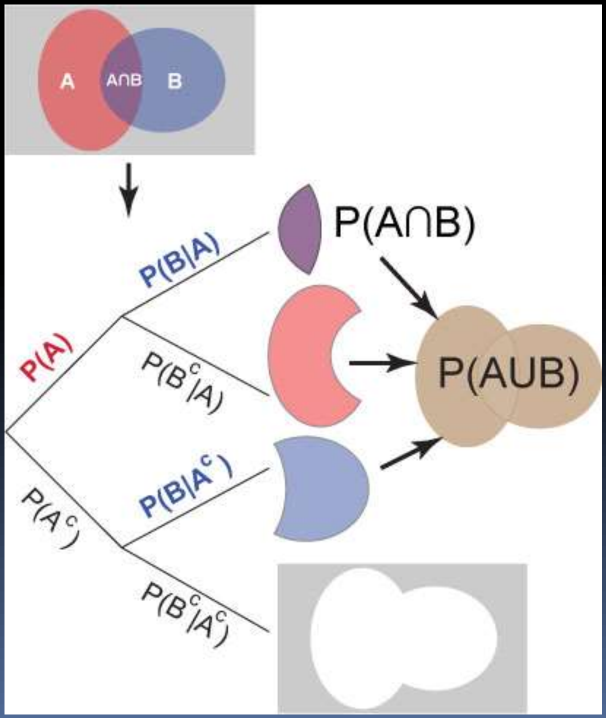
15-251: Great Theoretical Ideas in Computer Science

Fall 2016 Lecture 18

October 27, 2016

Probability 2:

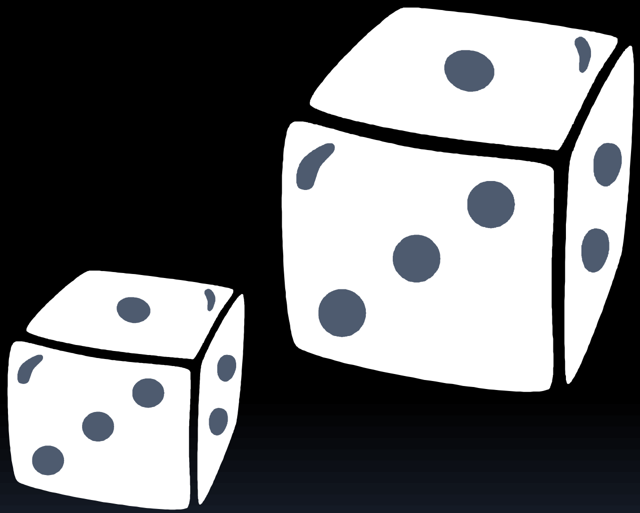
Random variables and Expectations



E[X+Y] = E[X] + E[Y]

Review

Some useful sample spaces…



1) A fair coin

sample space $\Omega = \{H, T\}$

$Pr[H] = ½$, $Pr[T] = ½$.

2) A "bias-p" coin

sample space $\Omega = \{H, T\}$

$Pr[H] = p$, $Pr[T] = 1-p$.

3) Two independent bias-p coin tosses

sample space Ω = {HH, HT, TH, TT}

x	$\Pr[x]$
$\langle T,T \rangle$	$(1-p)^2$
$\langle T,H \rangle$	$(1-p)p$
$\langle H,T \rangle$	$(1-p)p$
$\langle H,H \rangle$	p^2

3) n bias-p coins

sample space $\Omega = \{H,T\}^n$

If outcome x in Ω has k heads and n-k tails

$$\Pr[x] = p^k (1-p)^{n-k}$$

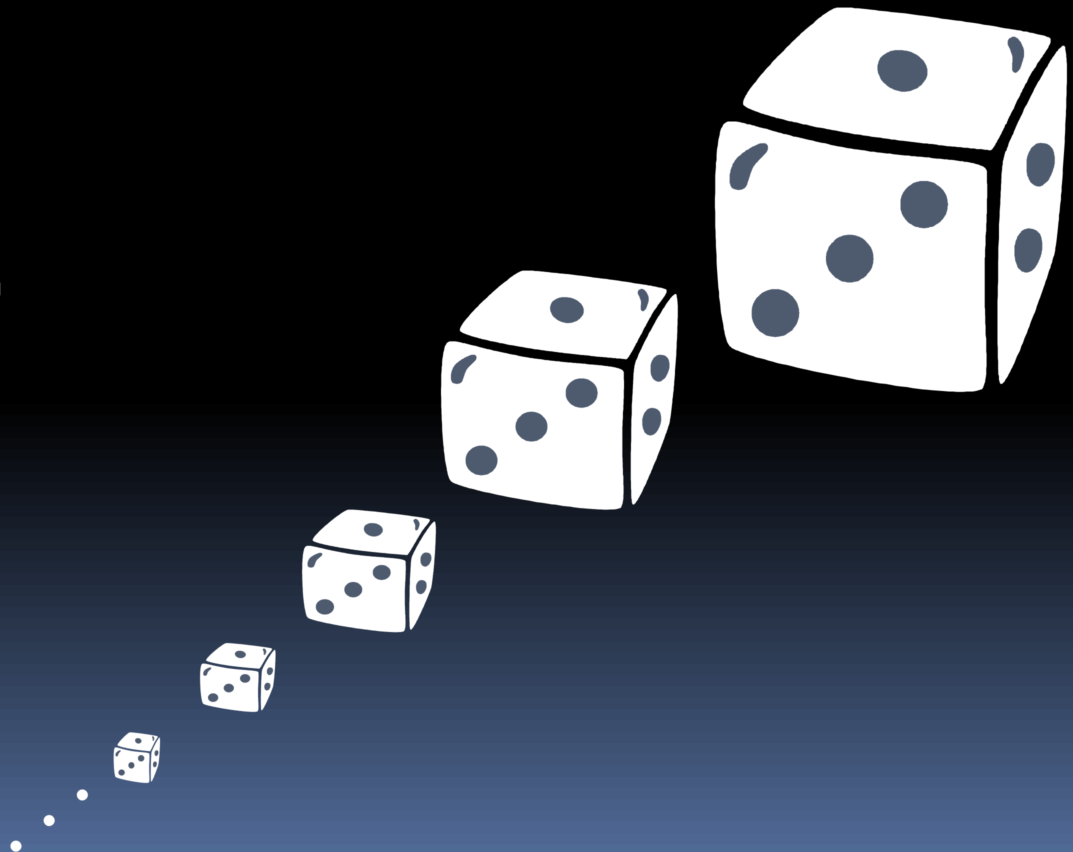
Event $E_k = \{x \in \Omega \mid x \text{ has k heads}\}$

$$\Pr[E_k] = \sum_{x \in E_k} \Pr[x] = \binom{n}{k} p^k (1-p)^{n-k}$$

"Binomial Distribution B(n,p) on {0,1,2,…,n}"

$$\Pr[k] = \binom{n}{k} p^k (1-p)^{n-k}$$

An Infinite sample space…



The "Geometric" Distribution

A bias-p coin is tossed until the first time that a head turns up.

sample space Ω = {H, TH, TTH, TTTH, …}

(shorthand $\Omega = \{1, 2, 3, 4, \dots\}$)

$\Pr_{Geom}[k] = (1-p)^{k-1}\, p$

(sanity check) $\sum_{k\geq 1} \Pr[k] = \sum_{k\geq 1} (1-p)^{k-1}\, p$

$= p * (1 + (1-p) + (1-p)^2 + \dots)$

$= p * 1/(1-(1-p)) = 1$

Independence of Events

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}

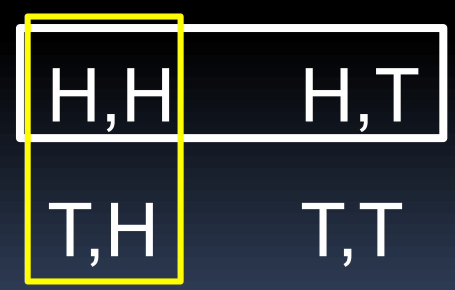
B = {second coin is heads}

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?

Feature Presentation: Random Variables

Random Variable

Let Ω be sample space in a probability distribution

A Random Variable is a function from Ω to reals

Examples:

F = value of first die in a two-dice roll

F(3,4) = 3, **F**(1,6) = 1

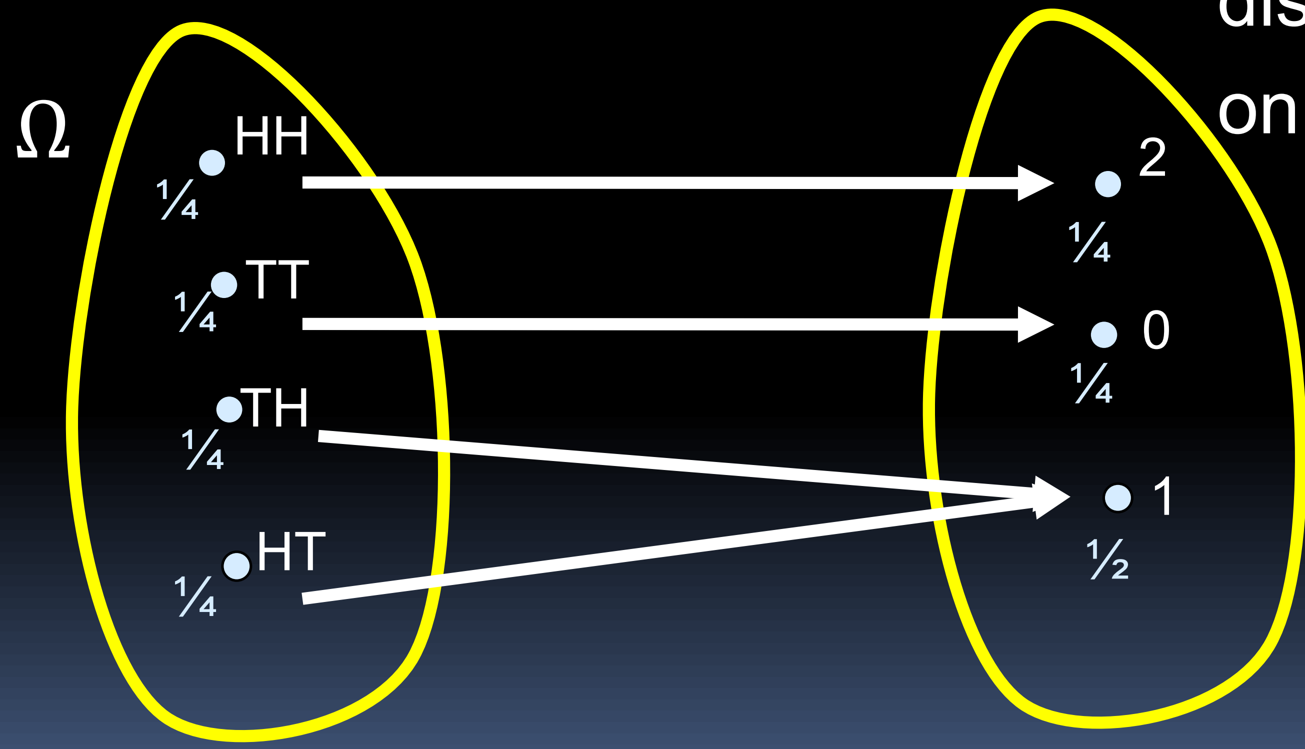
X = sum of values of the two dice

X(3,4) = 7, **X**(1,6) = 7

Two Coins Tossed

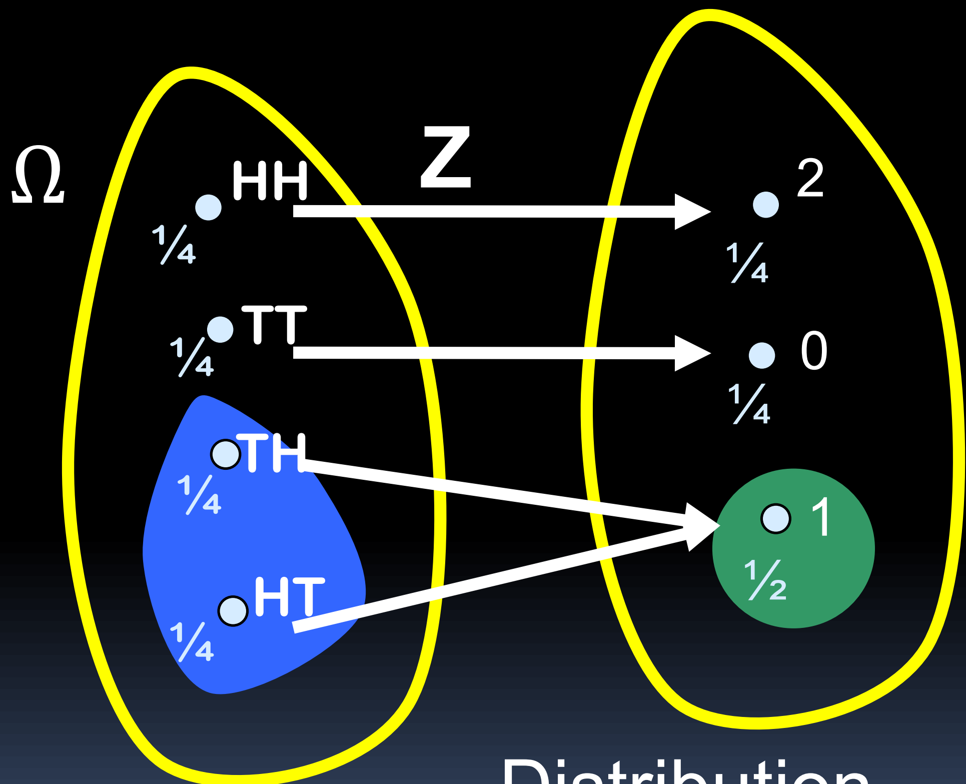
Z: {TT, TH, HT, HH} → {0, 1, 2} counts the number of heads

Induces distribution on {0,1,2}



Two Coins Tossed

Z: {TT, TH, HT, HH} → {0, 1, 2} counts # of heads



Distribution of **Z**

Pr[**Z**= a] =
Pr[{t ∈ Ω | **Z**(t) = a}]

Pr[**Z** = 1]

= Pr[{t ∈ Ω | **Z**(t) = 1}]

= Pr[{TH, HT}] = ½

Two Views of Random Variables

Input to the function is random

Think of a R.V. as

A function from sample space to the reals R

Or think of the induced distribution on R

Randomness is “pushed” to the values of the function

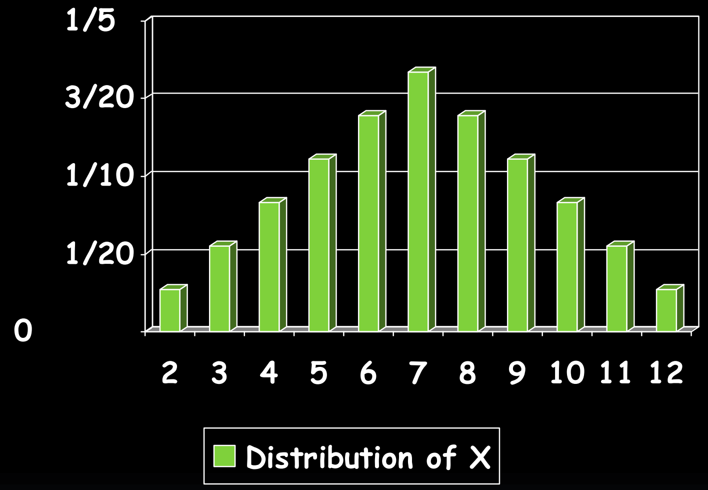
Given a distribution on some sample space Ω, a random variable transforms it into a distribution on reals

Two dice

I throw a white die and a black die. **X** = sum of both dice

Sample space =

{ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6),
(2,1), (2,2), (2,3), (2,4), (2,5), (2,6),
(3,1), (3,2), (3,3), (3,4), (3,5), (3,6),
(4,1), (4,2), (4,3), (4,4), (4,5), (4,6),
(5,1), (5,2), (5,3), (5,4), (5,5), (5,6),
(6,1), (6,2), (6,3), (6,4), (6,5), (6,6) }



function with **X**(1,1) = 2, **X**(1,2) = 3, …, **X**(6,6)=12

Random variables: two viewpoints

It is a function on the sample space

It is a variable with a probability distribution on its values

You should be comfortable with both views

Random Variables: introducing them

Retroactively:

"Let **D** be the random variable given by subtracting the first roll from the second."

D((1,1)) = 0, …, **D**((5, 3)) = −2, etc.

Random Variables: introducing them

In terms of other random variables:

“Let **Y** = $\mathbf{X}^2$+**D**.” ⇒ **Y**((5,3)) = 62

“Suppose you win $30 on a roll of double-6, and you lose $1 otherwise. Let **W** be the random variable representing your winnings.”

$$\mathbf{W} = 30 \cdot \mathbf{I} + (-1)\,(1-\mathbf{I}) = 31 \cdot \mathbf{I} - 1$$

Where **I**((6,6))=1 and **I**((x,y))=0 otherwise

Random Variables: introducing them

By describing its distribution:

"Let **X** be a Bernoulli(1/3) random variable."

- Means Pr[**X**=1]=1/3, Pr[**X**=0]=2/3

"Let **Y** be a Binomial(100,1/3) random variable."

"Let **T** be a random variable which is uniformly distributed (= each value equal probability) on the set {0,2,4,6,8}."

Random Variables to Events

E.g.: **S** = sum of two dice

"Let A be the event that **S** ≥ 10."

A = { (4,6), (5,5), (5,6), (6,4), (6,5), (6,6) }

Pr[**S** ≥ 10] = 6/36 = 1/6

Shorthand notation for (of **S** ≥ 10)

the **event** { ℓ : **S**(ℓ) ≥ 10 }.

Events to Random Variables

Definition:

Let A be an event. The indicator of A is the random variable **X** which is 1 when A occurs and 0 when A doesn't occur.

$$\mathbf{X} : \Omega \to \mathbb{R} \qquad \mathbf{X}(\ell) = \begin{cases} 1 & \text{if } \ell \in A \\ 0 & \text{if } \ell \notin A \end{cases}$$

Notational Conventions

Use letters like A, B, C for events

Use letters like **X, Y, f, g** for R.V.'s

R.V. = random variable

Independence of Random Variables

Definition:

Random variables **X** and **Y** are independent if the events "X = u" and "Y = v" are independent for all u,v∈ℝ.

(And similarly for more than 2 random variables.)

Random variables X_1, X_2, …, X_n are independent if *for all* reals a_1, a_2, …, a_n

$$\Pr(X_1 = a_1 \cap X_2 = a_2 \cap \cdots \cap X_n = a_n) = \prod_{i=1}^{n} \Pr(X_i = a_i)$$

Examples: Independence of r.v's

Two random variables **X** and **Y** are said to be independent if *for all* reals a, b,

$$\Pr[\mathbf{X} = a \cap \mathbf{Y} = b] = \Pr[\mathbf{X}=a] \Pr[\mathbf{Y}=b]$$

A coin is tossed twice.

$\mathbf{X}_i$ = 1 if the i[th] toss is heads and 0 otherwise.

Are $\mathbf{X}_1$ and $\mathbf{X}_2$ independent R.Vs ?

Yes.

Let $\mathbf{Y} = \mathbf{X}_1 + \mathbf{X}_2$. Are $\mathbf{X}_1$ and **Y** independent?

No.

Expectation

aka Expected Value

aka Mean

Expectation

Intuitively, expectation of **X** is what its average value would be if you ran the experiment millions and millions of times.

Definition:

Let **X** be a random variable in experiment with sample space Ω. Its expectation is:

$$\mathbf{E}[\mathbf{X}] = \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{X}(\ell)$$

Expectation — examples

Let **R** be the roll of a standard die.

$$\mathbf{E}[\mathbf{R}] = \frac{1}{6}\cdot 1 + \frac{1}{6}\cdot 2 + \frac{1}{6}\cdot 3 + \frac{1}{6}\cdot 4 + \frac{1}{6}\cdot 5 + \frac{1}{6}\cdot 6$$

$$= 3.5$$

Question: What is **Pr**[**R** = 3.5]?

Answer: 0. Don't always expect the expected!

Expectation — examples

"Suppose you win \$30 on a roll of double-6, and you lose \$1 otherwise. Let **W** be the random variable representing your winnings."

$$\mathbf{E}[\mathbf{W}] = \frac{1}{36}\cdot(-1) + \frac{1}{36}\cdot(-1) + \cdots + \frac{1}{36}\cdot(-1) + \frac{1}{36}\cdot 30$$

$$= -5/36 \approx -13.9¢$$

Expectation — examples

Let $\mathbf{R}_1$ = Throw of die 1, $\mathbf{R}_2$ = Throw of die 2

$\mathbf{S} = \mathbf{R}_1 + \mathbf{R}_2$.

$$\mathbf{E}[\mathbf{S}] = \frac{1}{36} \cdot (1+1) + \frac{1}{36} \cdot (1+2) + \cdots + \frac{1}{36} \cdot (6+6)$$

= lots of arithmetic ☹

= 7 (eventually)

One of the top tricks in probability...

Linearity of Expectation

Given an experiment,

let **X** and **Y** be any random variables.

Then **E**[**X**+**Y**] = **E**[**X**] + **E**[**Y**]

X and **Y** do ***not*** have to be independent!!

Linearity of Expectation

E[X+Y] = E[X] + E[Y]

Proof: Let **Z** = **X+Y** (another random variable).

Then $$\mathbf{E}[\mathbf{Z}] = \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{Z}(\ell)$$

$$= \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot (\mathbf{X}(\ell) + \mathbf{Y}(\ell))$$

$$= \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{X}(\ell) \quad + \quad \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{Y}(\ell)$$

$$= \mathbf{E}[\mathbf{X}] \quad + \quad \mathbf{E}[\mathbf{Y}]$$

Linearity of Expectation

$$\mathbf{E}[\mathbf{X}+\mathbf{Y}] = \mathbf{E}[\mathbf{X}] + \mathbf{E}[\mathbf{Y}]$$

Also:

$$\mathbf{E}[a\mathbf{X}+b] = a\mathbf{E}[\mathbf{X}]+b \quad \text{for any } a,b\in\mathbb{R}.$$

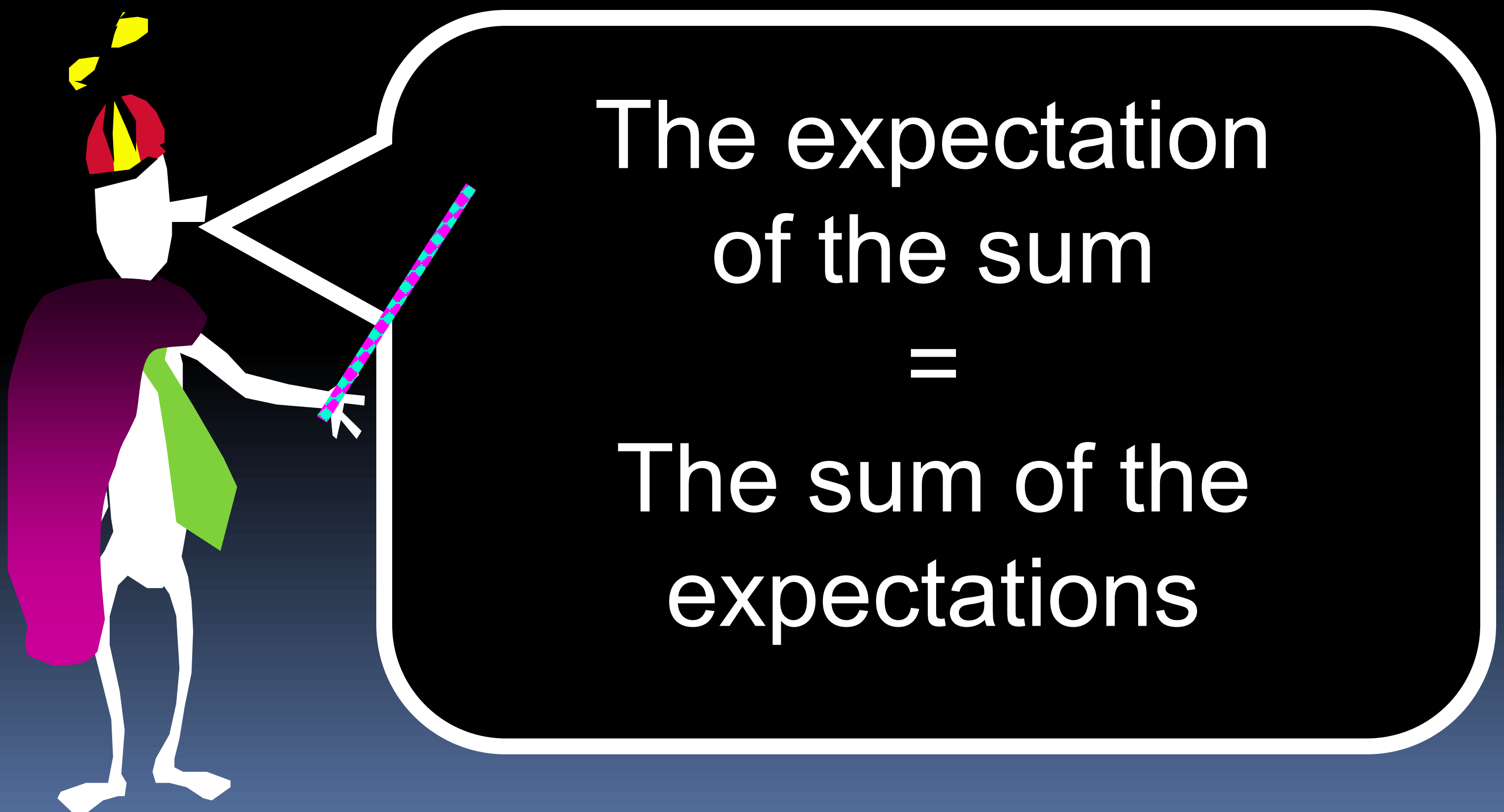
By Induction

$$\mathbf{E}[\mathbf{X}_1 + \cdots + \mathbf{X}_n] = \mathbf{E}[\mathbf{X}_1] + \cdots + \mathbf{E}[\mathbf{X}_n]$$

Remember…

$$E[X_1 + X_2 + \ldots + X_n] =$$

$$E[X_1] + E[X_2] + \ldots. + E[X_n],$$ ***always***



Linearity of Expectation example

Let $\mathbf{R}_1$ = Throw of die 1, $\mathbf{R}_2$ = Throw of die 2

$\mathbf{S} = \mathbf{R}_1 + \mathbf{R}_2$.

$$\mathbf{E}[\mathbf{S}] = \mathbf{E}[\mathbf{R}_1] + \mathbf{E}[\mathbf{R}_2]$$

$$= 3.5 + 3.5$$

$$= 7$$

Expectation of an Indicator

Fact: Let A be an event, let **X** be its indicator r.v.

Then **E**[**X**] = **Pr**[A].

Proof:

$$
\begin{aligned}
\mathbf{E}[\mathbf{X}] &= \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{X}(\ell) \\
&= \sum_{\ell \in A} \mathbf{Pr}[\ell] \cdot 1 + \sum_{\ell \notin A} \mathbf{Pr}[\ell] \cdot 0 \\
&= \sum_{\ell \in A} \mathbf{Pr}[\ell] \\
&= \mathbf{Pr}[A]
\end{aligned}
$$

Linearity of Expectation

+

Indicators

= best friends forever

Linearity of Expectation + Indicators

There are 251 students in a class.

The TAs randomly permute their midterms before handing them back.

Let **X** be the number of students getting their own midterm back.

What is **E**[**X**]?

Let's try 3 students first

Midterm they got	Student 1	Student 2	Student 3	Prob	# getting own midterm
	1	2	3	1/6	3
	1	3	2	1/6	1
	2	1	3	1/6	1
	2	3	1	1/6	0
	3	1	2	1/6	0
	3	2	1	1/6	1

$\therefore$ **E[X]** $= (1/6)(3+1+1+0+0+1) = 1$

Now let's do 251 students

		Um…		

Now let's do 251 students

Let A_i be the event that i^{th} students gets own midterm.

Let $\mathbf{X}_i$ be the indicator of A_i.

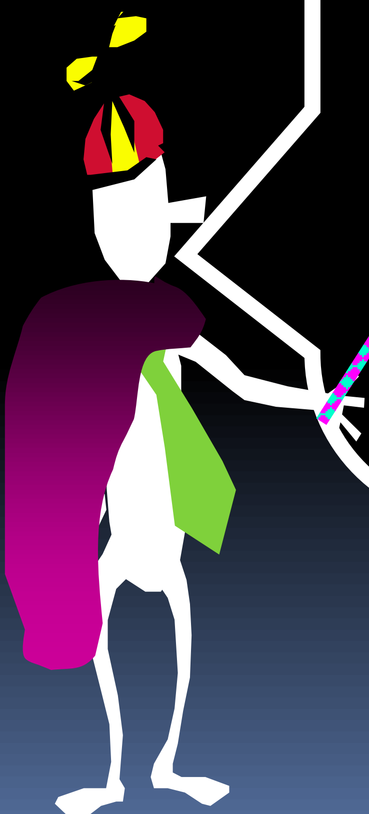
Then $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$

Thus $\mathbf{E}[\mathbf{X}] = \mathbf{E}[\mathbf{X}_1] + \mathbf{E}[\mathbf{X}_2] + \cdots + \mathbf{E}[\mathbf{X}_n]$

by linearity of expectation

$\mathbf{E}[\mathbf{X}_i] = \mathbf{Pr}[A_i]$, and $\mathbf{Pr}[A_i] = 1/251$ for each i.

$\therefore \mathbf{E}[\mathbf{X}] = 251 \cdot (1/251) = 1$



So, in expectation, 1 student will receive his/her midterm.

Pretty neat: it doesn't depend on how many students!

Question: were the X_i independent?

No! E.g., think of n=2

Another Formula for Expectation

For a r.v **X** over sample space Ω :

$$\mathbf{E}[\mathbf{X}] = \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{X}(\ell)$$

Also:

$$\mathbf{E}[\mathbf{X}] = \sum_{u \in \text{range}(\mathbf{X})} \mathbf{Pr}[\mathbf{X} = u] \cdot u$$

Remarks:

- range(**X**) = the set of real numbers **X** may take on
- "**X** = u" is an event
- some people (not us) take this as the *definition*

Expectation in two ways

$$E[\mathbf{X}] = \sum_{t \in \Omega} \Pr(t)\, \mathbf{X}(t) = \sum_{u} u \Pr[\mathbf{X} = u]$$

X is a function on the sample space

X has an associated prob. distribution on its values

(assuming **X** takes discrete values)

$$\mathbf{E}[\mathbf{X}] = \sum_{u \in \text{range}(\mathbf{X})} \mathbf{Pr}[\mathbf{X} = u] \cdot u$$

Proof by "counting two ways":

$$\begin{aligned}
\mathbf{E}[\mathbf{X}] &= \sum_{\ell \in \Omega} \mathbf{Pr}[\ell] \cdot \mathbf{X}(\ell) \\
&= \sum_{u \in \text{range}(\mathbf{X})} \sum_{\ell : \mathbf{X}(\ell) = u} \mathbf{Pr}[\ell] \cdot \mathbf{X}(\ell) \\
&= \sum_{u \in \text{range}(\mathbf{X})} \sum_{\ell : \mathbf{X}(\ell) = u} \mathbf{Pr}[\ell] \cdot u \\
&= \sum_{u \in \text{range}(\mathbf{X})} u \cdot \sum_{\ell : \mathbf{X}(\ell) = u} \mathbf{Pr}[\ell] \\
&= \sum_{u \in \text{range}(\mathbf{X})} u \cdot \mathbf{Pr}[\mathbf{X} = u]
\end{aligned}$$

Example

Question: Let **X** be a uniformly random integer between 1 and 10. Let **Y** = **X** mod 3.

What is **E[Y]**?

Poll

range(**Y**) = {0,1,2}

$\mathbf{E}[\mathbf{Y}] = \mathbf{Pr}[\mathbf{Y} = 0] \cdot 0 + \mathbf{Pr}[\mathbf{Y} = 1] \cdot 1 + \mathbf{Pr}[\mathbf{Y} = 2] \cdot 2$

$\mathbf{E}[\mathbf{Y}] = \mathbf{Pr}[\mathbf{Y} = 1] + 2\,\mathbf{Pr}[\mathbf{Y} = 2]$

$\mathbf{E}[\mathbf{Y}] = \mathbf{Pr}[\{1,4,7,10\}] + 2\,\mathbf{Pr}[\{2,5,8\}]$

$\mathbf{E}[\mathbf{Y}] = 4/10 + 2(3/10) = 1$

Example

Question: Let **X** be a uniformly random integer between 1 and 10. Let **Y** = **X** mod 3.

What is **E[Y]**?

range(**Y**) = {0,1,2}

E[**Y**] = **Pr**[**Y** = 0] · 0 + **Pr**[**Y** = 1] · 1 + **Pr**[**Y** = 2] · 2

Note: We didn't really care how **Y** was created.

We only needed **Pr[Y=u]** for each u ∈ range(**Y**).

If I return 251 randomly permuted midterms to 251 students, on average how many students get their back their own midterm?

Hmm…

$\sum_k k \cdot$ Pr[exactly k letters end up in correct envelopes]

= $\sum_k k \cdot$ (…aargh!!…)

Thank you,

Linearity of Expectation!

Type Checking

Pr[B] — B must be an event

E [X] — X must be a R.V.

cannot do Pr[R.V.] or E[event]

Operations on R.V.s

You can sum them, take differences,
or do most other math operations
(they are just functions!)

E.g., $(\mathbf{X} + \mathbf{Y})(t) = \mathbf{X}(t) + \mathbf{Y}(t)$

$(\mathbf{X} * \mathbf{Y})(t) = \mathbf{X}(t) * \mathbf{Y}(t)$

$(\mathbf{X}^{\mathbf{Y}})(t) = \mathbf{X}(t)^{\mathbf{Y}(t)}$

Expectation of a Sum of r.v.s

= Sum of their Expectations

even when r.v.s are not independent!

Expectation of a Product of r.v.s

vs. Product of their Expectations ?

Multiplication of Expectations

A coin is tossed twice.

$\mathbf{X}_i$ = 1 if the i^{th} toss is heads and 0 otherwise.

$E[\mathbf{X}_1] = E[\mathbf{X}_2] = 1/2$

$E[\mathbf{X}_1 \mathbf{X}_2] = 1/4$

$E[\mathbf{X}_1] E[\mathbf{X}_2] = 1/4$

<u>Lemma</u>: E[**XY**] = E[**X**] E[**Y**] if **X** and **Y** are *independent* random variables.

(And similar statement for > 2 r.v's)

Proof left as exercise.

Multiplication of Expectations

Consider a single toss of a coin.

X = 1 if heads turns up and 0 otherwise.

Set **Y** = 1 - **X**

E[**X**] = E[**Y**] = 1/2

X and **Y** are not independent

E[**X Y**] $\neq$ E[**X**] E[**Y**]

since **X Y** = 0 with probability 1

More examples of Computing Expectations

We flip n coins of bias p. What is the expected number of heads **H**?

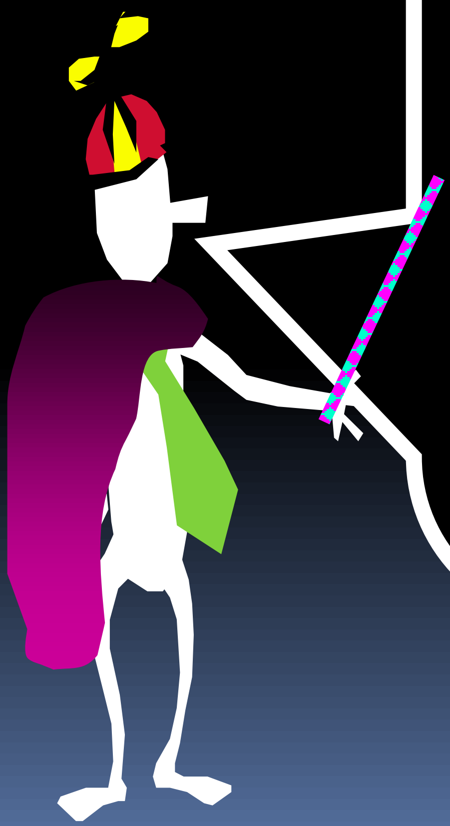
We could do this by summing

$$\sum_k k \Pr(H = k) = \sum_k k \binom{n}{k} p^k (1-p)^{n-k}$$

$$= np$$

But we know a better way!

Use Linearity of Expectation



General approach:

View thing you care about as expected value of some RV

Write this RV as sum of simpler RVs (often indicator RVs)

Solve for their expectations and add them up!

Back to example:

Let **H** = number of heads when n independent coins of bias p are flipped

Break **H** into n simpler RVs:

$$\mathbf{H_i} = \begin{cases} 1 & \text{if the } i^{th} \text{ coin is heads} \\ 0 & \text{if the } i^{th} \text{ coin is tails} \end{cases}$$

$E[\, \mathbf{H_i} \,] = p$

Note $\mathbf{H} = \sum_i \mathbf{H_i}$

$E[\, \mathbf{H} \,] = E[\, \sum_i \mathbf{H_i} \,] = \sum_i E[\, \mathbf{H_i} \,] = np$

Geometric Random Variables

X ~ Geometric(p)

What is **E[X]**?

Average number of p-biased coin flips until you get Heads: you might guess 1/p.

Proof: Direct calculation

E[X] = $\sum_{k\geq 1} k \cdot \Pr[\mathbf{X} = k] = \sum_{k\geq 1} k\, p\, (1-p)^{k-1}$

$= p\ \sum_{k\geq 1} k\, (1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$

An approach: Generating Functions

The Coupon Collector

There are n different kinds of coupons.

On each day, you get a random coupon.

(You may get duplicates.)

Let **X** be the # of days till you have them all.

What is **E[X]**?

The Coupon Collector

Let **X** be the # of days till you have them all.

What is **E[X]**?

Key idea: Let $\mathbf{X}_i$ be # of days it took you to go from i−1 to i coupons.

Key idea: $\mathbf{X} = \mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n$

$$\therefore \mathbf{E}[\mathbf{X}] = \mathbf{E}[\mathbf{X}_1] + \mathbf{E}[\mathbf{X}_2] + \cdots + \mathbf{E}[\mathbf{X}_n]$$

So we need to figure out $\mathbf{E}[\mathbf{X}_i]$.

The Coupon Collector

Key idea: Let X_i be # of days it took you to go from $i-1$ to i coupons.

When sitting on $i-1$ distinct coupons, each day you have probability $\frac{n-(i-1)}{n}$ of getting a new one.

$\therefore X_i \sim \text{Geometric}\left(\frac{n-(i-1)}{n}\right)$ $\therefore E[X_i] = \frac{n}{n-(i-1)}$

for example,

$$E[X_1] = \frac{n}{n} = 1, \quad E[X_2] = \frac{n}{n-1}, \quad \cdots, \quad E[X_n] = \frac{n}{1} = n$$

The Coupon Collector

$\therefore$ **E[X]** = **E[X_1]** + **E[X_2]** + ⋯ + **E[X_n]**

$$= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1}$$

$$= n\left(\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n}\right)$$

$\therefore$ **E[X]** = $n \cdot H_n$ $\therefore$ **E[X]** $\approx n \ln n$

where H_n = "the nth harmonic number"

$$= \frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{n} \approx \ln n$$

Using linearity of expectations in unexpected places…

10% of the surface of a sphere is colored green, and the rest is colored blue. Show that now matter how the colors are arranged, it is possible to inscribe a cube in the sphere so that all of its vertices are blue.

Solution

Pick a random cube. (Note: any particular vertex is uniformly distributed over surface of sphere).

Let $\mathbf{X_i}$ = 1 if i^{th} vertex is blue, 0 otherwise (indicator r.v.)

Let $\mathbf{X} = \mathbf{X_1} + \mathbf{X_2} + \ldots + \mathbf{X_8}$

$$E[\mathbf{X_i}] = Pr[\mathbf{X_i}=1] = \frac{9}{10}$$

$$E[\mathbf{X}] = 8 \cdot \frac{9}{10} > 7$$

So, must have some cubes where $\mathbf{X} = 8$!!

The general principle we used in this example:

Show the expected value of some random variable is "high"

Hence, there must be an outcome in the sample space where the random variable takes on a "high" value.

(Not everyone can be below the average.)

called **"the probabilistic method"**

(a ***very*** powerful & important tool)

Conditional expectations

Just like probabilities, we can also talk about expectations *conditioned on some event.*

E[**X** | A] = expectation of **X** conditioned on event A

It's just the expectation according to the conditional distribution!

$$E[\mathbf{X} \mid A] = \sum_{t \in A} \mathbf{X}(t) \frac{\Pr[t]}{\Pr[A]} = \sum_{k \in \text{range}(X)} k \Pr[\mathbf{X} = k \mid A]$$

<u>Law of total expectation</u>:

$$E[\mathbf{X}] = \Pr[A]\, E[\mathbf{X} \mid A] + \Pr[\overline{A}]\, E[\mathbf{X} \mid \overline{A}]$$

More generally, if $A_1, A_2, \ldots, A_n$ partition the sample space

$$E[\mathbf{X}] = E[\mathbf{X}|A_1]\Pr[A_1] + E[\mathbf{X}|A_2]\Pr[A_2] + \cdots + E[\mathbf{X}|A_n]\Pr[A_n]$$

Simple example: Law of total expectation

49.8% of population male

Average height: 5’11” (men) 5’5” (female)

What’s the average height of the whole population?

$$\mathrm{E}[\mathbf{H}] = \mathrm{E}[\mathbf{H}\,|\mathrm{M}]\,\Pr[\mathrm{M}] + \mathrm{E}[\mathbf{H}\,|\overline{M}]\,\Pr[\overline{M}]$$
$$= 5\frac{11}{12}\cdot 0.498 + 5\frac{5}{12}\cdot 0.502$$

Markov's inequality

"Not too many people can be well above the average."

Suppose **X** is a non-negative r.v. with E[**X**] = 10

How often can **X** be 20 or higher?

i.e., How high can Pr [**X** ≥ 20] be?

$E[\mathbf{X}] = E[\mathbf{X} \mid \mathbf{X} \geq 20]\ \Pr[\mathbf{X} \geq 20] + E[\mathbf{X} \mid \mathbf{X} < 20]\ \Pr[\mathbf{X} < 20]$

$\geq E[\mathbf{X} \mid \mathbf{X} \geq 20]\ \Pr[\mathbf{X} \geq 20] \quad \geq 20 \Pr[\mathbf{X} \geq 20]$

So $\Pr[\mathbf{X} \geq 20] \leq E[\mathbf{X}]/20 = ½$.

Markov's inequality: For a non-negative r.v. **X**,

$$\Pr[\mathbf{X} \geq a] \leq \frac{E[\mathbf{X}]}{a} \quad \text{for every } a > 0.$$

Study Bee

- Basic sample spaces
- Binomial & Geometric dist.
- Random variables
  - their dual views
- Independence of R.Vs
- Expectation of R.Vs
- Linearity of Expectation
- Basic use of the probabilistic method

Supplementary material

[Another linearity of expectation example and Birthday paradox]

Enemybook

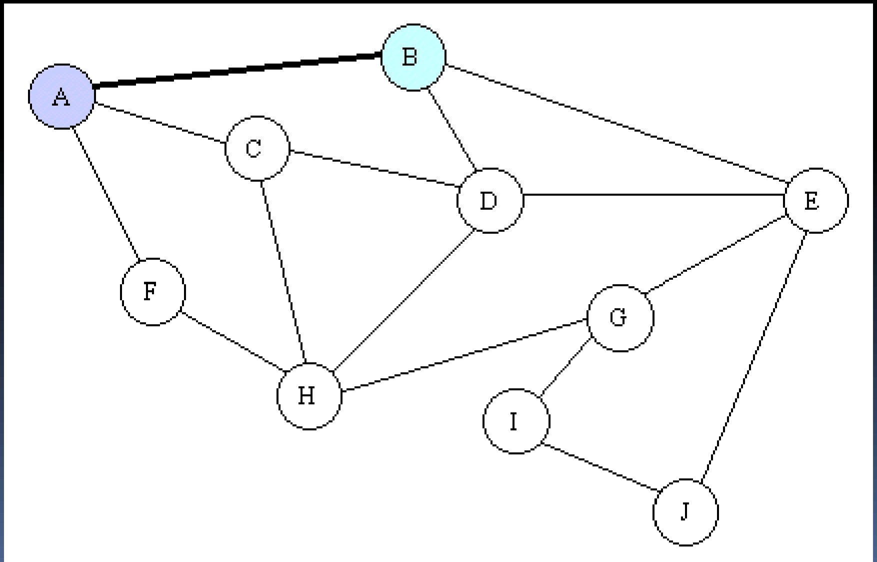
www.enemybook.org

Enemybook
Keep your friends close and your enemies closer...

Enemybook is an ***anti-social utility*** *that* ***disconnects you*** *to the so-called friends around you.*

On Enemybook, Enemyships connect pairs of people

Suppose there are n students with m enemyships between them



Enemybook Schism

Suppose there are n students with m enemyships between them

We would like to devise a schism in enemybook.

i.e., split the students into two teams so that many enemyships are broken.

Prove that, no matter what the enemybook network, we can always do this in a way that breaks at least $m/2$ enemyships

Enemybook Schism

Prove that, no matter what the enemybook network, we can always devise a partition into two teams that breaks at least ½ the enemyships

Here's a simple (almost dumb) thing to try:

For each student, place him/her in team 1 or 2 randomly (independent of other students)

Let **X** = number of enemyships broken

E[**X**] = ?

Indicators + Linearity to the rescue

For each of the m enemyships e,

let B_e be the event that it's broken,

let $\mathbf{X}_e$ be the indicator r.v for B_e.

$$\mathbf{X} = \sum_{\text{enemyships } e} \mathbf{X}_e$$

$$\therefore \ \mathbf{E}[\mathbf{X}] = \sum_e \mathbf{E}[\mathbf{X}_e] = \sum_e \mathbf{Pr}[B_e]$$

$\mathbf{Pr}[B_e] = 1/2$ (broken if 1,2 or 2,1)

Indicators + Linearity to the rescue

For each of the m enemyships e,

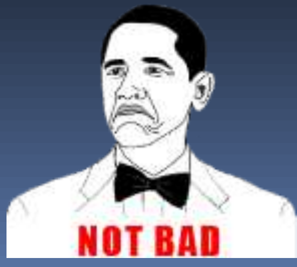
let B_e be the event that it's broken,

let $\mathbf{X_e}$ be the indicator rv for B_e.

$$\mathbf{X} = \sum_{\text{enemyships } e} \mathbf{X}_e \quad \therefore \; \mathbf{E}[\mathbf{X}] = \sum_e \mathbf{E}[\mathbf{X}_e] = \sum_e \mathbf{Pr}[B_e]$$

$$\mathbf{Pr}[B_e] = 1/2 \qquad \therefore \; \mathbf{E}[\mathbf{X}] = (1/2)m$$

By the probabilistic method, there must exist schisms that separate *at least* m/2 pairs.



Birthday Problem

Question:

There are m students in a room ($m \leq 365$).

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.

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

Chain rule:

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

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.

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

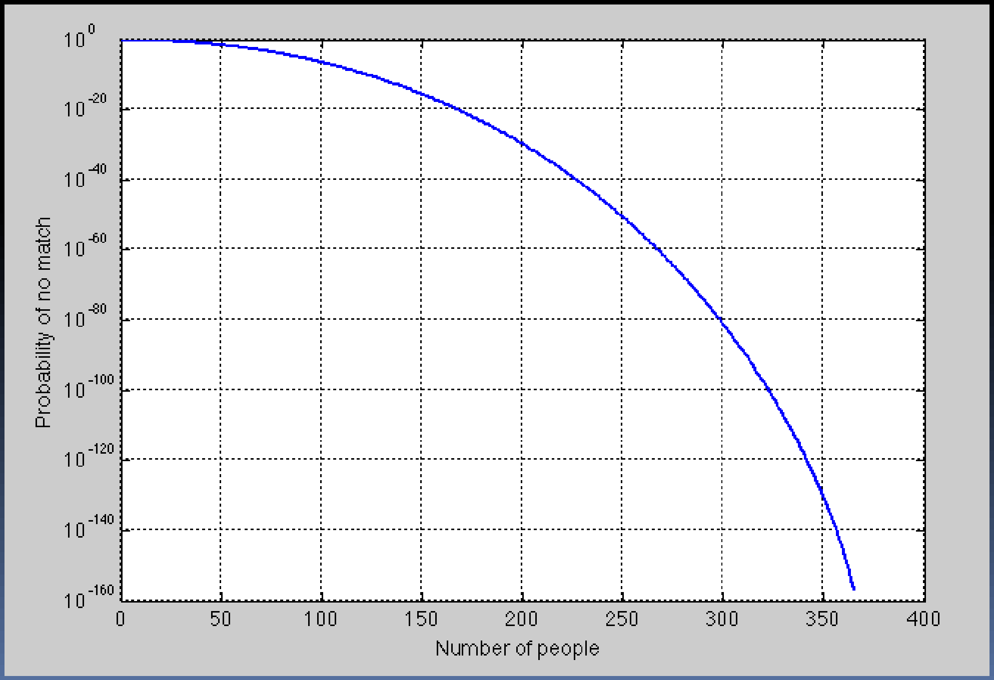
$$= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{m-1}{365}\right)$$

This is the final answer.

Birthday Problem — Analysis

Pr[all m students have different bdays]

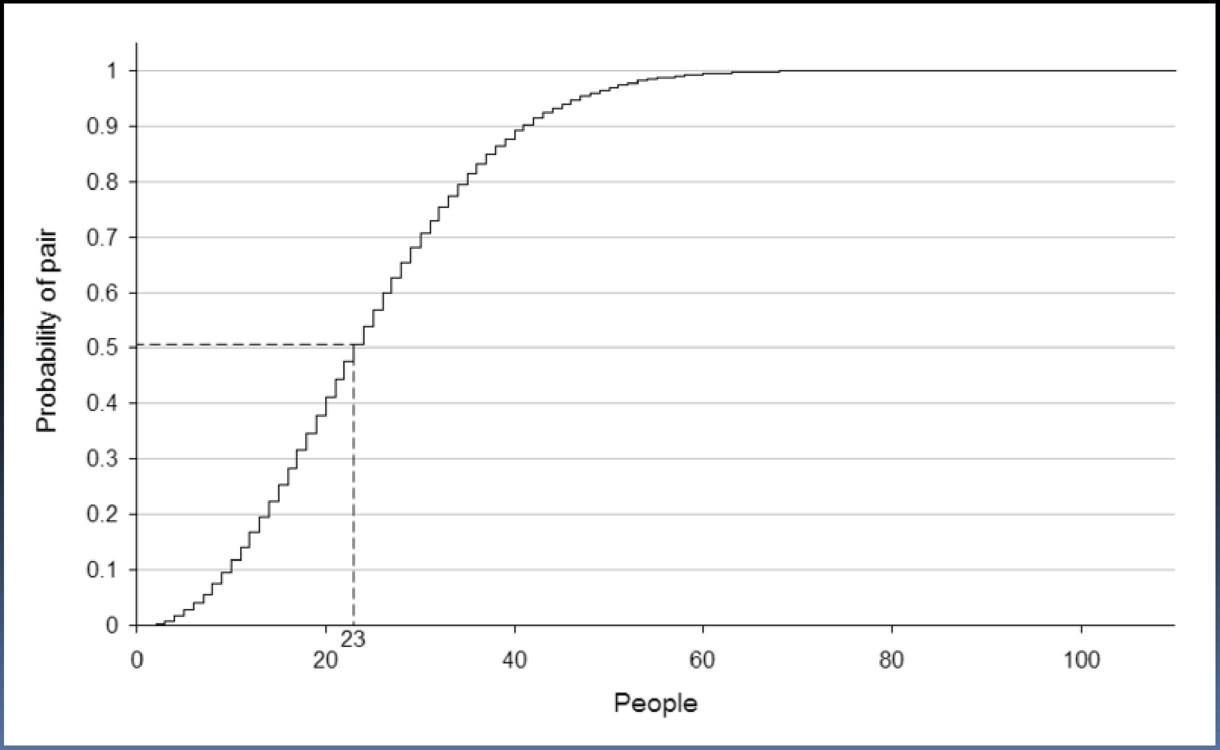
$$= 1 \cdot \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{m-1}{365}\right)$$



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 ?

I'll just tell you: for $m \approx \sqrt{N}$