Great Theoretical Ideas In Computer Science

Steven Rudich, Anupam Gupta — CS 15-251 — Spring 2005

Lecture 21 — March 29, 2005 — Carnegie Mellon University

Today, we will learn about a formidable tool in probability that will allow us to solve problems that seem <u>really really messy</u>…

If I randomly put 100 letters into 100 addressed envelopes, <u>on average</u> how many letters will end up in their correct envelopes?

Hmm…

$\sum_k k \cdot Pr(\textit{exactly } k \text{ letters end up in correct envelopes})$

$= \sum_k k \cdot$ (…aargh!!…)

On average, in class of size m, how many pairs of people will have the same birthday?

$\sum_k k \cdot Pr(\text{exactly } k \text{ collisions})$

$= \sum_k k \cdot$ (…aargh!!!!…)

The new tool is called "Linearity of Expectation"

Expectatus Linearitus

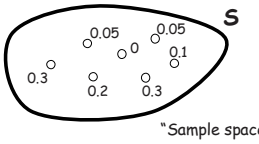
Random Variable

To use this new tool, we will also need to understand the concept of a <u>Random Variable</u>

Today's goal: not too much material, but to understand it well.

Probability Distribution

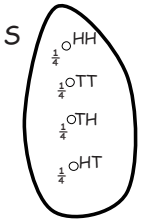
A (finite) probability distribution D
- a finite set S of elements (samples)
- each $x \in S$ has weight or probability $p(x) \in [0,1]$



weights must sum to 1

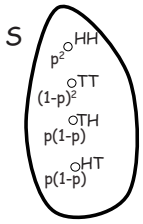
"Sample space"

Flip penny and nickel (unbiased)

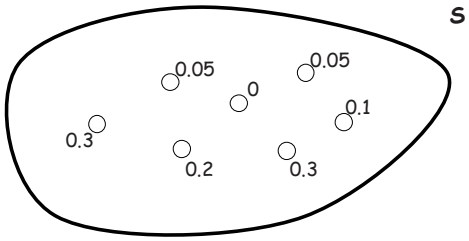


Flip penny and nickel (biased)

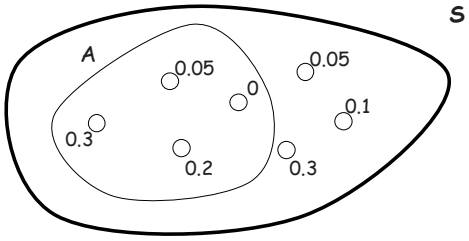
heads probability = p



Probability Distribution



An event is a subset



$Pr[A] = \sum_{x \in A} p(x) = 0.55$

Running Example

I throw a white die and a black die.

Sample space S =

(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) }

Pr(x) = 1/36
$\forall x \in S$

E = event that sum ≤ 3

Pr[E] = |E|/|S| = proportion of E in S = 3/36

New concept: Random Variables

Random Variables

Random Variable: a (real-valued) function on S

Examples:

X = value of white die.
X(3,4) = 3, X(1,6) = 1 etc.

Y = sum of values of the two dice.
Y(3,4) = 7, Y(1,6) = 7 etc.

W = (value of white die)$^{\text{value of black die}}$
W(3,4) = 3^4 Y(1,6) = 1^6

Z = (1 if two dice are equal, 0 otherwise)
Z(4,4) = 1, Z(1,6) = 0 etc.

Toss a white die and a black die.

Sample space S =
{ (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) }

E.g., tossing a fair coin n times

S = all sequences of $\{H, T\}^n$
D = uniform distribution on S
$\Rightarrow D(x) = (\frac{1}{2})^n$ for all $x \in S$

Random Variables (say n = 10)
X = # of heads
X(HHHTTHTHTT) = 5
Y = (1 if #heads = #tails, 0 otherwise)
Y(HHHTTHTHTT) = 1, Y(THHHHTTTTT) = 0

Notational conventions

Use letters like A, B, E for events.

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

R.V. = random variable

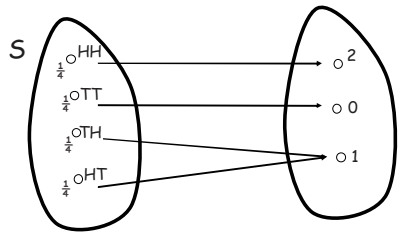
Two views of random variables

Think of a R.V. as

- a function from S to the reals P
- or think of the induced distribution on P

Two coins tossed

X: {TT, TH, HT, HH} -> {0, 1, 2}
counts the number of heads



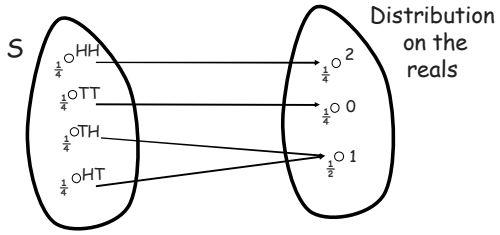
Two views of random variables

Think of a R.V. as

- a function from S to the reals P
- or think of the induced distribution on P

Two coins tossed

X: {TT, TH, HT, HH} -> {0, 1, 2}
counts the number of heads



Two views of random variables

Think of a R.V. as

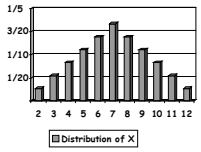
- a function from S to the reals P
- or think of the induced distribution on P

Two dice

I throw a white die and a black die.

Sample space S =

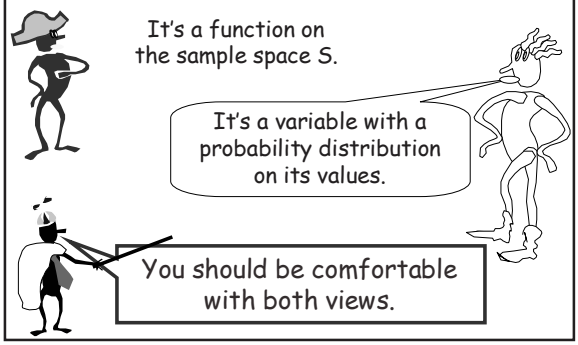
(1,1),	(1,2),	(1,3),	(1,4),	(1,5),	(1,6),
(2,1),	(2,2),	(2,3),	(2,4),	(2,5),	(2,6),
(3,1),	(3,2),	(3,3),	(3,4),	(3,5),	(3,6),
(4,1),	(4,2),	(4,3),	(4,4),	(4,5),	(4,6),
(5,1),	(5,2),	(5,3),	(5,4),	(5,5),	(5,6),
(6,1),	(6,2),	(6,3),	(6,4),	(6,5),	(6,6) }



X = sum of both dice

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

It's a floor wax *and* a dessert topping



From Random Variables to Events

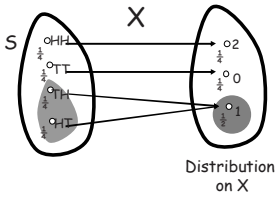
For any random variable X and value a, we can define the event A that X=a.

$$Pr(A) = Pr(X=a) = Pr(\{x \in S | X(x)=a\}).$$

Two coins tossed

X: {TT, TH, HT, HH} -> {0, 1, 2}
counts the number of heads

$\Pr(X = a) = \Pr(\{x \in S \mid X(x) = a\})$



$\Pr(X = 1)$

$= \Pr(\{x \in S \mid X(x) = 1\})$

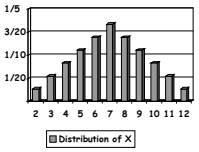
$= \Pr(\{TH, HT\}) = \frac{1}{2}$.

Two dice

I throw a white die and a black die. X = sum

Sample space S =

(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) }



$\Pr(X = 6)$
$= \Pr(\{x \in S \mid X(x) = 6\})$
$= 5/36$.

$\Pr(X = a) = \Pr(\{x \in S \mid X(x) = a\})$

From Random Variables to Events

For any random variable X and value a, we can define the <u>event</u> A that X=a.

$\Pr(A) = \Pr(X=a) = \Pr(\{x \in S \mid X(x)=a\})$.

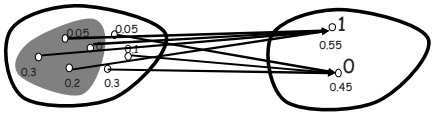
X has a distribution on its values

X is a function on the sample space S

From Events to Random Variables

For any event A, can define the <u>indicator random variable</u> for A:

$X_A(x) = 1$ if $x \in A$
$= 0$ if $x \notin A$



Definition: expectation

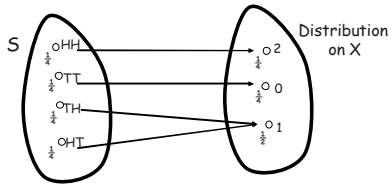
The <u>expectation</u>, or <u>expected value</u> of a random variable X is written as E[X], and is

$$E[X] = \sum_{x \in S} \Pr(x)X(x) = \sum_{k} k \Pr(X = k)$$

X is a function on the sample space S

X has a distribution on its values

Thinking about expectation



$\sum_{x \in S} X(x) \Pr(x)$

$\sum_{k} k \Pr(X = k)$

$E[X] = \frac{1}{4}*0 + \frac{1}{4}*1 + \frac{1}{4}*1 + \frac{1}{4}*2 = 1.$

$E[X] = \frac{1}{4}*0 + \frac{1}{2}*1 + \frac{1}{4}*2 = 1.$

A quick calculation…

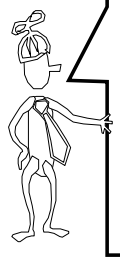
What if I flip a coin 3 times? Now what is the expected number of heads?

$E[X] = (1/8) \times 0 + (3/8) \times 1 + (3/8) \times 2 + (1/8) \times 3 = 1.5$

But Pr[X = 1.5] = 0…

Moral: don't always expect the expected.
Pr[X = E[X]] may be 0 !

Type checking



A Random Variable is the type of thing you might want to know an expected value of.

If you are computing an expectation, the thing whose expectation you are computing is a random variable.

Indicator R.V.s: $E[X_A] = Pr(A)$

For event A, the indicator random variable for A:

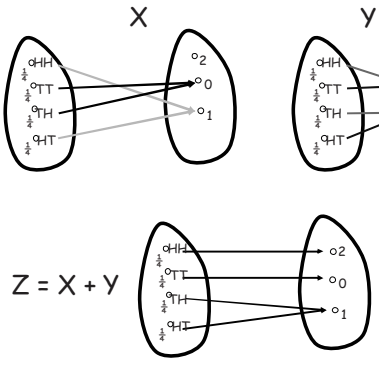
$X_A(x) = 1$ if $x \in A$
$= 0$ if $x \notin A$

$E[X_A] = 1 \times Pr(X_A = 1) = Pr(A)$

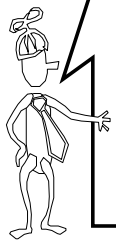
Adding Random Variables

If X and Y are random variables (on the same set S), then
Z = X + Y is also a random variable.
$Z(x) \equiv X(x) + Y(x)$

E.g., rolling two dice.
X = 1st die, Y= 2nd die,
Z = sum of two dice.



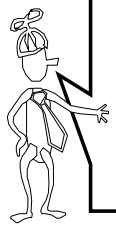
Adding Random Variables



Example: Consider picking a random person in the world. Let X = length of the person's left arm in inches. Y = length of the person's right arm in inches. Let Z = X+Y. Z measures the combined arm lengths.

Formally, S = {people in world},
D = uniform distribution on S.

Independence



Two random variables X and Y are independent if for every a,b, the events X=a and Y=b are independent.

How about the case of
X=1st die, Y=2nd die?
X = left arm, Y=right arm?

Linearity of Expectation

If Z = X+Y, then

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

Even if X and Y are not independent.

Linearity of Expectation

If Z = X+Y, then

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

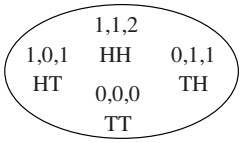
Proof:

$E[X] = \sum_{x \in S} Pr(x)\, X(x)$
$E[Y] = \sum_{x \in S} Pr(x)\, Y(x)$
$E[Z] = \sum_{x \in S} Pr(x)\, Z(x)$
but $Z(x) = X(x) + Y(x)$

Linearity of Expectation

E.g., 2 fair flips:
X = 1st coin, Y = 2nd coin.
Z = X+Y = total # heads.

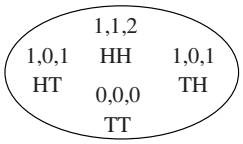
What is E[X] ? E[Y] ? E[Z] ?



Linearity of Expectation

E.g., 2 fair flips:
X = at least one coin heads,
Y = both coins are heads, Z = X+Y

Are X and Y independent?
What is E[X] ? E[Y] ? E[Z] ?

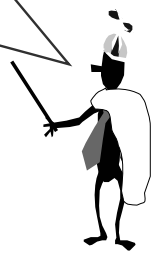


By induction

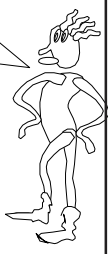
$E[X_1 + X_2 + \ldots + X_n] =$
$E[X_1] + E[X_2] + \ldots. + E[X_n]$

the expectation
of the sum
=
the sum of the
expectations

It is finally time to show off our probability prowess…



If I randomly put 100 letters into 100 addressed envelopes, on average how many letters will end up in their correct envelopes?



Hmm…

Σ_k k·Pr(*exactly* k letters end up in correct envelopes)
= Σ_k k· (…aarghh!…)

Use Linearity of Expectation

Let A_i be the event the i^{th} letter ends up in its correct envelope.

Let X_i be the indicator R.V. for A_i.

$$X_i = \begin{cases} 1 \text{ if } A_i \text{ occurs} \\ 0 \text{ otherwise} \end{cases}$$

Let $Z = X_1 + \ldots + X_{100}$.

We are asking for E[Z].

Use Linearity of Expectation

Let A_i be the event the i^{th} letter ends up in the correct envelope.
Let X_i be the indicator R.V. for A_i.
Let $Z = X_1 + \ldots + X_n$. We are asking for E[Z].

What is E[X_i]?
E[X_i] = Pr(A_i) = 1/100.

What is E[Z]?
E[Z] = E[X_1 + … + X_{100}]
= E[X_1] + … + E[X_{100}]
= 1/100 + … + 1/100 = 1.

Use Linearity of Expectation

So, in expectation, 1 card will be in the same position as it started.

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

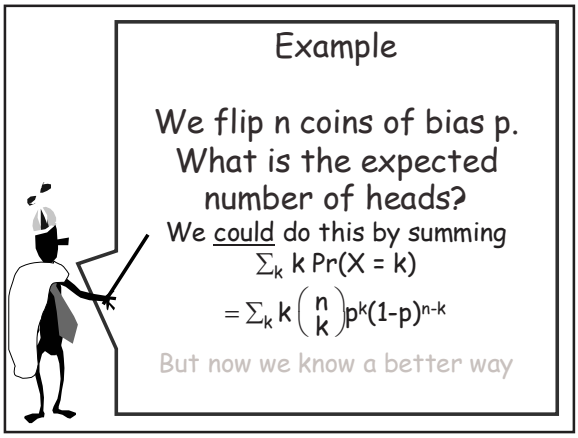
Question: were the X_i independent?

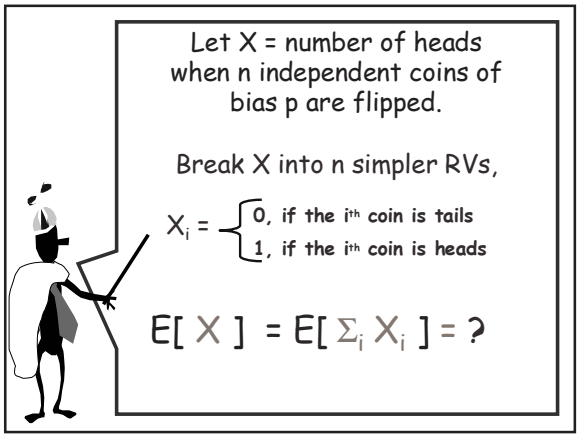
No! E.g., think of n=2.

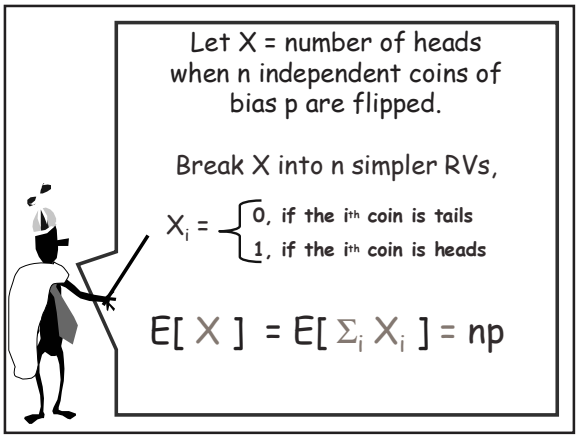
Use Linearity of Expectation

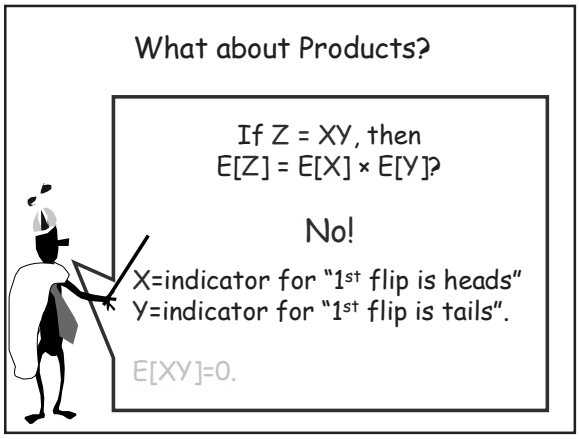
General approach:

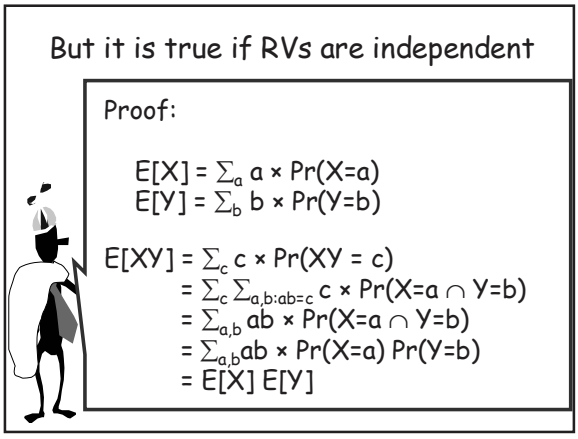
- View thing you care about as expected value of some RV.
- Write this RV as sum of simpler RVs (typically indicator RVs).
- Solve for their expectations and add them up!

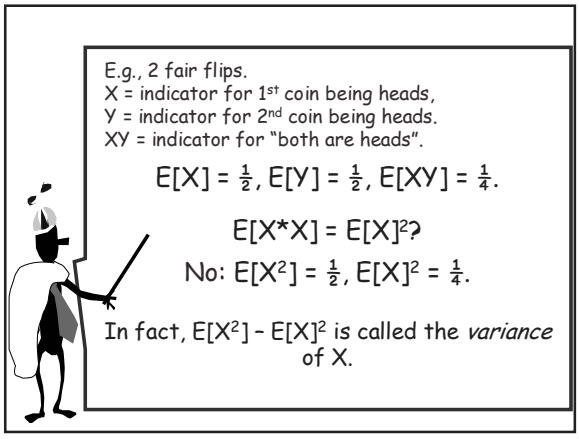

Example
We flip n coins of bias p. What is the expected number of heads?
We could do this by summing
$\sum_k k \Pr(X = k)$
$= \sum_k k \binom{n}{k} p^k (1-p)^{n-k}$
But now we know a better way

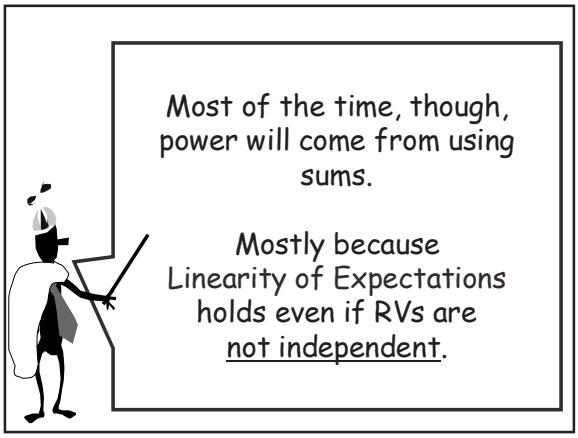

Let X = number of heads when n independent coins of bias p are flipped.
Break X into n simpler RVs,
$X_i = \begin{cases} 0, \text{ if the } i^{th} \text{ coin is tails} \\ 1, \text{ if the } i^{th} \text{ coin is heads} \end{cases}$
$E[X] = E[\sum_i X_i] = ?$

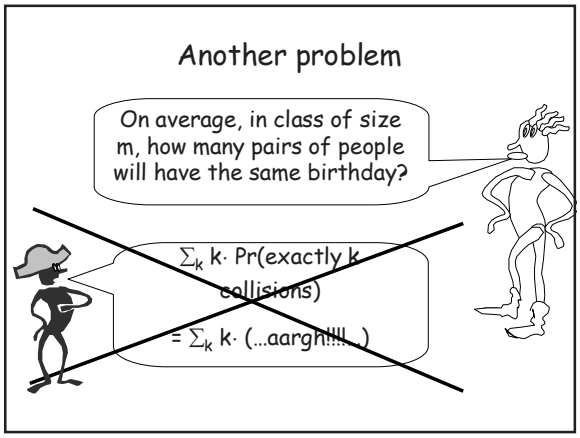

Let X = number of heads when n independent coins of bias p are flipped.
Break X into n simpler RVs,
$X_i = \begin{cases} 0, \text{ if the } i^{th} \text{ coin is tails} \\ 1, \text{ if the } i^{th} \text{ coin is heads} \end{cases}$
$E[X] = E[\sum_i X_i] = np$

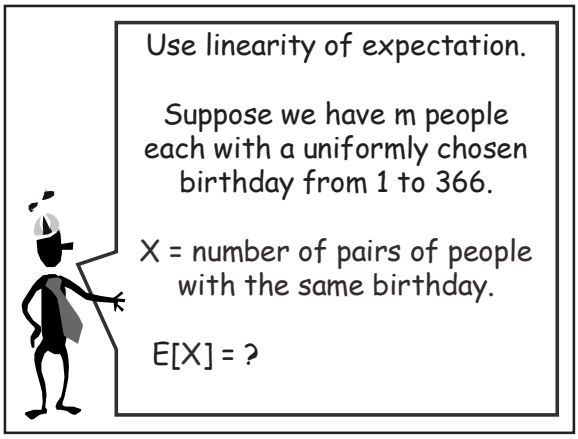

What about Products?
If Z = XY, then E[Z] = E[X] × E[Y]?
No!
X=indicator for "1st flip is heads"
Y=indicator for "1st flip is tails".
E[XY]=0.

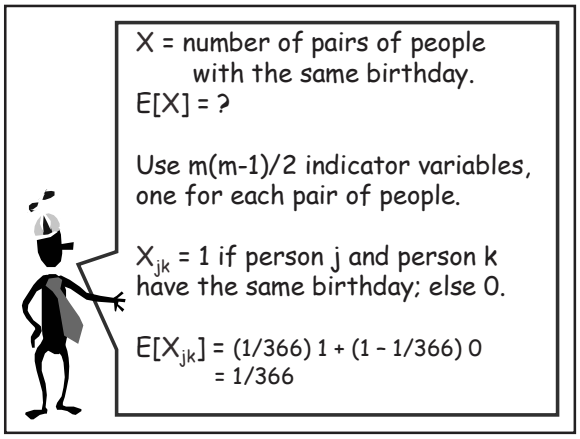

But it is true if RVs are independent
Proof:
$E[X] = \sum_a a \times \Pr(X=a)$
$E[Y] = \sum_b b \times \Pr(Y=b)$
$E[XY] = \sum_c c \times \Pr(XY = c)$
$= \sum_c \sum_{a,b:ab=c} c \times \Pr(X=a \cap Y=b)$
$= \sum_{a,b} ab \times \Pr(X=a \cap Y=b)$
$= \sum_{a,b} ab \times \Pr(X=a) \Pr(Y=b)$
$= E[X] E[Y]$

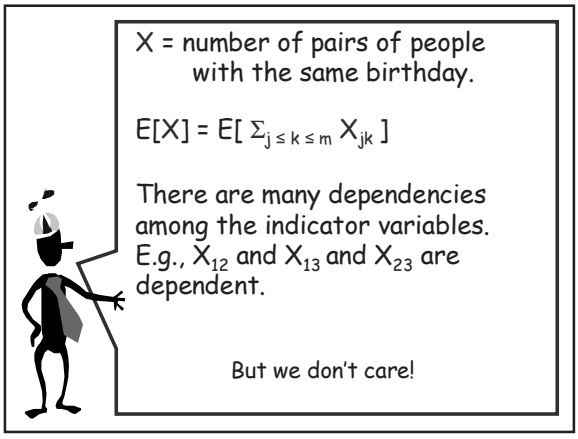

E.g., 2 fair flips.
X = indicator for 1st coin being heads,
Y = indicator for 2nd coin being heads.
XY = indicator for "both are heads".
$E[X] = \frac{1}{2}, E[Y] = \frac{1}{2}, E[XY] = \frac{1}{4}.$
$E[X*X] = E[X]^2$?
No: $E[X^2] = \frac{1}{2}, E[X]^2 = \frac{1}{4}.$
In fact, $E[X^2] - E[X]^2$ is called the *variance* of X.

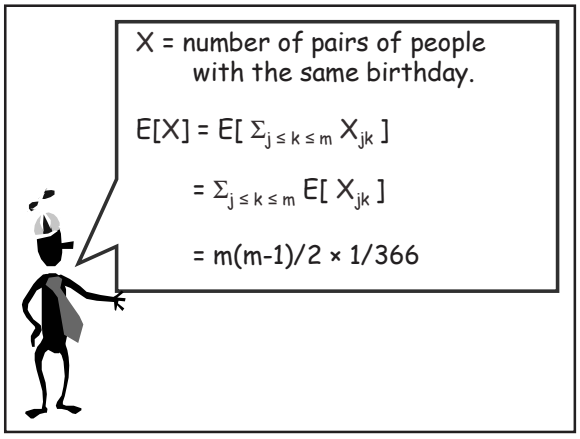

Most of the time, though, power will come from using sums.
Mostly because Linearity of Expectations holds even if RVs are not independent.


Another problem
On average, in class of size m, how many pairs of people will have the same birthday?
Σ_k k· Pr(exactly k collisions)
= Σ_k k· (…aargh!!!! …)

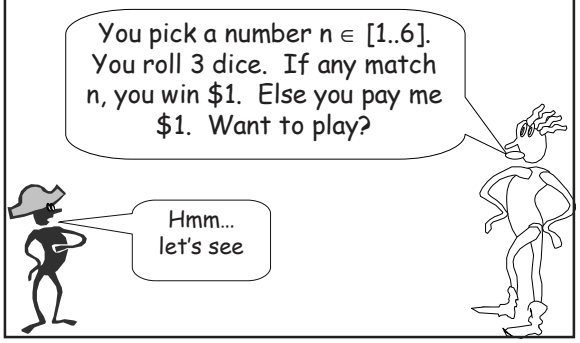

Use linearity of expectation.
Suppose we have m people each with a uniformly chosen birthday from 1 to 366.
X = number of pairs of people with the same birthday.
E[X] = ?


X = number of pairs of people with the same birthday.
E[X] = ?
Use m(m-1)/2 indicator variables, one for each pair of people.
X_{jk} = 1 if person j and person k have the same birthday; else 0.
$E[X_{jk}]$ = (1/366) 1 + (1 – 1/366) 0
= 1/366


X = number of pairs of people with the same birthday.
E[X] = E[$\Sigma_{j \le k \le m} X_{jk}$]
There are many dependencies among the indicator variables. E.g., X_{12} and X_{13} and X_{23} are dependent.
But we don't care!


X = number of pairs of people with the same birthday.
E[X] = E[$\Sigma_{j \le k \le m} X_{jk}$]
= $\Sigma_{j \le k \le m}$ E[X_{jk}]
= m(m-1)/2 × 1/366

Step right up…



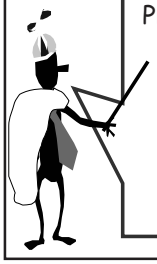
Analysis

A_i = event that i-th die matches.
X_i = indicator RV for A_i.

Expected number of dice that match:
$E[X_1+X_2+X_3] = 1/6+1/6+1/6 = \frac{1}{2}$.

But this is not the same as Pr(at least one die matches).

Analysis

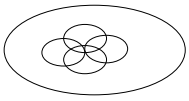


Pr(at least one die matches)
= 1 – Pr(none match)
= $1 - (5/6)^3 = 0.416$.

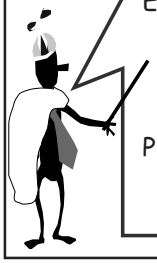
What's going on?

Say we have a collection of events A_1, A_2, ….

How does E[# events that occur] compare to Pr(at least one occurs)?



What's going on?



E[# events that occur]
= $\sum_k$ Pr(k events occur) × k
= $\sum_{(k>0)}$ Pr(k events occur) × k

Pr(at least one event occurs)
= $\sum_{(k>0)}$ Pr(k events occur)

What's going on?

Moral #1: be careful you are modeling problem correctly.

Moral #2: watch out for carnival games.