15-251: Great Theoretical Ideas in Computer Science
Lecture 17

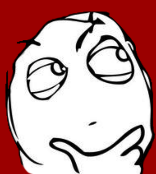
# Probability 1

## 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

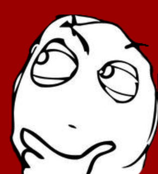
Antoine Gombaud,
AKA "Chevalier de Méré"

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”

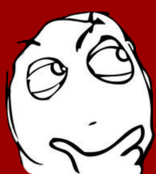
Hmm.

I keep losing money!

## 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?



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”

Probability Theory
=
Analyzing Code with
Random Number Generators

---

"Méré throws four 6-sided dice."

⇩

```
die1 = RandInt(6)
die2 = RandInt(6)
die3 = RandInt(6)
die4 = RandInt(6)
```

---

"Méré throws an 8-sided die
and a 3-sided die."

⇩

```
die1 = RandInt(8)
die2 = RandInt(3)
```

**def:** 'Experiment' = some randomized code

---

"A patient has a 10% chance
of having a certain disease..."

⇩

```
x = RandInt(10)
if x == 1 then
  patient.hasDisease = 1
else
  patient.hasDisease = 0
```

---

"A patient has a 10% chance
of having a certain disease..."

⇩

```
patient.hasDisease = Bernoulli(.1)
```

`Bernoulli(p)` returns 1 with probability p,
0 with probability 1−p

---

"Antoine flips two fair coins."

⇩

```
coin1 = Bernoulli(1/2)
coin2 = Bernoulli(1/2)
```

```
( if Bernoulli(1/2) == 0
      then coin1 = Heads
      else coin1 = Tails )
```

The two random generators we allow:

`RandInt(m)` returns 1, 2, 3, …, m with probability 1/m each

`Bernoulli(p)` returns 1 with probability p, 0 with probability 1−p

NOT ALLOWED IN 15-251:

~~`Uniform(0,1)` returns a random real number between 0 and 1~~

## How to Analyze Random Code

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.

**STEP 1**: Translate to code.

```
flip = Bernoulli(1/2)
if flip == 0 (Heads) then
  die = RandInt(3)
else
  die = RandInt(4)
```

## How to Analyze Random Code

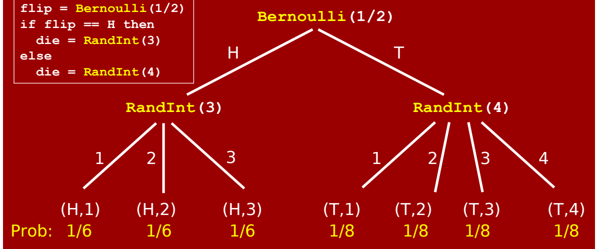
Mary flips a fair coin. If it's heads, she rolls a 3-sided die. If it's tails, she rolls a 4-sided die.

**STEP 2**: Draw a probability tree.

```
flip = Bernoulli(1/2)
if flip == 0 (Heads) then
  die = RandInt(3)
else
  die = RandInt(4)
```

## How to Analyze Random Code

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




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 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.

## How to Analyze Random Code

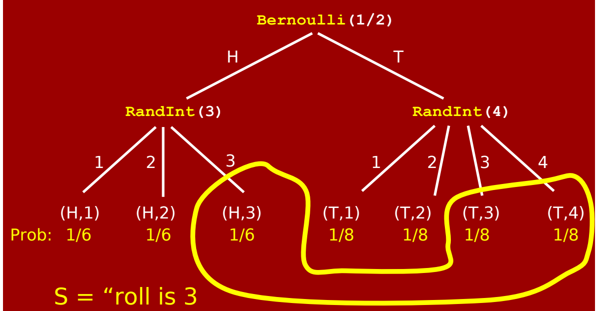
Mary flips a fair coin. If it's heads, she rolls a 3-sided die. If it's tails, she rolls a 4-sided die.

What is the probability die roll is 3 or higher?

**Event:**

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

**Pr**[S] = sum of the probabilities of the outcomes in S.




S = "roll is 3 or higher"

**Pr**[S] = 1/6 + 1/8 + 1/8 = 5/12

## 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?

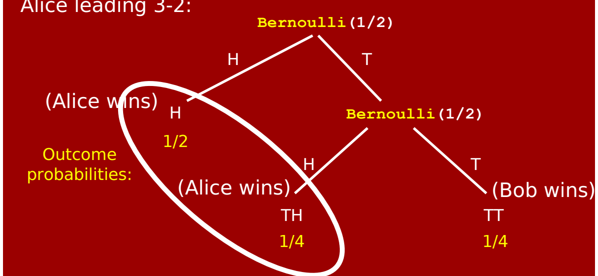
## France, 1654

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

## Events and Probabilities: Facts

Since **Pr**[A] = sum of probs of outcomes in A, …

If A ⊆ B then **Pr**[A] ≤ **Pr**[B]

*"not A"* **Pr**[$A^c$] = 1 − **Pr**[A]

*"A or B"* **Pr**[A ∪ B] = **Pr**[A] + **Pr**[B] − **Pr**[A ∩ B]

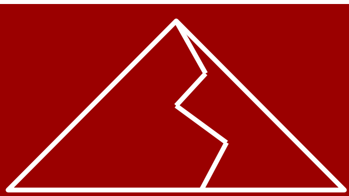
**FALLACY:** **Pr**[A ∪ B] = **Pr**[A] + **Pr**[B]

**True:** **Pr**[A ∪ B] ≤ **Pr**[A] + **Pr**[B]

## France, 1654

Let's bet:

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



(4,6,1,2) Prob. $1/6^4$

Let W be the event that Méré wins
Easier to compute **Pr**[$W^c$].
$W^c$ = { all outcomes with no 1's }
$|W^c| = 5^4$
∴ **Pr**[$W^c$] = $5^4/6^4$
∴ **Pr**[W] = $1-5^4/6^4 \approx 51.8\%$

## France, 1654

Let's bet:
I will roll two dice 24 times.
I win if I get a double-1's.

**Pr**[Méré wins] =

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

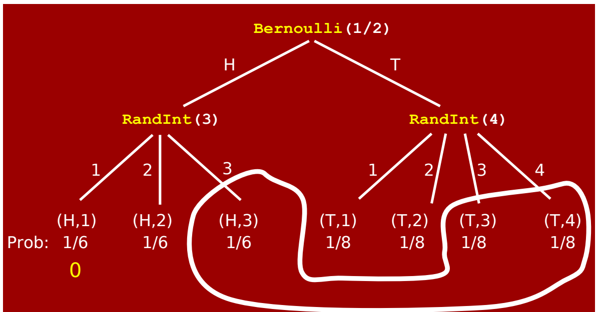
≈ 49.1%

## Conditioning

= Revising probabilities based on 'partial information'

'Partial information' = an event

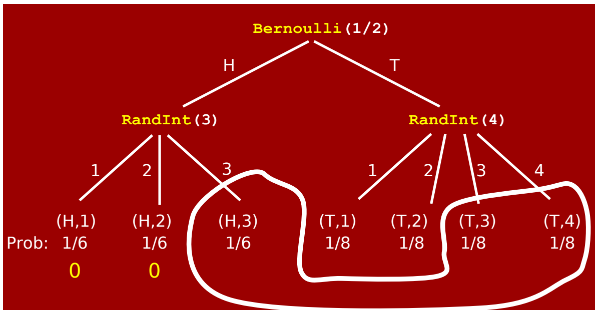
'Conditioning on event A' is like assuming/promising A occurs.




Condition on S, the event "roll is 3 or higher"

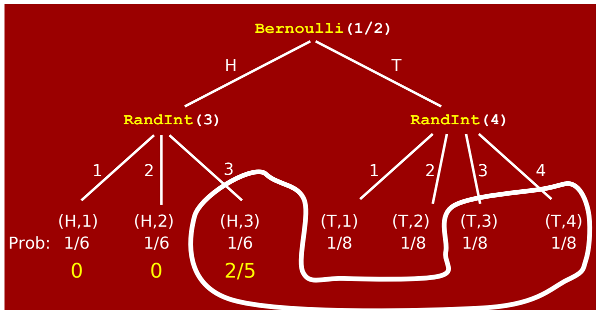
**Pr** [(H,1) | S] = 0

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



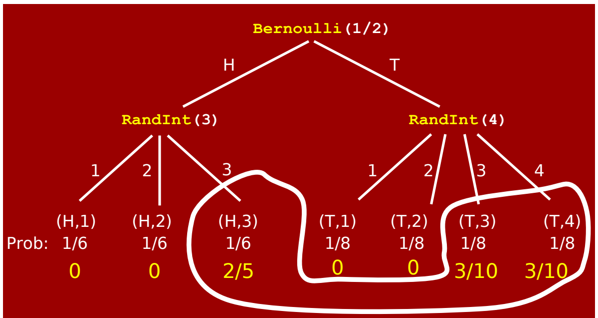

Condition on S, the event "roll is 3 or higher"

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



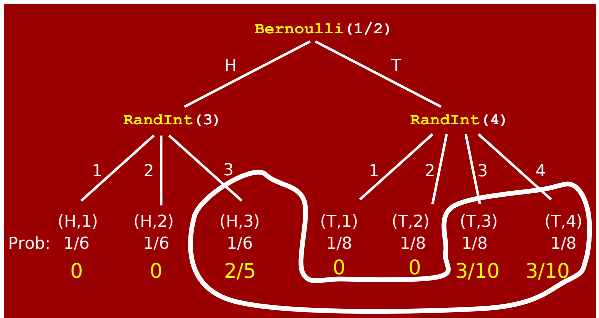

Condition on S, the event "roll is 3 or higher"

**Pr** [(H,3) | 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.

$\textbf{Pr}\,[A \mid S] =$ 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 $\ell$ is

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

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

# "Chain Rule"

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

*"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"

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

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

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

etc.

# Silver and Gold: a problem

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

Mark 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?

## Silver and Gold: a problem

Let $G_1$ be the event that the first chosen coin is gold.
Let $G_2$ be the event that the second coin in the bag is gold.
The problem is asking us to find $\mathbf{Pr}[G_2 \mid G_1]$.

$\mathbf{Pr}[G_1] = 3/6 = 1/2$ (each coin equally likely to be first)

$\mathbf{Pr}[G_1 \cap G_2] = 1/3$ (if and only if gold-gold bag picked)

$$\therefore \mathbf{Pr}[G_2 \mid G_1] = \frac{\mathbf{Pr}[G_2 \cap G_1]}{\mathbf{Pr}[G_1]} = \frac{1/3}{1/2} = 2/3$$

## Law of Total Probability

### or, how to actually calculate stuff

$$\mathbf{Pr}[B] = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B \mid A] + \mathbf{Pr}[A^c] \cdot \mathbf{Pr}[B \mid A^c]$$

"Regarding event B — either A occurs
(this has probability $\mathbf{Pr}[A]$),
in which case B occurs with probability $\mathbf{Pr}[B \mid A]$;

*or*, A does not occur
(this has probability $\mathbf{Pr}[A^c] = 1-\mathbf{Pr}[A]$),
in which case B occurs with probability $\mathbf{Pr}[B \mid A^c]$."

## Law of Total Probability

### or, how to actually calculate stuff

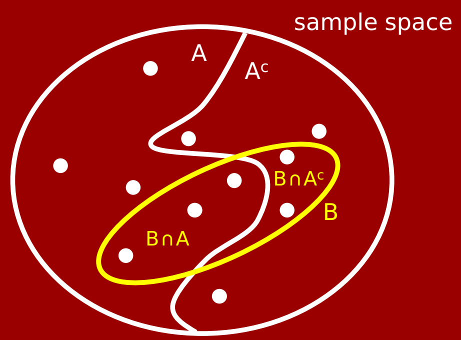
$$\mathbf{Pr}[B] = \mathbf{Pr}[A] \cdot \mathbf{Pr}[B \mid A] + \mathbf{Pr}[A^c] \cdot \mathbf{Pr}[B \mid A^c]$$

Proof:

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

Similarly, $\mathbf{Pr}[A^c] \cdot \mathbf{Pr}[B \mid A^c] = \mathbf{Pr}[B \cap A^c]$

Each outcome in B is in exactly one of $B \cap A$, $B \cap A^c$
Thus $\mathbf{Pr}[B] = \mathbf{Pr}[B \cap A] + \mathbf{Pr}[B \cap A^c]$




## Law of Total Probability

### more general version

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 $A_{100}$, …, $A_{600}$ partition the sample space.

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

**Pr**[B | $A_k$] = 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 **Pr**[B] =

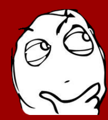
**Pr**[$A_{100}$]**Pr**[B | $A_{100}$] + · · · + **Pr**[$A_{600}$]**Pr**[B | $A_{600}$]

= **Pr**[$A_{100}$] (1/6) + · · · + **Pr**[$A_{600}$] (1/6)

= (1/6) (**Pr**[$A_{100}$] + · · · + **Pr**[$A_{600}$]) = 1/6.

## Trickier Problem

“I roll 101 regular dice. What is the probability their sum is divisible by **5**?”



Answer: $\frac{32665931175003545304834513357902891026857185523647743577153598318474857073869}{163329655875017726524172566789514455134285927618238717885767991592374285369344}$

$\approx .20000000000000000000000000000000000000000000000000000000000000000000000000001$

I solved this using linear algebra (Lecture 23…)

## Independence

def: We say events A, B are independent if
**Pr**[A∩B] = **Pr**[A] **Pr**[B]

Except in the pointless case of **Pr**[A] or **Pr**[B] is 0,
equivalent to **Pr**[A | B] = **Pr**[A],
or to **Pr**[B | A] = **Pr**[B].

## Independence Problem

Question:

I flip two coins. Let A be event “first flip is heads”, let B be event “even number of heads”. Are A and B independent?

Answer #1:

Yes! **Pr**[A] = 1/2, **Pr**[B] = 1/2,
**Pr**[A∩B] = **Pr**[(H,H)] = 1/4.

And (1/2)(1/2) = 1/4.

## Independence Problem

Question:

I flip two coins. Let A be event “first flip is heads”, let B be event “even number of heads”. Are A and B independent?

Answer #2:

Who cares? This is a pointless question.

You managed to calculate the 3 probabilities; who cares if two multiply to give the third?

## The Secret “Principle of Independence”

Suppose you have a block of randomized code with two parts.

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

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

## Birthday Problem

Question:

There are m students in a room ($m \le 365$).
What's the probability they
all have different birthdays?

Modeling:

Ignore Feb. 29. Assume days equally likely.
Assume no twins in the class.

```
for i = 1...m
    student[i].bday = RandInt(365)
```

## Birthday Problem — Analysis

Let $A_i$ be event that student i's bday differs from the bday of all previous students.

Let D be event that all bdays are different.

$$D = A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_m$$

Chain rule:

$\mathbf{Pr}[D] = \mathbf{Pr}[A_1]\,\mathbf{Pr}[A_2|A_1]\,\mathbf{Pr}[A_3|A_1 \cap A_2]\,\mathbf{Pr}[A_4| \cdots$ etc.]

So what is $\Pr[A_i \mid A_1 \cap A_2 \cap \cdots \cap A_{i-1}]$ ?

## Birthday Problem — Analysis

Let $A_i$ be event that student i's bday differs from the bday of all previous students.

So what is $\Pr[A_i \mid A_1 \cap A_2 \cap \cdots \cap A_{i-1}]$ ?

$A_1 \cap A_2 \cap \cdots \cap A_{i-1}$ means first $i-1$ students all had different birthdays.

$i-1$ out of 365 occupied when ith bday chosen.

$$\mathbf{Pr}[A_i \mid A_1 \cap A_2 \cap \cdots \cap A_{i-1}] = \frac{365-(i-1)}{365} = 1 - \frac{i-1}{365}$$

## Birthday Problem — Analysis

Let $A_i$ be event that student i's bday differs from the bday of all previous students.

Let D be event that all bdays are different.

$\mathbf{Pr}[D] = \mathbf{Pr}[A_1]\,\mathbf{Pr}[A_2|A_1]\,\mathbf{Pr}[A_3|A_1 \cap A_2]\,\mathbf{Pr}[A_4| \cdots$ etc.]

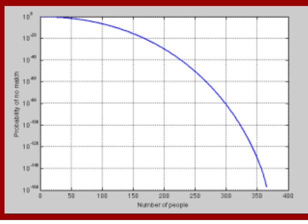
$$= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{m-1}{365}\right)$$

This is the final answer.

## Birthday Problem — Analysis

**Pr**[all m students have different bdays]

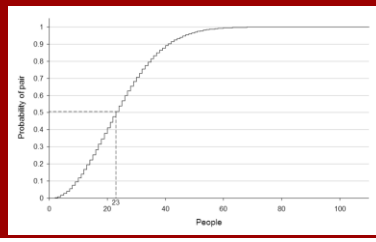
$$= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{m-1}{365}\right)$$



## Birthday Problem — Analysis

**Pr**[in m students, some pair share a bday]

$$= 1 - 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{m-1}{365}\right)$$




## Birthday Problem —

Sometimes called the Birthday “Paradox”, because 23 seems surprisingly small.

## Birthday Problem — Analysis

What if there are N possible “birthdays”?

**Pr**[in m students, some pair share a “bday”]

$$= 1 - 1 \cdot \left(1 - \frac{1}{N}\right) \cdot \left(1 - \frac{2}{N}\right) \cdots \left(1 - \frac{m-1}{N}\right)$$

For what value of m is this ≈ 1/2 ?

This is not a calculus class, so I’ll just tell you:

for $m \approx \sqrt{N}$

## Birthday Problem —

Sometimes called the Birthday “Paradox”, because 23 seems surprisingly small.

Sometimes called the Birthday “Attack” in theoretical cryptography. Why…?

## Cryptographic Hash Functions

“Scrambles” any string S into a k-bit ‘hash’ f(S)

- Given f(S), should be ‘hard’ to recover S.
- Should be ‘hard’ to find a “collision”: two strings $S_1 \neq S_2$ with $f(S_1) = f(S_2)$.

Applications: authentication schemes, data integrity schemes, digital signatures, e-cash…

## Cryptographic Hash Functions

1991: Rivest publishes MD5. (k=128)

1993: NSA publishes SHA-0. (k=160)

1995: NSA: “Um, never mind.
Please use SHA-1 instead.”
SHA-1 was/is widely used: SSL, PGP, …

2001: NSA also introduces SHA-2
(variants with k=224,256,384,512)

2012: Non-NSA introduces SHA-3

## Birthday Attack

Imagine trying to find a collision for SHA-1:
Take a huge number of strings, hash them all, hope that two hash to the same 160 bits.

If SHA-1 is really safe, each hash f(S) should be like RandInt($2^{160}$).

This is like the Birthday Problem with N = $2^{160}$ !

So # tries before good chance of collision:

$$\approx \sqrt{2^{160}} = 2^{80} = 1208925819614629174706176$$

## Birthday Attack

Everybody knows this.
$2^{80}$ is considered safely “too large”.

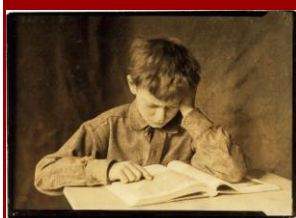
A crypto hash function is considered “broken” if you can beat the Birthday Attack.

Xiaoyun Wang (王小云)

2005: SHA-1 collisions in $2^{69}$
Later (w/ coauthors): in $2^{63}$

SHA-1 = broken
(phased out of SSL by 2017)

## Study Guide



Definitions:
RandInt, Bernoulli experiment
sample space, outcome
event, probability
conditioning
Law of Total Probability
independence

Solving problems:
how to find probabilities
how to condition
proving independence