex}[X_1] + b$
- You should be able to easily prove these for yourself at home.

Variance & Standard Deviation

- The *variance* **Var**[X] = $\sigma^2(X)$ of a random variable X is the expected value of the *square* of the difference between the value of X and its expectation value **Ex**[X]:

$$\mathbf{Var}[X] :\equiv \sum_{s \in S} \left(X(s) - \mathbf{Ex}_p[X] \right)^2 p(s)$$

- The *standard deviation* or *root-mean-square* (RMS) *difference* of X is $\sigma(X) :\equiv \mathbf{Var}[X]^{1/2}$.

Example 15

- What is the variance of the random variable *X*, where *X* is the number that comes up when a die is rolled?
- $V(X) = E(X^2) - E(X)^2$

$E(X^2) = 1/6[1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2] = 91/6$

$V(X) = 91/6 - (7/2)^2 = 35/12 \approx 2.92$