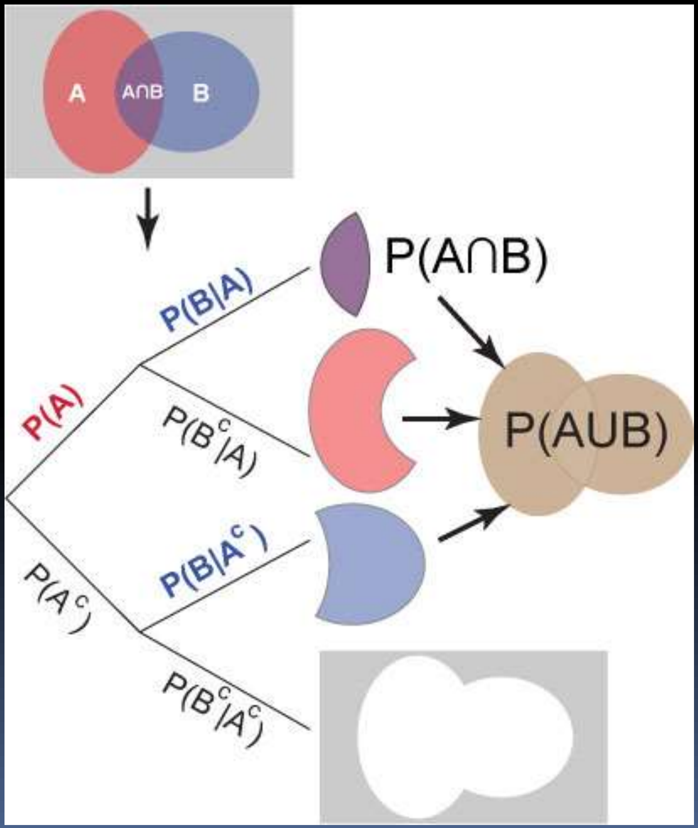
15-251: Great Theoretical Ideas in Computer Science

Fall 2016 Lecture 17

October 25, 2016

Probability 1



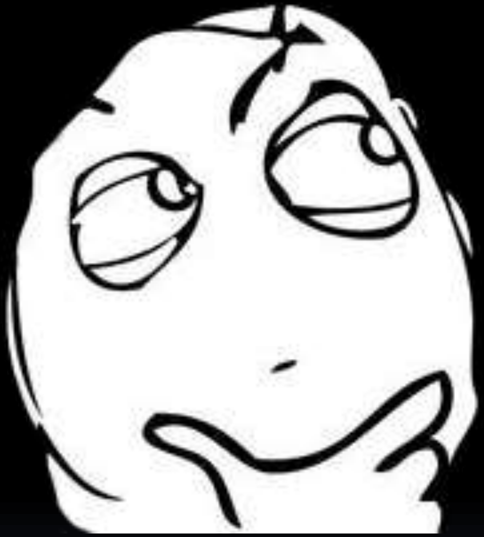
France, 1654

“Chevalier de Méré”
AKA Antoine Gombaud

Let’s bet:

I will roll a die four times.
I win if I get a 1.

(not actually Méré)



France, 1654

Antoine Gombaud,
AKA "Chevalier de Méré"

Hmm.

No one wants to take this bet any more.

France, 1654

Antoine Gombaud,
AKA “Chevalier de Méré”

New bet:

I will roll two dice, 24 times.
I win if I get double-1’s.

France, 1654

Antoine Gombaud,
AKA “Chevalier de Méré”

Hmm.

I keep losing money!

Problem of Points

Alice and Bob are flipping a coin. Alice gets a point for Heads, Bob a point for Tails. First one to 4 points wins the stake of 100 francs.

Alice is leading 3-2 when gendarmes arrive to break up the game. How should they divide the stakes?

Pascal

Fermat

Probability Theory is Born

Moral of the Story

Analyzing gambling is not
a side-benefit of probability.

Probability was invented
to analyze gambling.

This is not
“Great Theoretical Ideas
in Gambling”

This is
“Great Theoretical Ideas
in Computer Science”

Why study probability?

Randomness is essential for computer science!

- Modeling/simulation **requires** randomness.
- Cryptography **requires** randomness.
- Some very basic problems (e.g., Primality, Polynomial Factorization) *seem* to be solvable faster using randomness.
- **Many** algorithms these days use randomness; "deterministic" algorithms seem quaint!

Teams A and B are equally good

In any one game, each is equally likely to win

What is most likely length of a "best of 7" series?

Flip coins until either 4 heads or 4 tails
Is this more likely to take 6 or 7 flips?

Poll

6 and 7 Are Equally Likely

To reach either one, after 5 games, it must be 3 to 2

½ chance it ends 4 to 2; ½ chance it doesn't

Teams A is now better than team B

The odds of A winning are 6:5

i.e., in any game, A wins with probability 6/11

What is the chance that A will beat B in the “best of 7” world series?

We’ll come back to it; let’s start with basics

Probability Theory
=
Analyzing Behavior of Experiments that Have Randomness in Them

Simple random experiments:

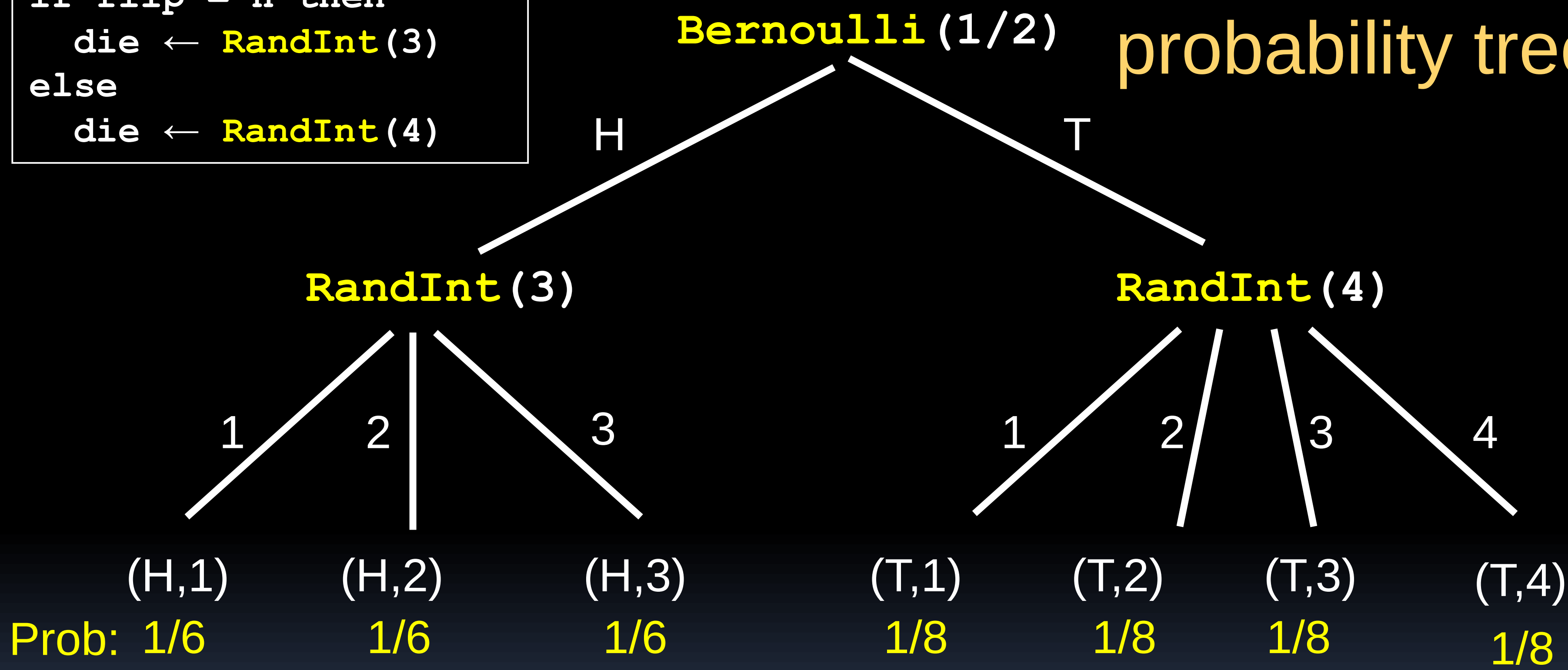
- Toss a fair coin: Bernoulli(1/2)
- Toss a fair 6-sided die: RandInt(6)
- Toss a biased coin: Bernoulli(0.51)
- Pick a random student from Fall’16 edition of 15-251: RandInt(104)

These experiments can be combined and repeated many times, in adaptive fashions

Mary flips a fair coin. If it's heads, she rolls a 3-sided die. If it's tails, she rolls a 4-sided die.

```
flip ← Bernoulli(1/2)
if flip = H then
  die ← RandInt(3)
else
  die ← RandInt(4)
```

We can draw a probability tree.



Have branching for each call to a generator

Label the leaves with **"outcomes"**

Under each, write its probability: multiply along the path

Outcome:

A leaf in the probability tree.

I.e., a possible sequence of values of all calls to random generators in an execution.

Sample Space:

The **set** of all outcomes.

E.g., { (H,1), (H,2), (H,3), (T,1), (T,2), (T,3), (T,4) }

Probability:

Each outcome has a nonnegative probability.

Sum of all outcomes' probabilities always 1.

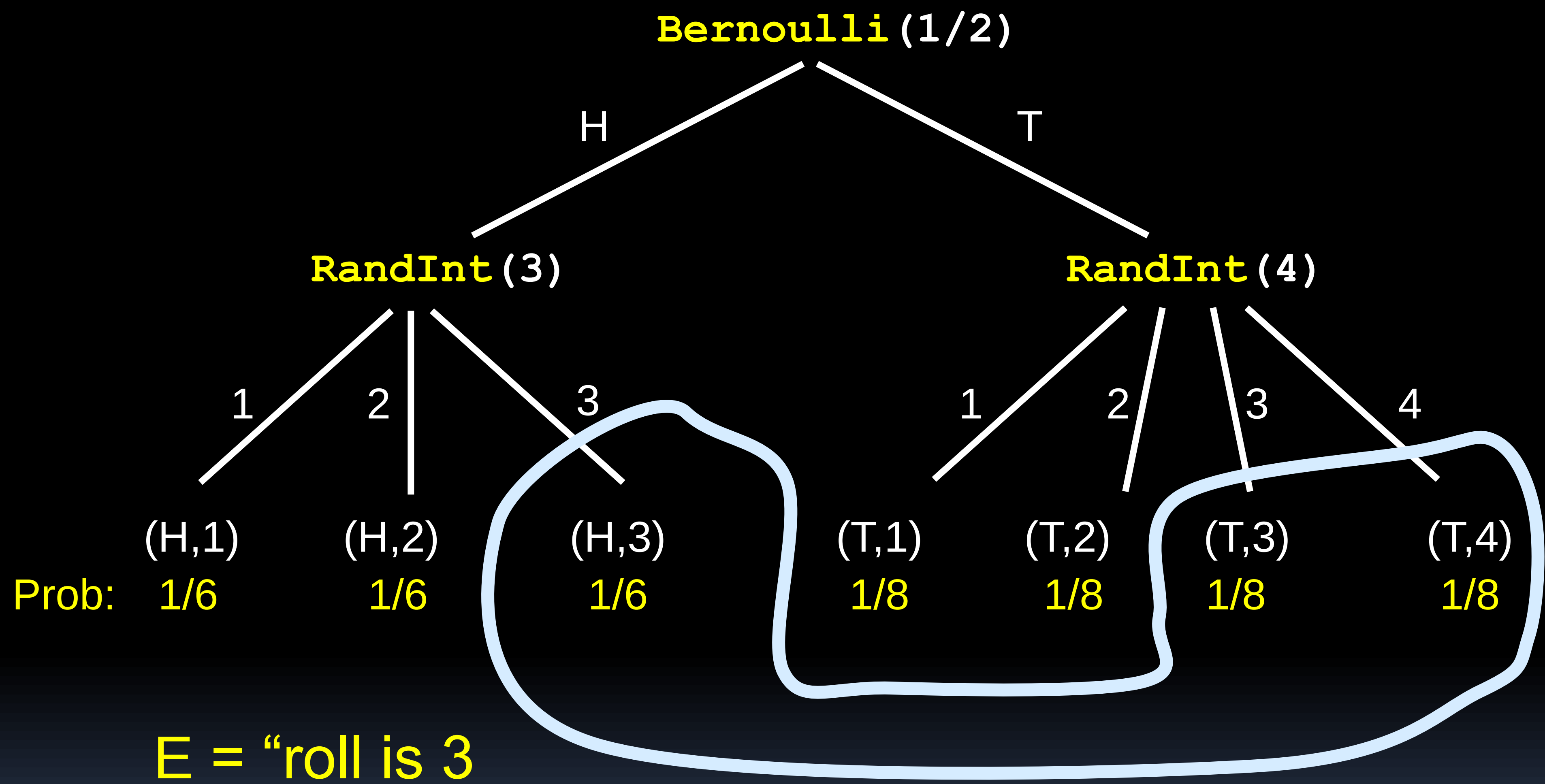
Mary flips a fair coin. If it's heads, she rolls a 3-sided die. If it's tails, she rolls a 4-sided die.

What is the probability die roll is 3 or higher?

Event:

A ***subset*** of outcomes.
In our example, E = { (H,3), (T,3), (T,4) }.

Pr[E] = sum of the probabilities of the outcomes in E.



E = "roll is 3 or higher"

Pr[E] = 1/6 + 1/8 + 1/8 = 5/12

A fair coin is tossed 100 times in a row

What is the probability that we get exactly 50 heads?

What is the <u>sample space</u> Ω?

Answer: $\{H,T\}^{100}$
(the set of all outcomes)

Each sequence in Ω is equally likely, and hence has probability

$$\frac{1}{|\Omega|} = \frac{1}{2^{100}}$$

The Language of Probability

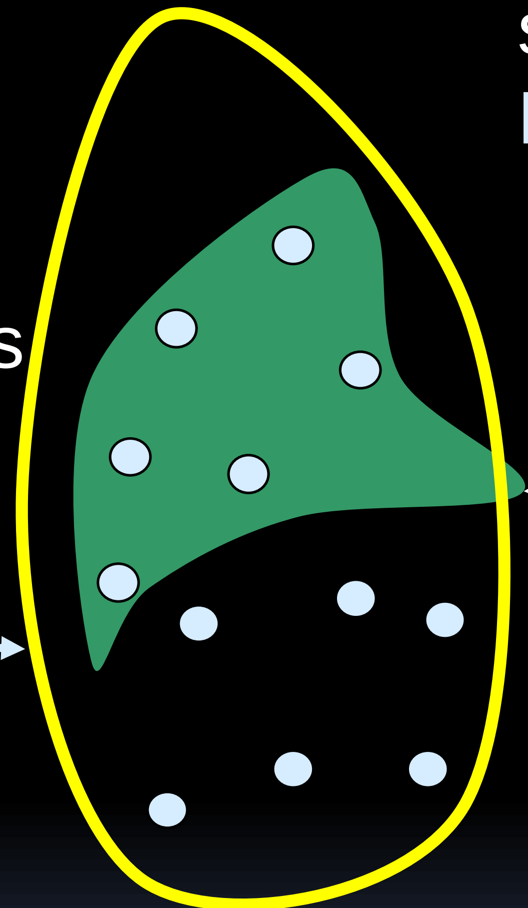
"What is the probability that we get exactly 50 heads?"

Let $E = \{x \text{ in } \Omega \mid x \text{ has 50 heads}\}$ be the <u>event</u> that we see half heads.

$$\Pr[E] = |E|/|\Omega| = |E|/2^{100}$$

Event E = Subset of sequences with 50 H’s and 50 T’s

Sample space Ω
all 2^{100} sequences
$\{H,T\}^{100}$



Probability of event E = proportion of E in Ω

$$\binom{100}{50} / 2^{100} = 0.07958923739\ldots$$

Finite Probability Distribution

A (finite) probability distribution D is a finite set Ω of elements, where each element $t \in \Omega$ has a *non-negative* real weight, proportion, or

probability p(t)

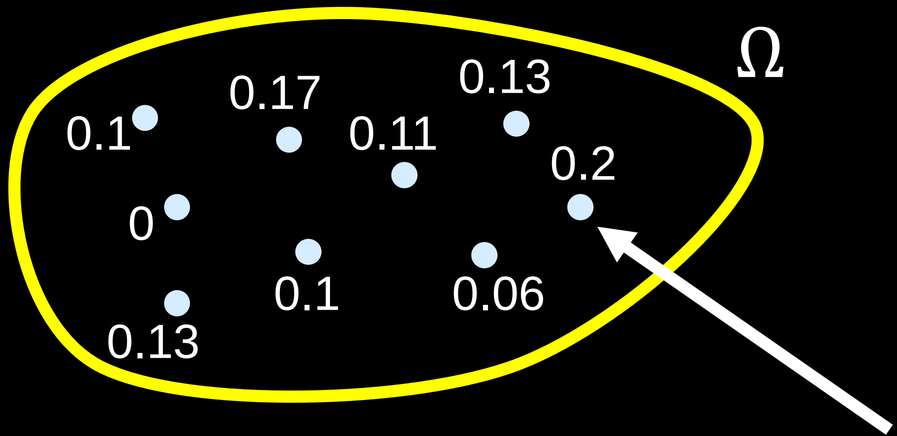
The weights must satisfy:

$$\sum_{t \in \Omega} p(t) = 1$$

For convenience we will define D(t) = p(t)

Ω is often called the **sample space** and elements t in Ω are called samples or outcomes

Sample Space



Sample space

weight or
probability of t

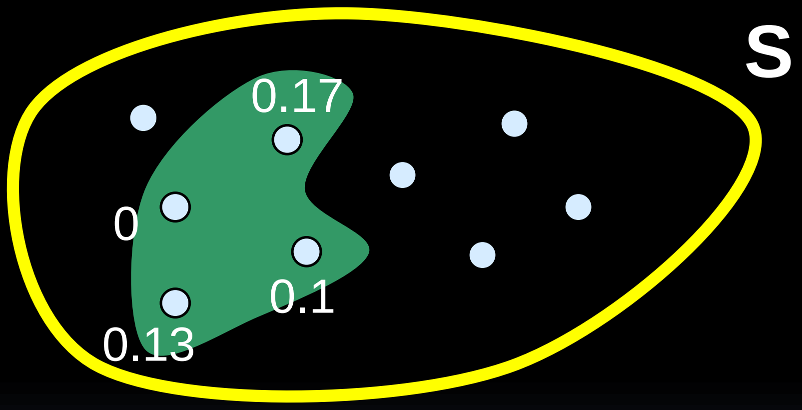
D(t) = p(t) = 0.2

Events

Any set $E \subseteq \Omega$ is called an event

Probability of event E is

$$Pr_D[E] = \sum_{t \in E} p(t)$$



$Pr_D[E] = 0.4$

Uniform Distribution

If each element has equal probability, the distribution is said to be uniform

$$\Pr_D[E] = \sum_{t \in E} p(t) = \frac{|E|}{|\Omega|}$$

France, 1654

Alice and Bob are flipping a coin. Alice gets a point for Heads, Bob a point for Tails. First one to 4 points wins the stake of 100 francs.

Alice is leading 3-2 when gendarmes arrive to break up the game. How should they divide the stakes?

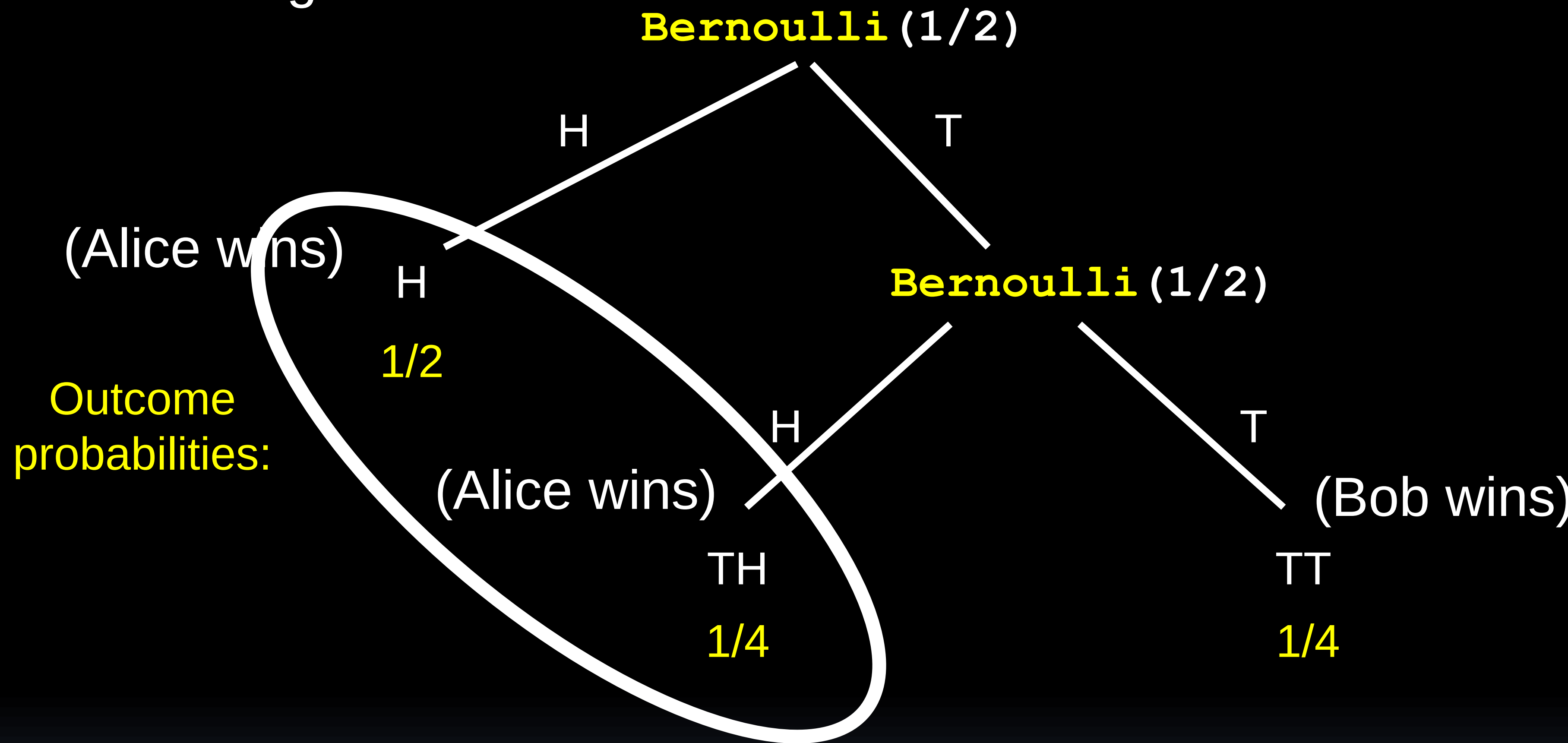
France, 1654

It seems fair that Alice should get

(100 francs) x **Pr**[Alice would win].

So let’s compute that!

Alice leading 3-2:



Event A = “Alice wins” = { H, TH }

Pr[A] = 1/2 + 1/4 = 3/4

Axioms of probability

Note that probabilities satisfy the following properties:

1) $\Pr[\Omega] = 1$

2) $\Pr[E] \geq 0$ for all events E

3) $\Pr[A \cup B] = \Pr(A) + \Pr(B)$,
for **disjoint** events A and B

} axioms of probability

Hence, $\Pr[\overline{A}] = 1 - \Pr[A]$
(Prob. of "not A", i.e., event A does not occur)

Some more useful facts

For any events A and B,

$$\Pr[A] = \Pr[A \cap B] + \Pr[A \cap \overline{B}]$$

Inclusion-Exclusion!

For any events A and B,

$$\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$$

Corollary: For any events A and B,

$$\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$$

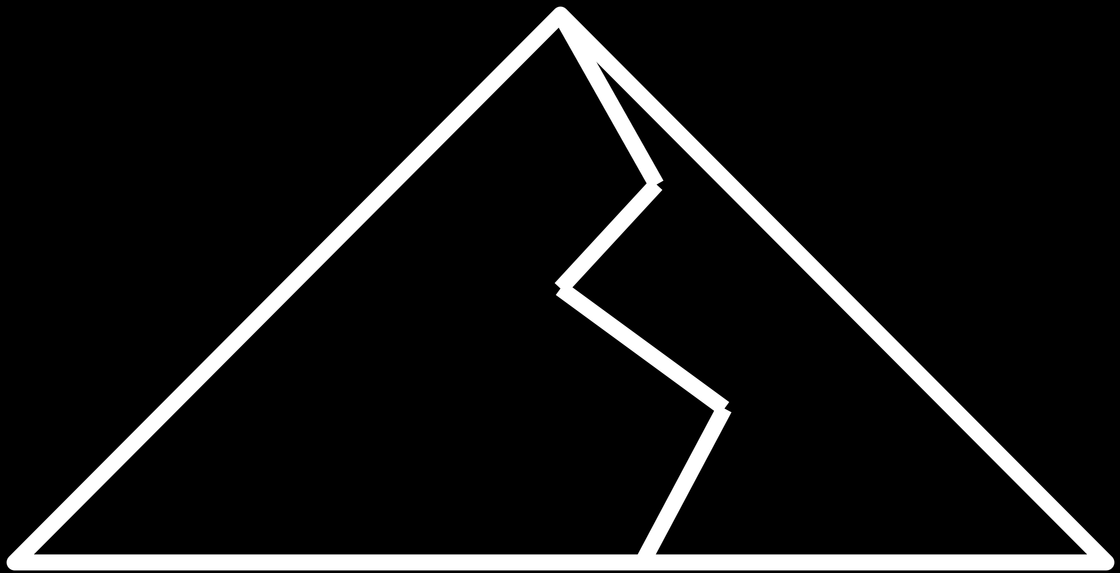
"Union-Bound"

"Boole's inequality"

$$\Pr(A_1 \cup A_2 \cup \cdots \cup A_n) \leq \Pr(A_1) + \Pr(A_2) + \cdots + \Pr(A_n)$$

Let’s bet:

I will roll a die four times.
I win if I get a 1.



Let W be the event that Méré wins

Easier to compute $\Pr[\overline{W}]$

$\overline{W}$ ={ all outcomes with no 1's }

$|\overline{W}| = 5^4$

$\therefore \Pr(\overline{W}) = 5^4/6^4$

$\therefore$ **Pr**[W] = $1 - 5^4/6^4 \approx 51.8\%$

Let’s bet:

I will roll two dice 24 times.
I win if I get a double-1’s.

Pr[Méré wins] =

$1-35^{24}/36^{24}$

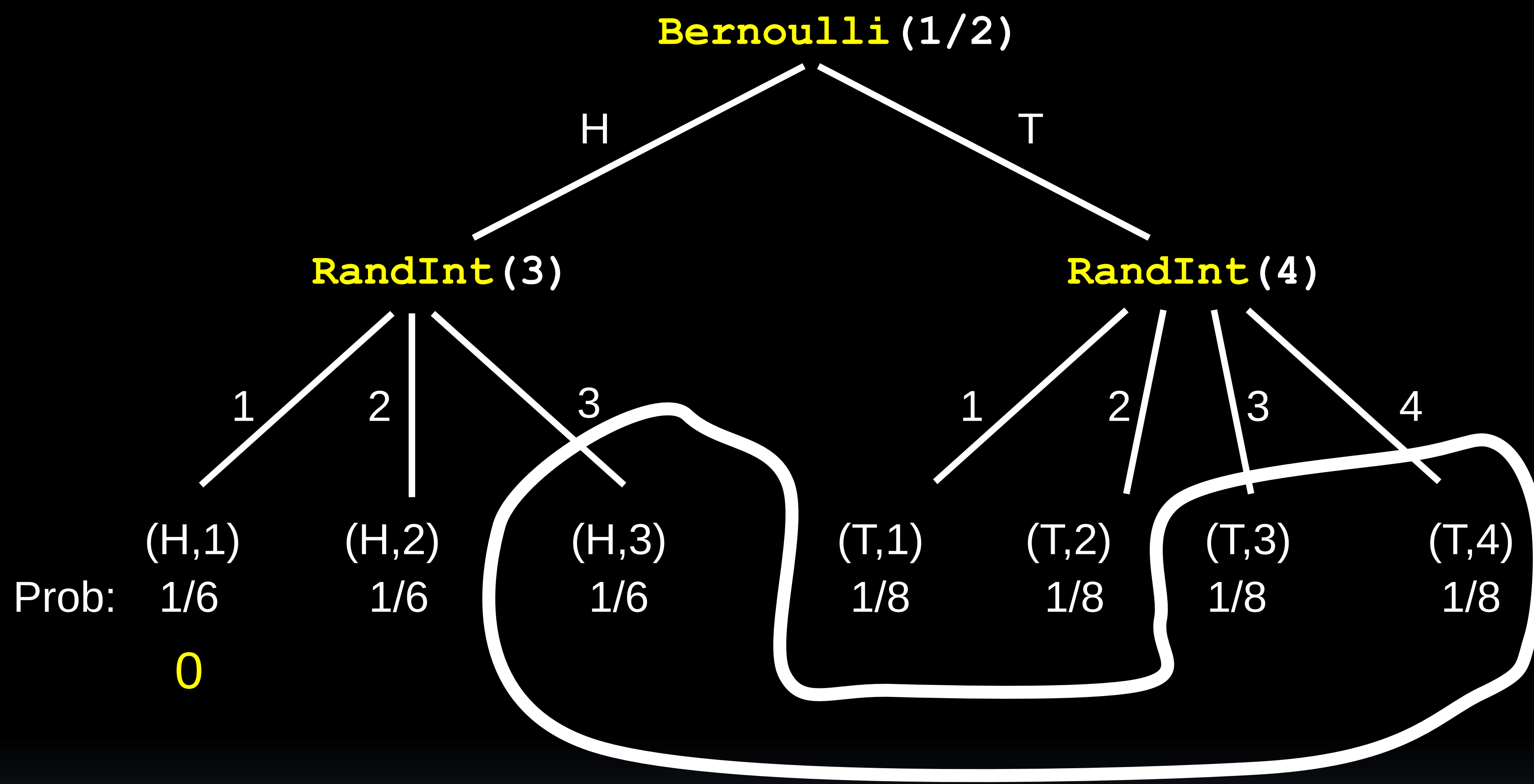
≈ 49.1%

Conditioning

= Revising probabilities based on ‘partial information’

‘Partial information’ = an event

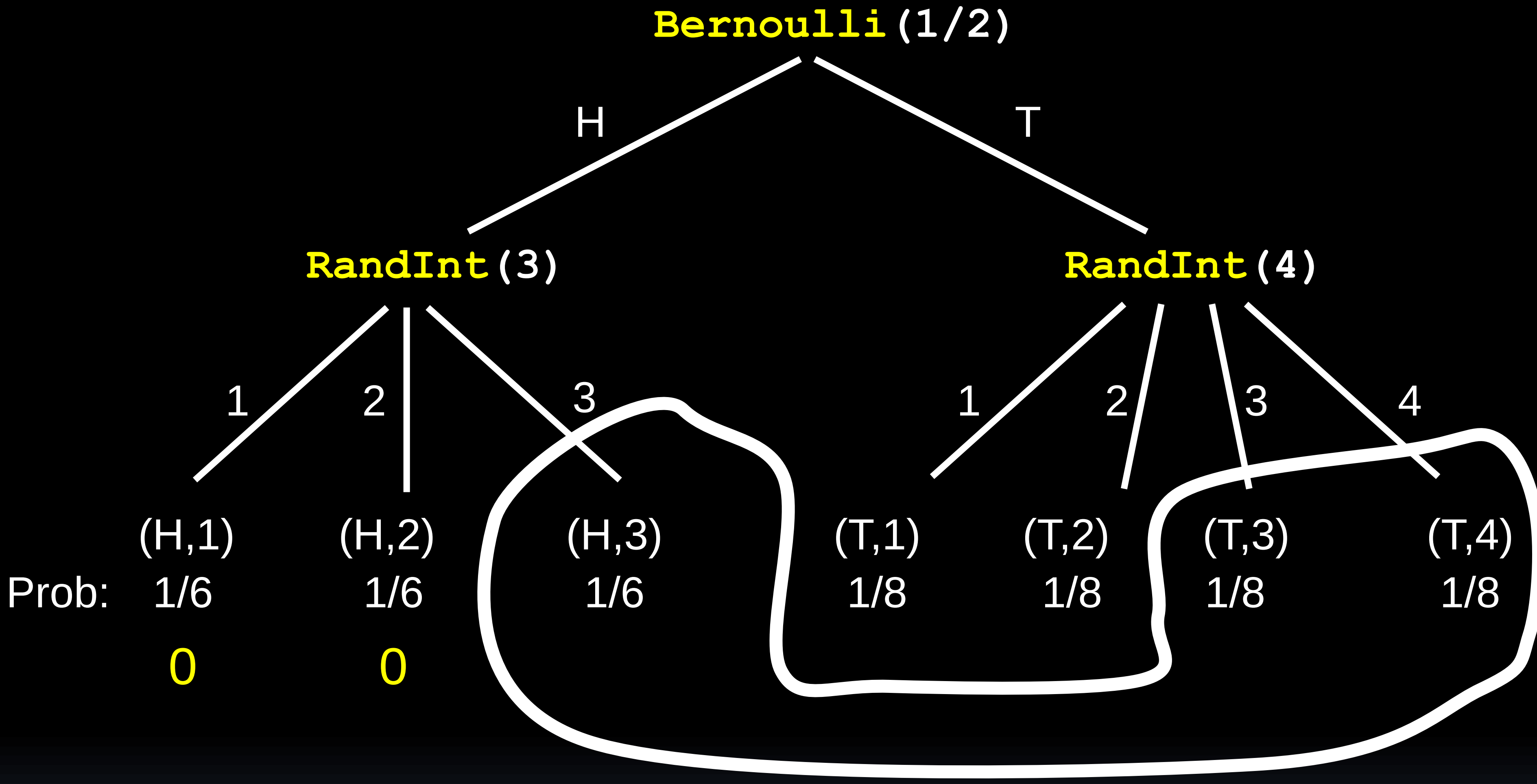
‘Conditioning on event A’ is like assuming/promising A occurs.



Condition on S, the event “roll is 3 or higher”

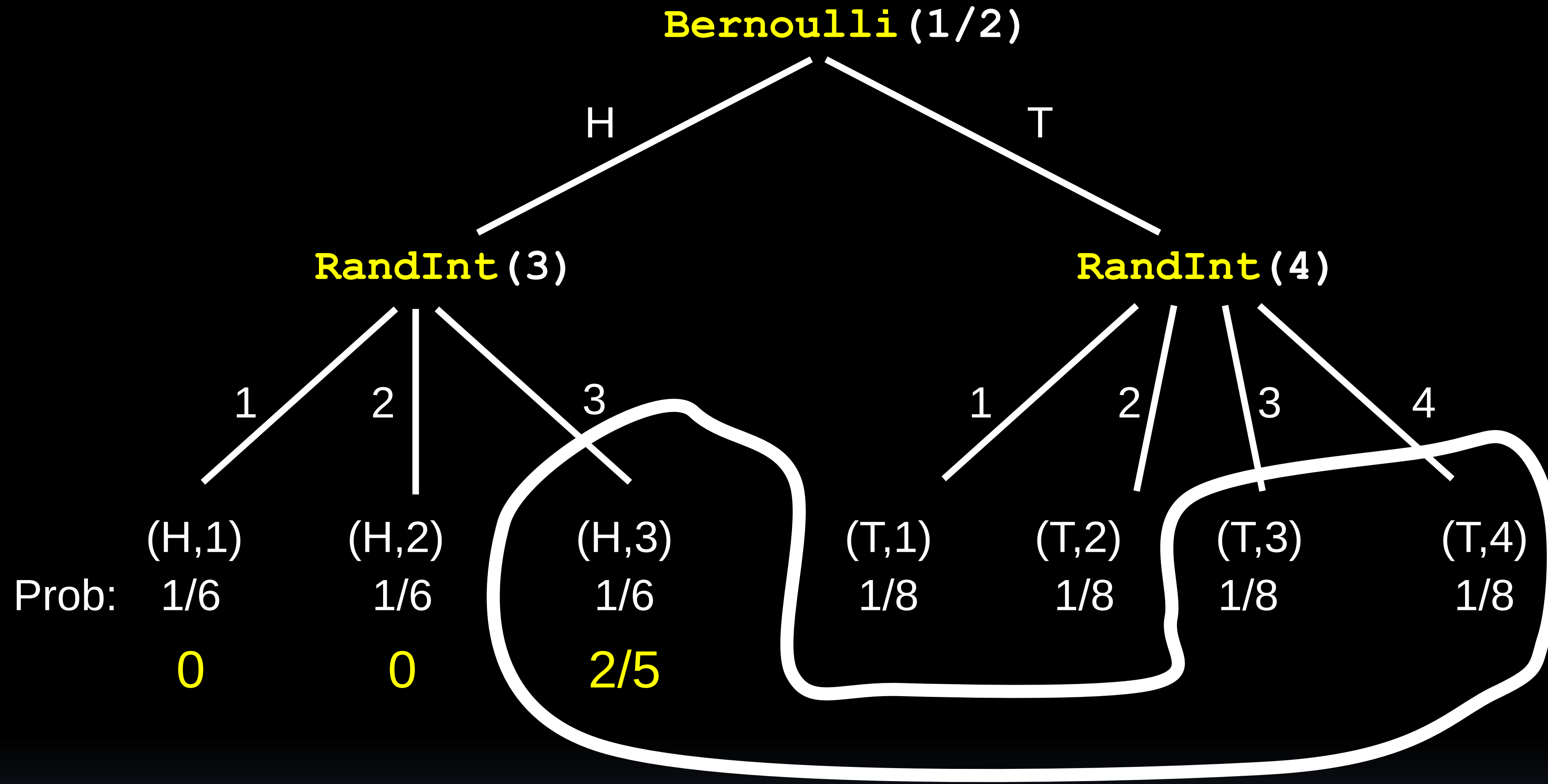
Pr [(H,1) | S] = 0

“probability of outcome (H,1) conditioned on event S”



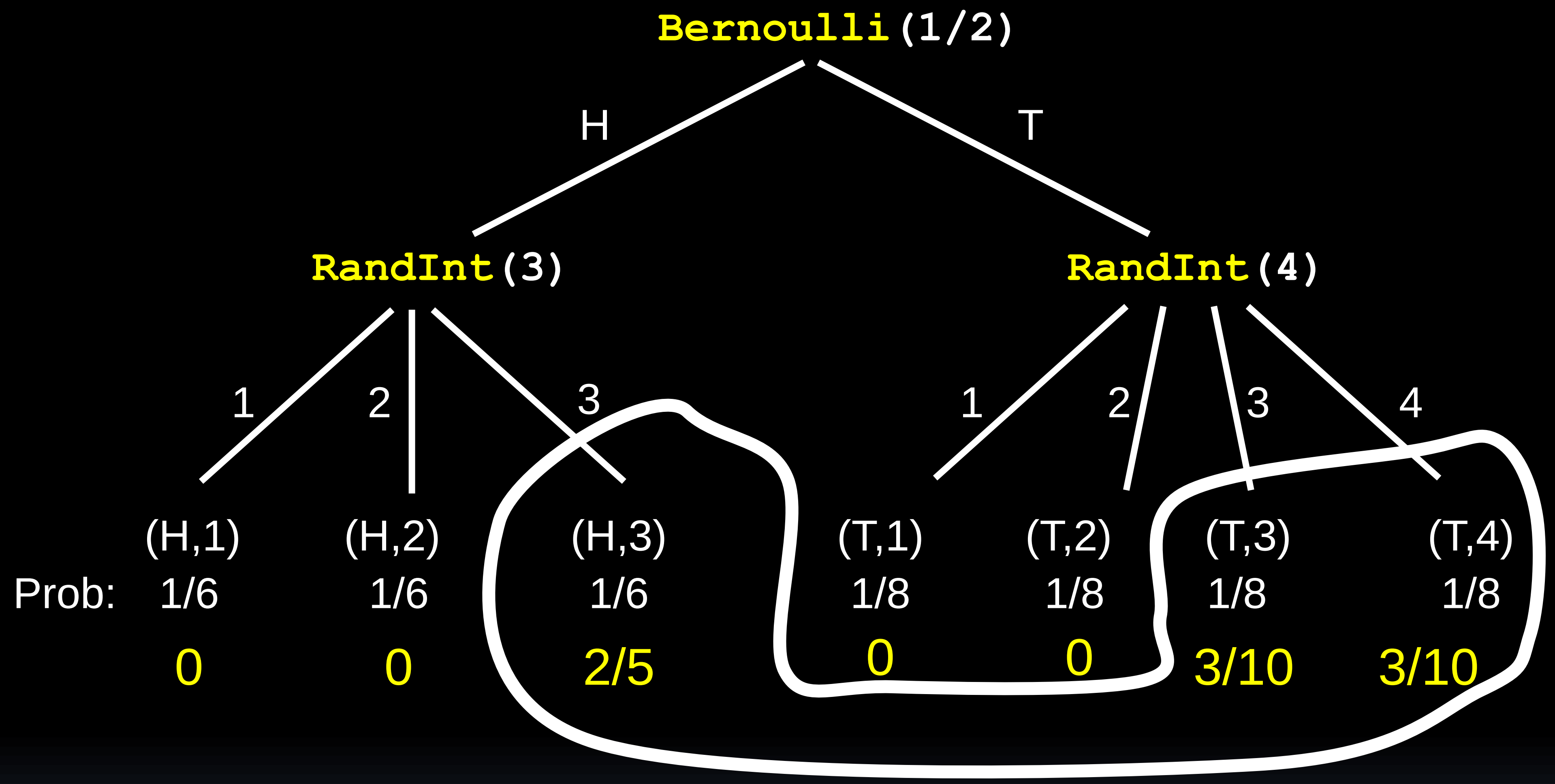
Condition on S, the event “roll is 3 or higher”

Pr [(H,2) | S] = 0



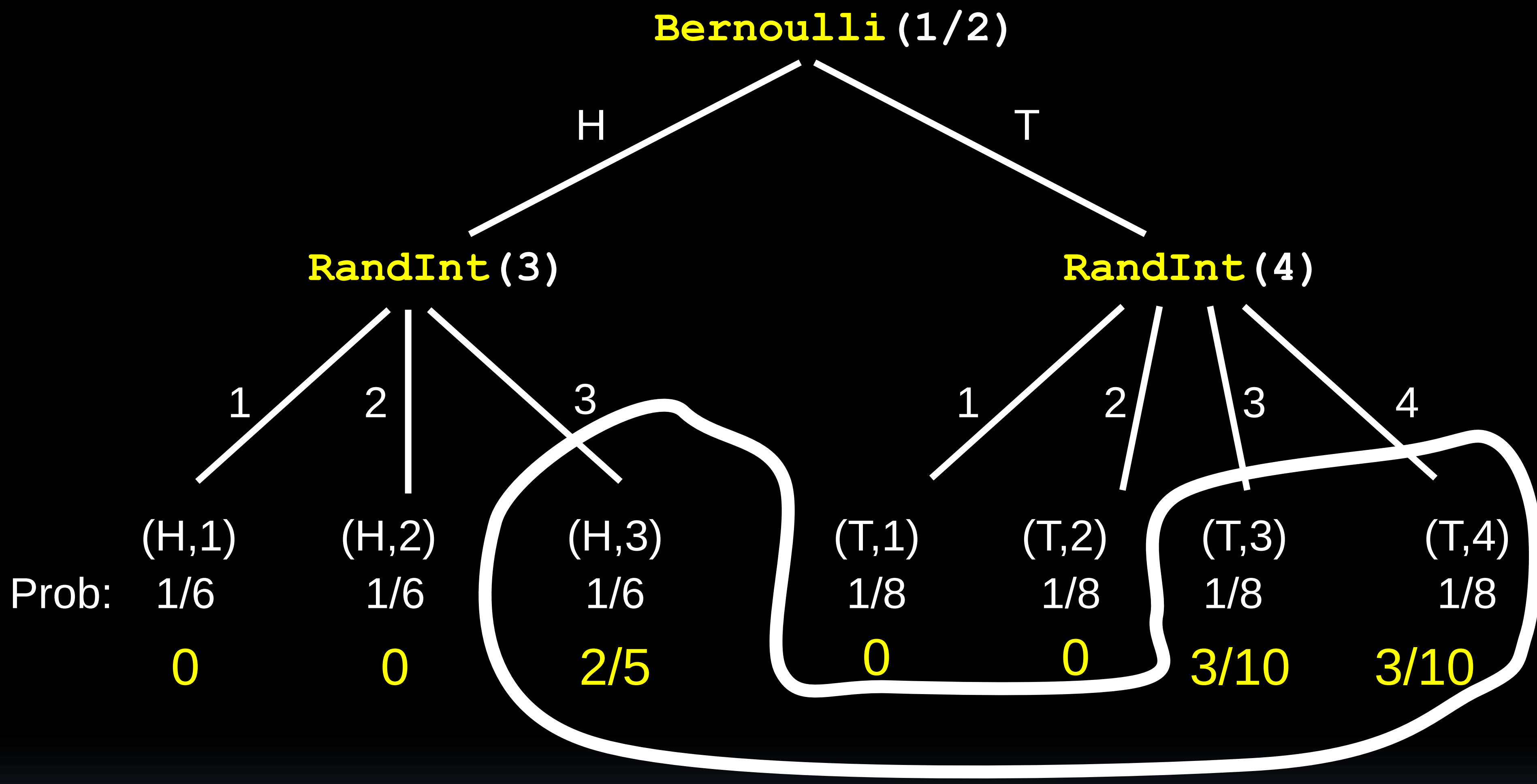
Condition on S, the event “roll is 3 or higher”

$$\textbf{Pr}\,[(H,3) \mid S] = \frac{1/6}{5/12} = 2/5$$



Condition on S, the event “roll is 3 or higher”

$$\textbf{Pr}\,[(T,3) \mid S] = \frac{1/8}{5/12} = 3/10$$



Condition on S, the event “roll is 3 or higher”

Let A be the event that Tails was flipped.

Pr [A | S] = 0+0+3/10+3/10 = 3/5

Conditioning: formally

Given an experiment, let A be an event.

(with nonzero probability)

The conditional probability of outcome ℓ is

$$\mathbf{Pr}[\ell \mid A] = \begin{cases} 0 & \text{if } \ell \notin A, \\ \frac{\mathbf{Pr}[\ell]}{\mathbf{Pr}[A]} & \text{if } \ell \in A. \end{cases}$$

$$\therefore \mathbf{Pr}[B \mid A] = \sum_{\ell \in B} \mathbf{Pr}[\ell \mid A] = \sum_{\ell \in B \cap A} \frac{\mathbf{Pr}[\ell]}{\mathbf{Pr}[A]} = \frac{\mathbf{Pr}[B \cap A]}{\mathbf{Pr}[A]}$$

"Chain Rule"

$$\mathbf{Pr}[A \cap B] = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B \mid A]$$

"For A ***and*** *B to occur, first A must occur (probability* **Pr**[A] *), and then B must occur given that A occurred (probability* **Pr**[B | A] *)."*

"Chain Rule"

$$\mathbf{Pr}[A \cap B] = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B \mid A]$$

$$\mathbf{Pr}[A \cap B \cap C] = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B \mid A] \cdot \mathbf{Pr}[C \mid A \cap B]$$

$$\mathbf{Pr}[A \cap B \cap C \cap D] = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B \mid A] \cdot \mathbf{Pr}[C \mid A \cap B] \cdot \mathbf{Pr}[D \mid A \cap B \cap C]$$

etc.

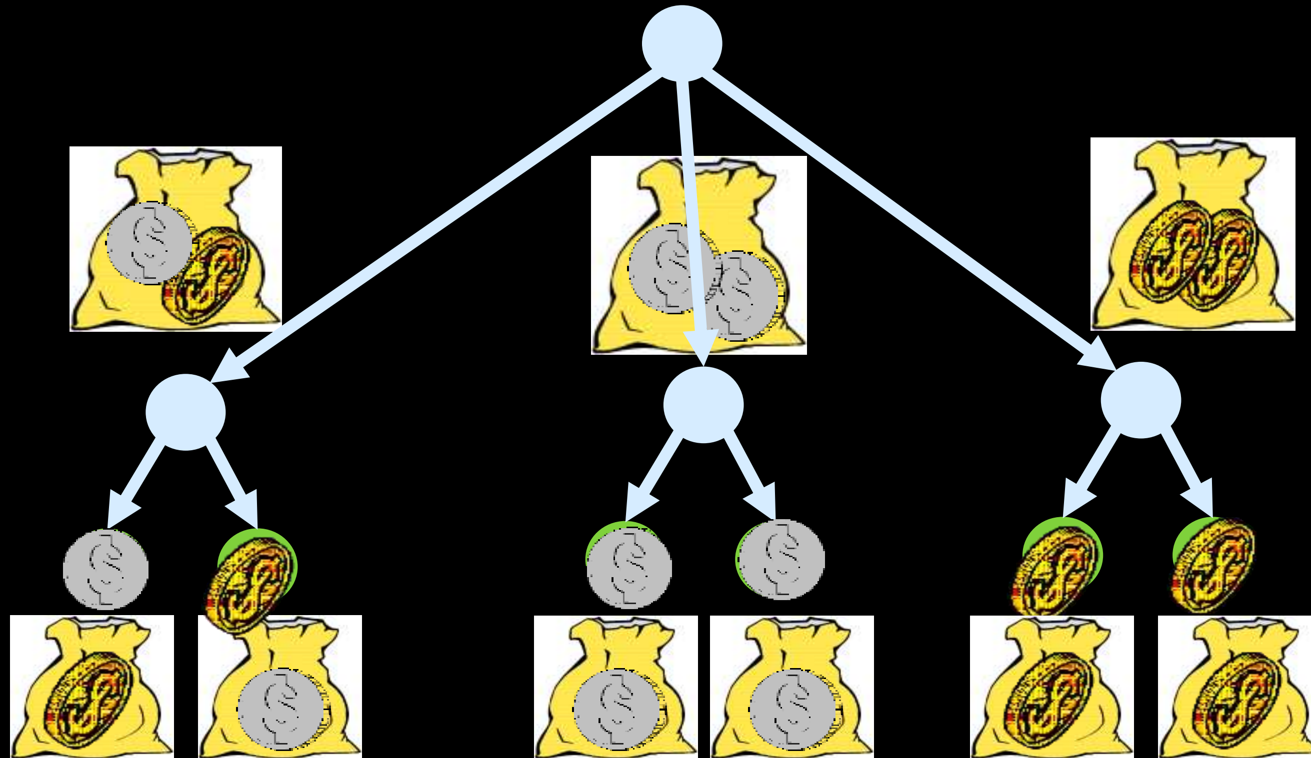
Silver and Gold: a problem

One bag contains two silver coins.
Another contains two gold coins.
Another contains one silver and one gold.

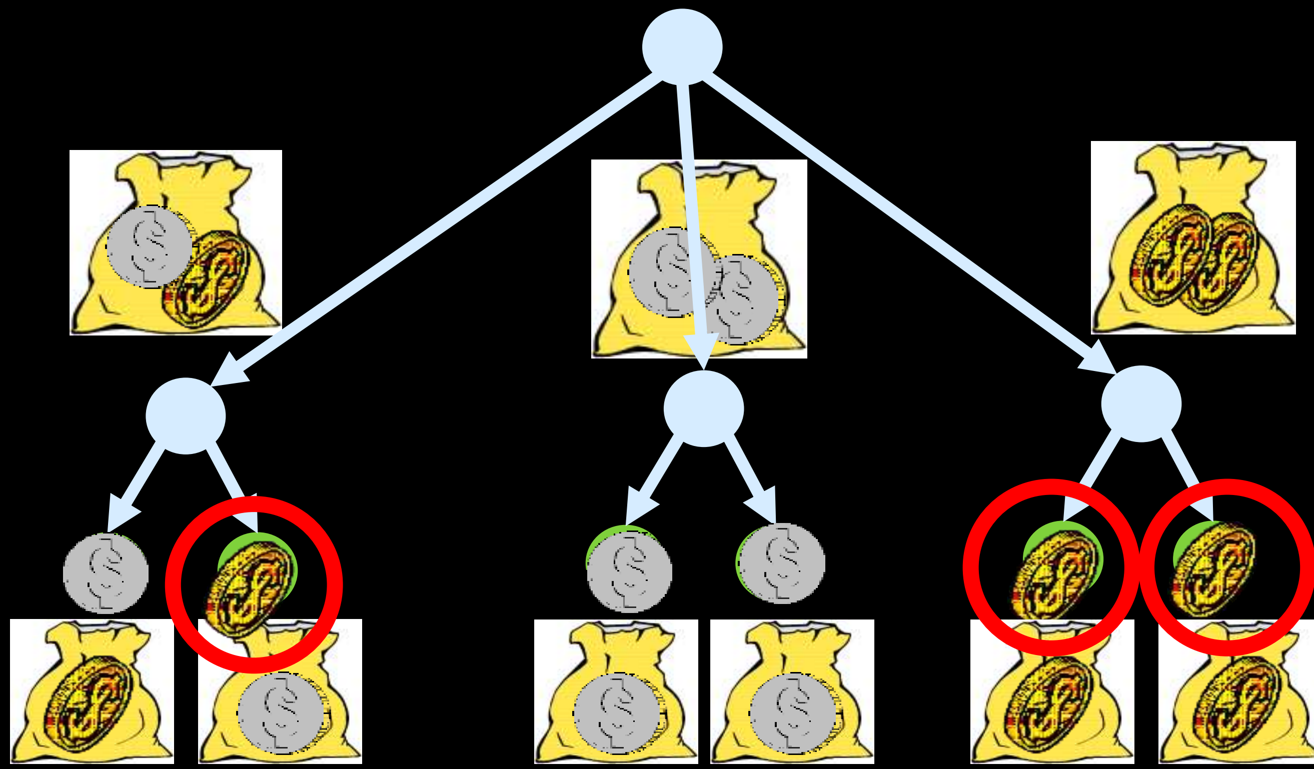
Poll

Anil picks a bag at random,
then picks a coin from it at random.

It turns out to be gold. What is the probability the *other* coin in his bag is gold?



3 choices of bag
2 ways to pick one of its two coins
6 equally likely paths

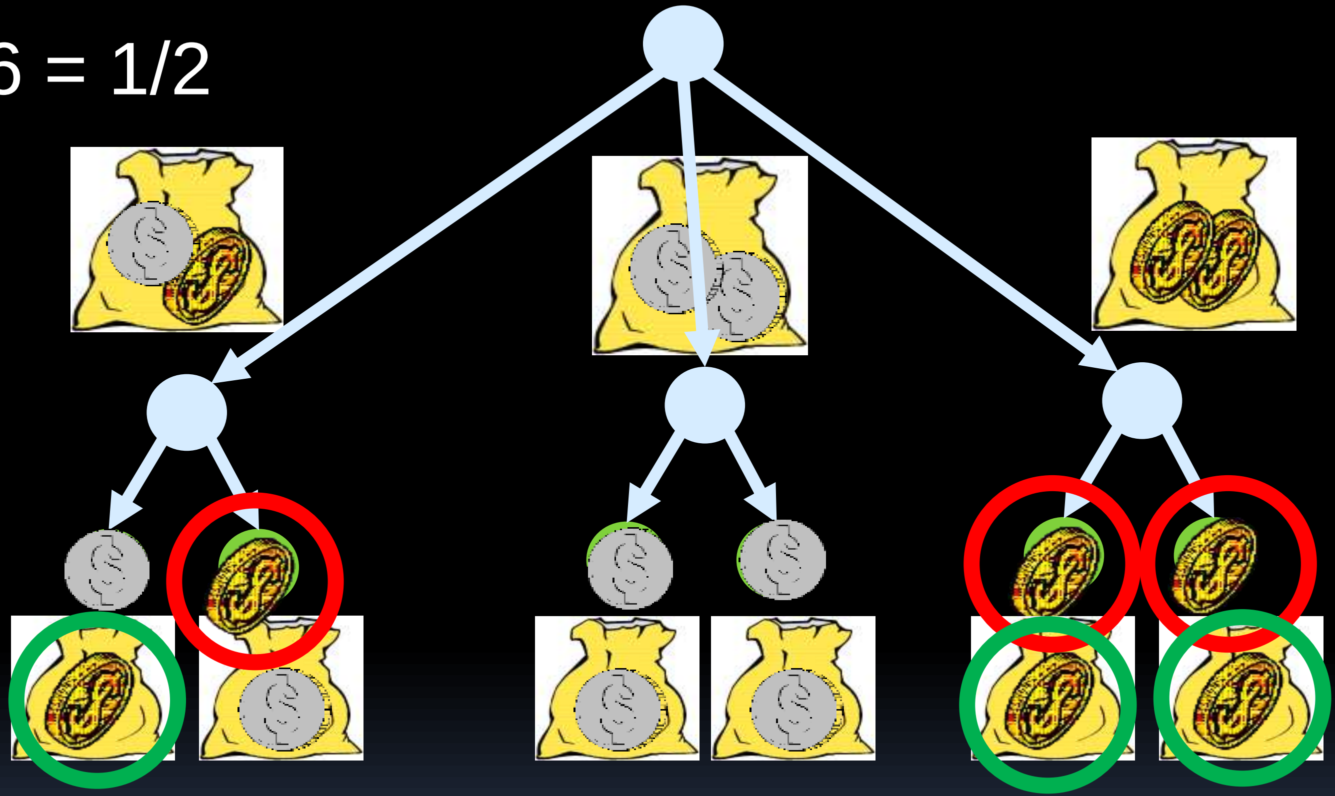


Given that we see a gold,
2/3 of possible paths have gold
as other coin!

Formally…

Let G_1 be the event that the coin pulled out is gold

$Pr[G_1] = 3/6 = 1/2$

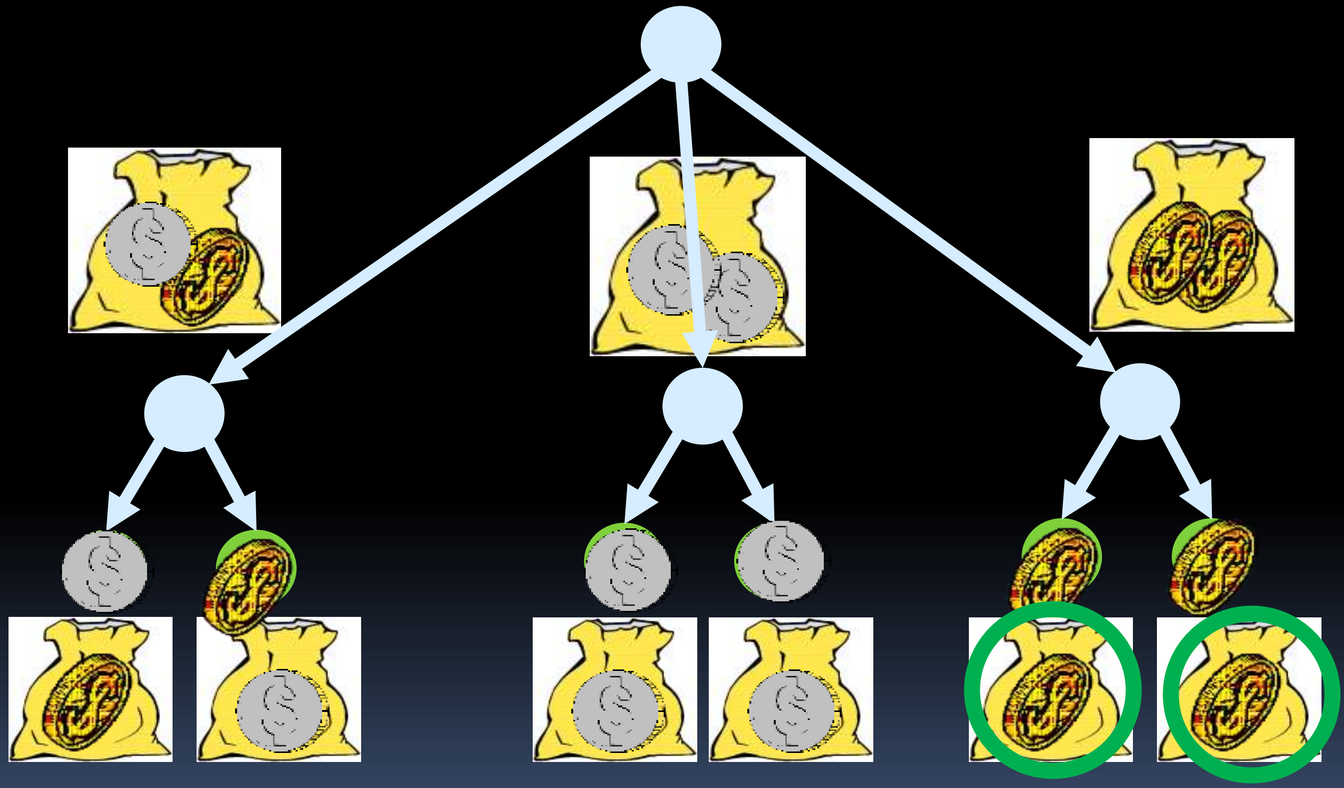


Let G_2 be the event that the second coin in the chosen bag is gold

$Pr[G_2] = 3/6 = 1/2$

Joint probability

$\Pr[G_1 \cap G_2] = \; 2/6 = 1/3$



Conditional probability

Pr[second coin is gold | coin we pulled out is gold]

$= \Pr[G_2 \mid G_1] = \Pr[G_1 \cap G_2] / \Pr[G_1]$

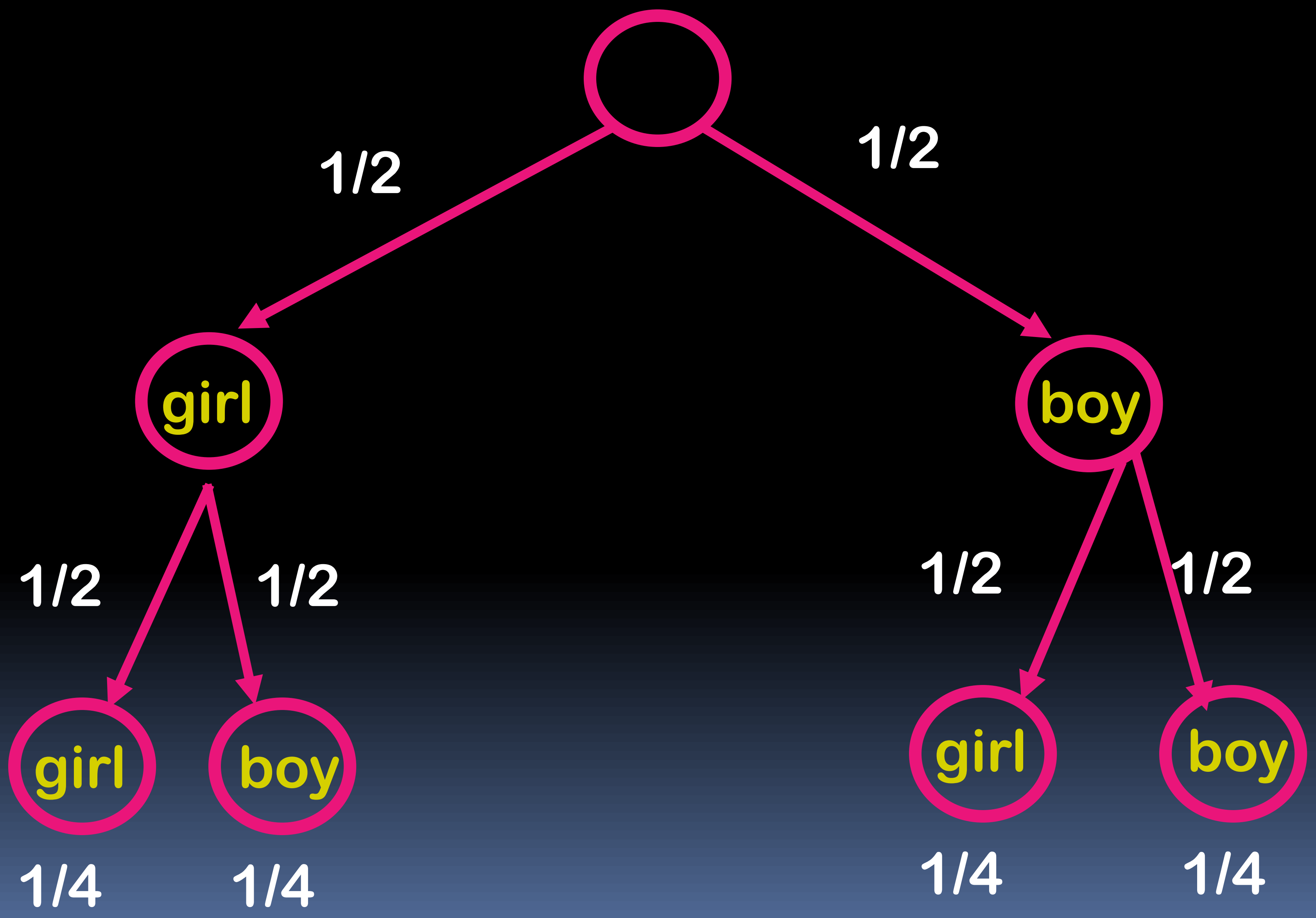
$= (1/3) / (1/2)$

$= 2/3$

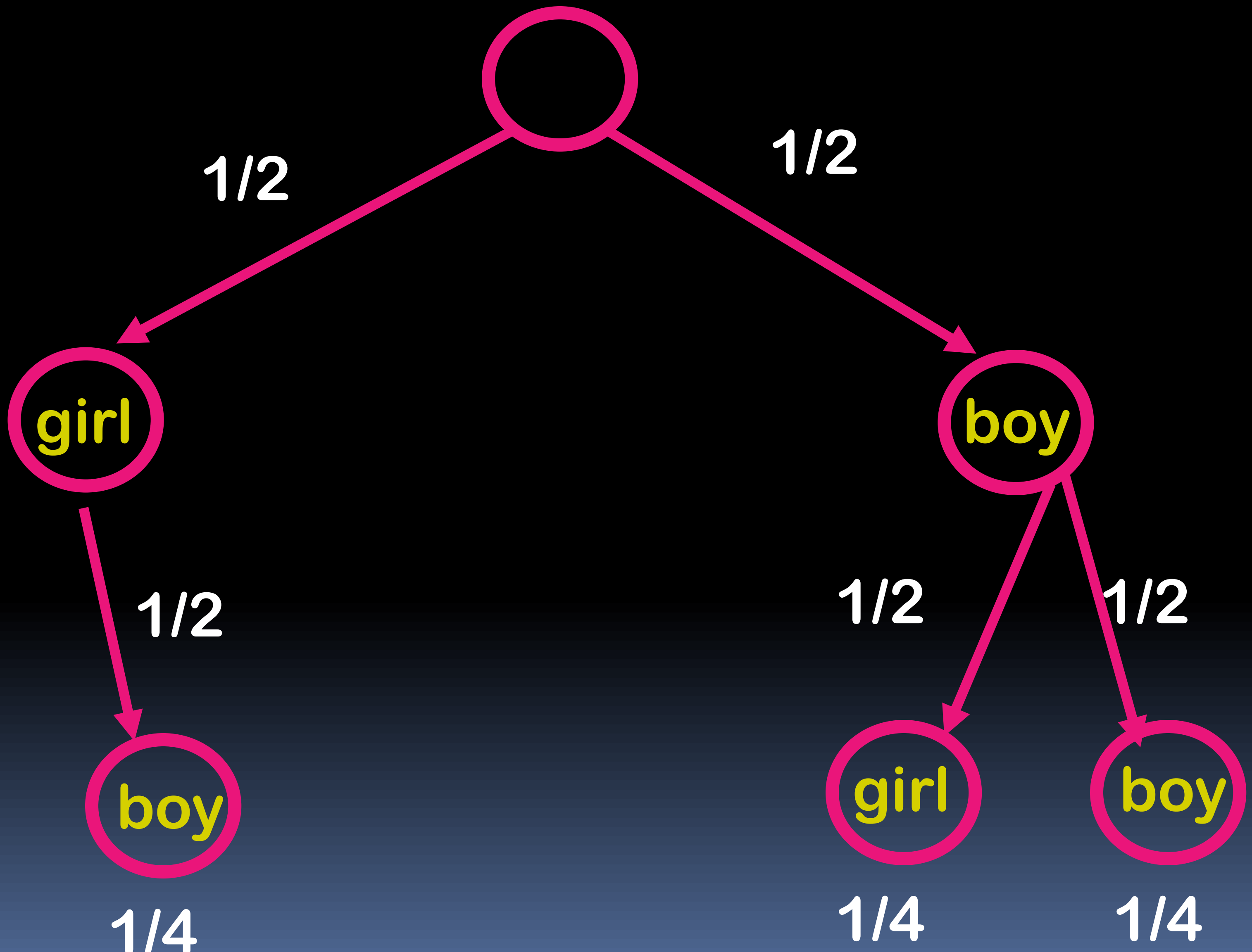
Boys and Girls

Consider a family with two children. Given that one of the children is a boy, what is the probability that both children are boys?

1/3


1/2
1/2
girl
boy
1/2
1/2
1/2
1/2
girl
boy
girl
boy
1/4
1/4
1/4
1/4

conditioning on at least one boy…



Boys and Girls

Consider a family with two children. Given that the first child is a boy, what is the probability that both children are boys?

1/2

Law of total probability

Suppose we pick a random CMU junior.

Let Pr[student has taken 251]= 1/4

Also, let

- Prob. that a student understands conditional probability well given they took 251 = 9/10.
- Prob. student understands conditional prob. well given they didn’t take 251 = 1/2

What’s the probability that a random junior at CMU understands conditional probability well?

Sample space = CMU juniors

Event E = student has taken 251 (Pr[E] = ¼)

Event C = student understands conditional probability well.

- Pr[C | E] = 9/10
- Pr[C | $\overline{E}$] = ½

What is Pr[C] ?

Pr[C] = Pr[C ∩ E] + Pr [C ∩ $\overline{E}$]

= Pr [C | E] Pr [E] + Pr [C | $\overline{E}$] Pr [$\overline{E}$]

= 9/10 * ¼ + ½ * ¾ = 0.6

Law of total probability

$$\Pr[C] = \Pr[C \mid E]\Pr[E] + \Pr[C \mid \overline{E}]\Pr[\overline{E}]$$

More generally, let events $A_1, \ldots, A_n$ be a partition of the sample space, meaning each outcome is in exactly one.

Then for any event B,

$$\mathbf{Pr}[B] = \mathbf{Pr}[A_1] \cdot \mathbf{Pr}[B \mid A_1] + \cdots + \mathbf{Pr}[A_n] \cdot \mathbf{Pr}[B \mid A_n]$$

Example

"I roll 101 regular dice. What is the probability their sum is divisible by 6?"

Trick: "Condition on" the sum of the first 100.

Let A_k be event "the first 100 dice sum to k".

Then $A_{100}, \ldots, A_{600}$ partition the sample space.

Let B be event "sum of all 101 divisible by 6".

$\mathbf{Pr}[B \mid A_k] = 1/6$ for any k,

because conditioned on the first 100 summing to k, the final sum equally likely to be k+1, k+2, …, k+6; exactly one of these is div. by 6

So $\mathbf{Pr}[B] =$

$$\mathbf{Pr}[A_{100}]\mathbf{Pr}[B \mid A_{100}] + \cdots + \mathbf{Pr}[A_{600}]\mathbf{Pr}[B \mid A_{600}]$$

$$= \mathbf{Pr}[A_{100}]\,(1/6) + \cdots + \mathbf{Pr}[A_{600}]\,(1/6)$$

$$= (1/6)\,(\mathbf{Pr}[A_{100}] + \cdots + \mathbf{Pr}[A_{600}]) = 1/6.$$

a posteriori probabilities
Bayes Rule

A posteriori probability

A conditional probability Pr [A | B] is called *a posteriori* if event A precedes B in time

Probability that

- it was cloudy this morning,
  given it is raining in the afternoon
- I got a pair initially,
  given I ended up with a full house

Mathematically, *no different* from ordinary conditional probabilities …

An example (courtesy "*Chances Are*," NY Times, Apr 2010)

Before going on vacation, you ask your spacey friend to water your ailing plant.

- Without water, the plant has a 90% chance of dying.
- Even with proper watering, it has a 20% chance of dying.
- And the probability that your friend will forget to water it is 30%.

You return from vacation to find your plant dead ☹
What's the chance that your friend forgot to water it?

W = event that friend watered the plant

- W^c = the complement event

D = event that plant died

65.853…%

Given data:

$\Pr[D \mid W^c] = 0.9$ $\Pr[D \mid W] = 0.2$ $\Pr[W] = 0.7$

We want to know $\Pr[W^c \mid D]$ $(= 1 - \Pr[W \mid D])$

$\Pr[W \mid D] = \Pr[W \cap D] / \Pr[D] = \Pr[D \mid W] \Pr[W] / \Pr[D]$

$$= \frac{\Pr[D \mid W] \Pr[W]}{\Pr[D \mid W] \Pr[W] + \Pr[D \mid W^c] \Pr[W^c]}$$

Law of total prob.

$$= \frac{0.2 \times 0.7}{0.2 \times 0.7 + 0.9 \times 0.3} = \frac{14}{41}$$

Bayes rule

$$\Pr[A|B] = \frac{\Pr[B|A]\cdot\Pr[A]}{\Pr[B]} = \frac{\Pr[B|A]\cdot\Pr[A]}{\Pr[B|A]\cdot\Pr[A]+\Pr[B|\overline{A}]\cdot\Pr[\overline{A}]}$$

More generally, if events A_1, A_2, … A_n partition the sample space

$$\Pr[A_i|B] = \frac{\Pr[B|A_i]\cdot\Pr[A_i]}{\Pr[B]} = \frac{\Pr[B|A_i]\cdot\Pr[A_i]}{\sum_{j=1}^{n}\Pr[B|A_j]\cdot\Pr[A_j]}$$

Back to the World Series

The odds of A winning a game are 6:5

i.e., in any game, A wins with probability 6/11,
Independent of previous outcomes

What is the chance that A will beat B in the "best of 7" world series?

Team A beats B with probability 6/11 in each game

(implicit assumption: true for each game, independent of past.)

Sample space $\Omega = \{W, L\}^7$

$\Pr(x) = p^k(1-p)^{7-k}$ if there are k W's in x, where p=6/11

Want event E = "team A wins at least 4 games"

$E = \{x \in \Omega \mid x \text{ has at least 4 W's}\}$

$$\Pr[E] = \sum_{x \in E} \Pr[x] = \sum_{i=4}^{7} \binom{7}{i} \left(\frac{6}{11}\right)^{i} \left(\frac{5}{11}\right)^{7-i}$$

$= 0.5986\ldots$

Question:

Why is it permissible to assume that the two teams play a full seven-game series even if one team wins four games before seven have been played?

Independence

def: We say events A, B are independent if

$$\mathbf{Pr}[A \cap B] = \mathbf{Pr}[A]\ \mathbf{Pr}[B]$$

Except in the pointless case of **Pr**[A] or **Pr**[B] is 0,

equivalent to $\mathbf{Pr}[A \mid B] = \mathbf{Pr}[A]$,

or to $\mathbf{Pr}[B \mid A] = \mathbf{Pr}[B]$.

Two fair coins are flipped

A = {first coin is heads}

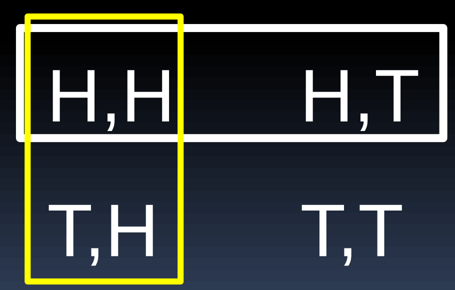
B = {second coin is heads}

Are A and B independent?

Pr[A] =

Pr[B] =

$\Pr[A \cap B]$ =



Two fair coins are flipped

A = {first coin is heads}

C = {two coins have different outcomes}

Are A and C independent?

Pr[A] =

Pr[C] =

Pr[A | C] =

Two fair coins are flipped

A = {first coin is heads}

$\overline{A}$ = {first coin is tails}

Are A and $\overline{A}$ independent?

H,H H,T

T,H T,T

The Secret "Principle of Independence"

Suppose you have an experiment with two parts (eg. two non-interacting blocks of code).

Suppose A is an event that only depends on the first part, B only on the second part.

Suppose you **prove** that the two parts *cannot* affect each other.
(E.g., equivalent to run them in opposite order.)

Then A and B are independent.

And you **may deduce** that **Pr**[A | B] = **Pr**[A].

Independence of Multiple Events

def: $A_1, \ldots, A_5$ are independent if

$$\mathbf{Pr}[A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5] = \mathbf{Pr}[A_1]\,\mathbf{Pr}[A_2]\,\mathbf{Pr}[A_3]\,\mathbf{Pr}[A_4]\,\mathbf{Pr}[A_5]$$

& $$\mathbf{Pr}[A_1 \cap A_2 \cap A_3 \cap A_4] = \mathbf{Pr}[A_1]\,\mathbf{Pr}[A_2]\,\mathbf{Pr}[A_3]\,\mathbf{Pr}[A_4]$$

& $$\mathbf{Pr}[A_1 \cap A_3 \cap A_5] = \mathbf{Pr}[A_1]\,\mathbf{Pr}[A_3]\,\mathbf{Pr}[A_5]$$

& in fact, the definition requires

$$\mathbf{Pr}\left[\bigcap_{i \in S} A_i\right] = \prod_{i \in S} \mathbf{Pr}[A_i] \quad \text{for all } S \subseteq \{1, 2, 3, 4, 5\}$$

Independence of Multiple Events

def: $A_1, \ldots, A_5$ are independent if

$$\mathbf{Pr}\left[\bigcap_{i \in S} A_i\right] = \prod_{i \in S} \mathbf{Pr}[A_i] \quad \text{for all } S \subseteq \{1, 2, 3, 4, 5\}$$

Similar 'Principle of Independence' holds

(5 blocks of code which don't affect each other)

Consequence: anything like

$$\mathbf{Pr}[A_1 \mid (A_2 \cup A_3) \cap (A_4^c \cup A_5)] = \mathbf{Pr}[A_1]$$

A little exercise

Can you give an example of a sample space and 3 events A_1, A_2, A_3 in it such that each pair of events A_i, A_j are independent, but A_1, A_2, A_3 together aren't independent?

Study Guide

Definitions:

Random experiments
Bernoulli, RandInt
sample space, outcome
event, probability
conditioning
Law of Total Probability
Bayes rule
independence

Solving problems:

how to find probabilities
how to condition
proving independence