15-251
Great Theoretical Ideas in Computer Science

Lecture 22:
Intro to Randomness and Probability Theory 2

April 6th, 2017

SUMMARY SO FAR

Real World → Code → Probability Tree

=

Mathematical Model

- set of outcomes Ω
- a prob. distribution

Events

Conditional probability:

$$\Pr[A \mid B] = \Pr[A \cap B] / \Pr[B]$$

Chain rule:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B \mid A]$$

Law of total probability:

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

Independent events:

$$\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$$

Union bound:

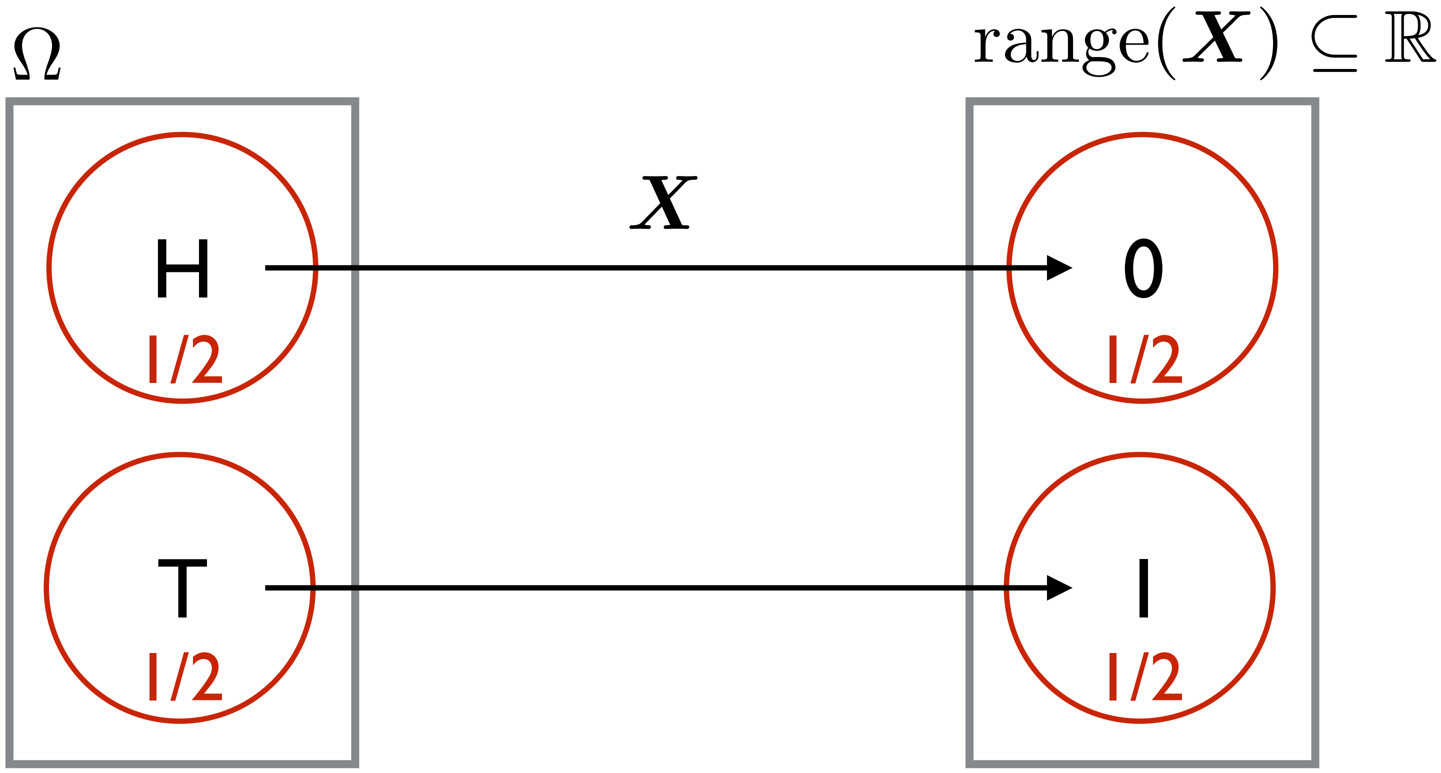
$$\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$$

Random Variables

What is a Random Variable?

Transformation of Ω to $\mathbb{R}$

i.e. a function $\boldsymbol{X} : \Omega \to \mathbb{R}$

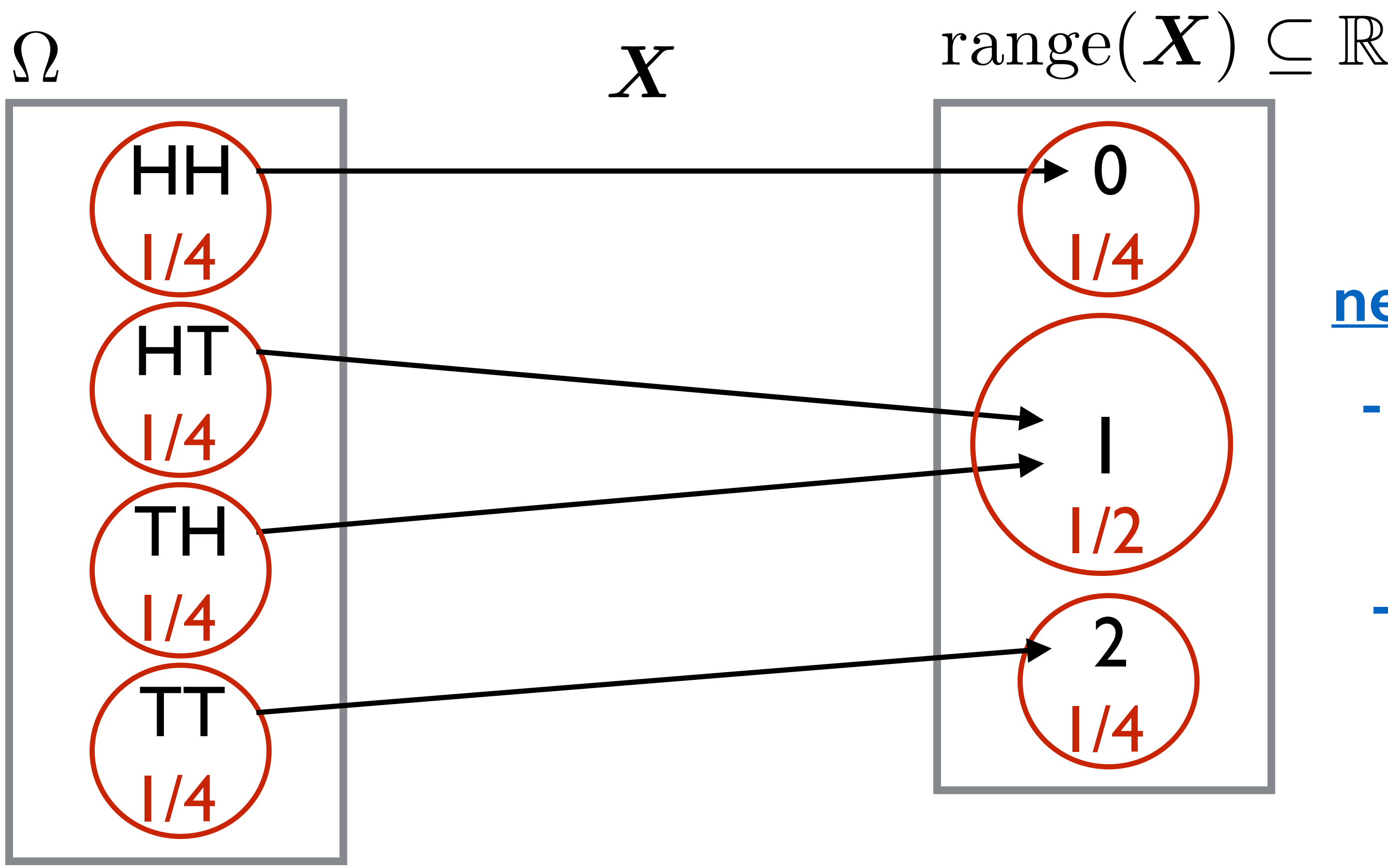


<u>typical description:</u> $\boldsymbol{X}$ = number of Tails

What is a Random Variable?

Transformation of Ω to $\mathbb{R}$

i.e. a function $\boldsymbol{X} : \Omega \to \mathbb{R}$



new prob. space:

- new sample space (values X can take)
- new prob. distr.

typical description: X = number of Tails

What is a Random Variable?

Transformation of Ω to $\mathbb{R}$

i.e. a function $S : \Omega \to \mathbb{R}$

$\Omega =$

{(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)}

Distribution:

for each $\ell \in \Omega$:

$$\Pr[\ell] = 1/36$$

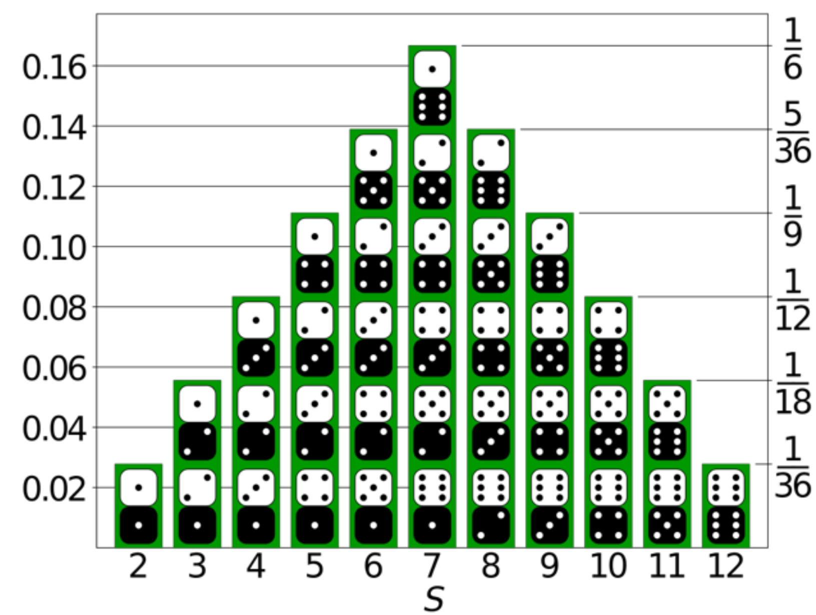
('uniform distribution')

S = sum of two dice

$\Omega' = \text{range}(S) =$

{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

Distribution:



Why?

- Often we are interested in numerical outcomes

  (e.g. number of Tails we see if we toss *n* coins)

  but initially outcomes are best expressed non-numerically.

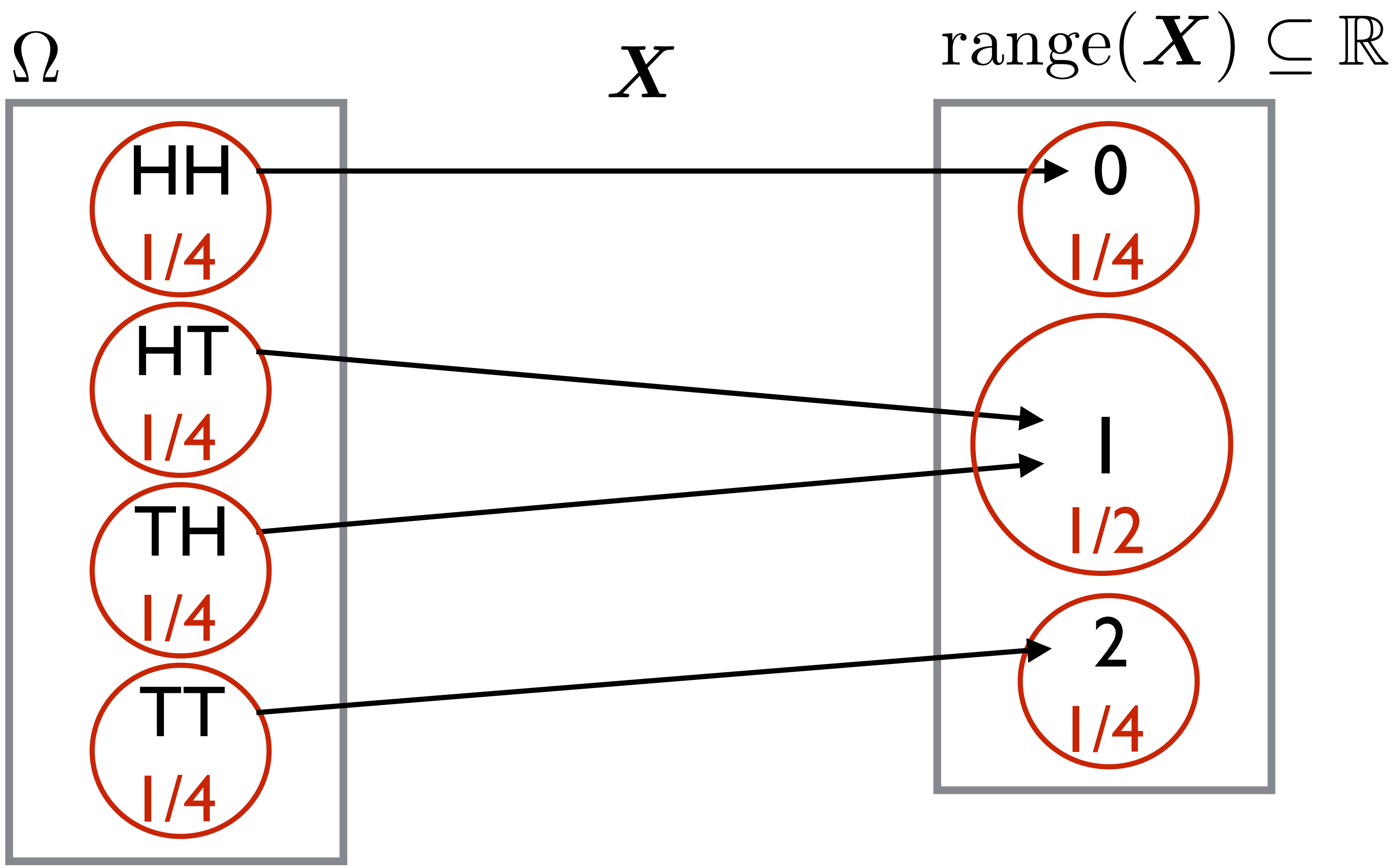
  (e.g. an outcome is a sequence of *n* coin tosses)

- We like talking about "expected values" (averages).

What is a Random Variable?

Transformation of Ω to $\mathbb{R}$

i.e. a function $\boldsymbol{X} : \Omega \to \mathbb{R}$



What is the "average number" of Tails?

$$0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} = 1$$

What is a Random Variable?

2nd Definition:

A **random variable** is a variable in some **randomized code**
(more accurately, the variable's value at the end of the execution)

of type 'real number'.

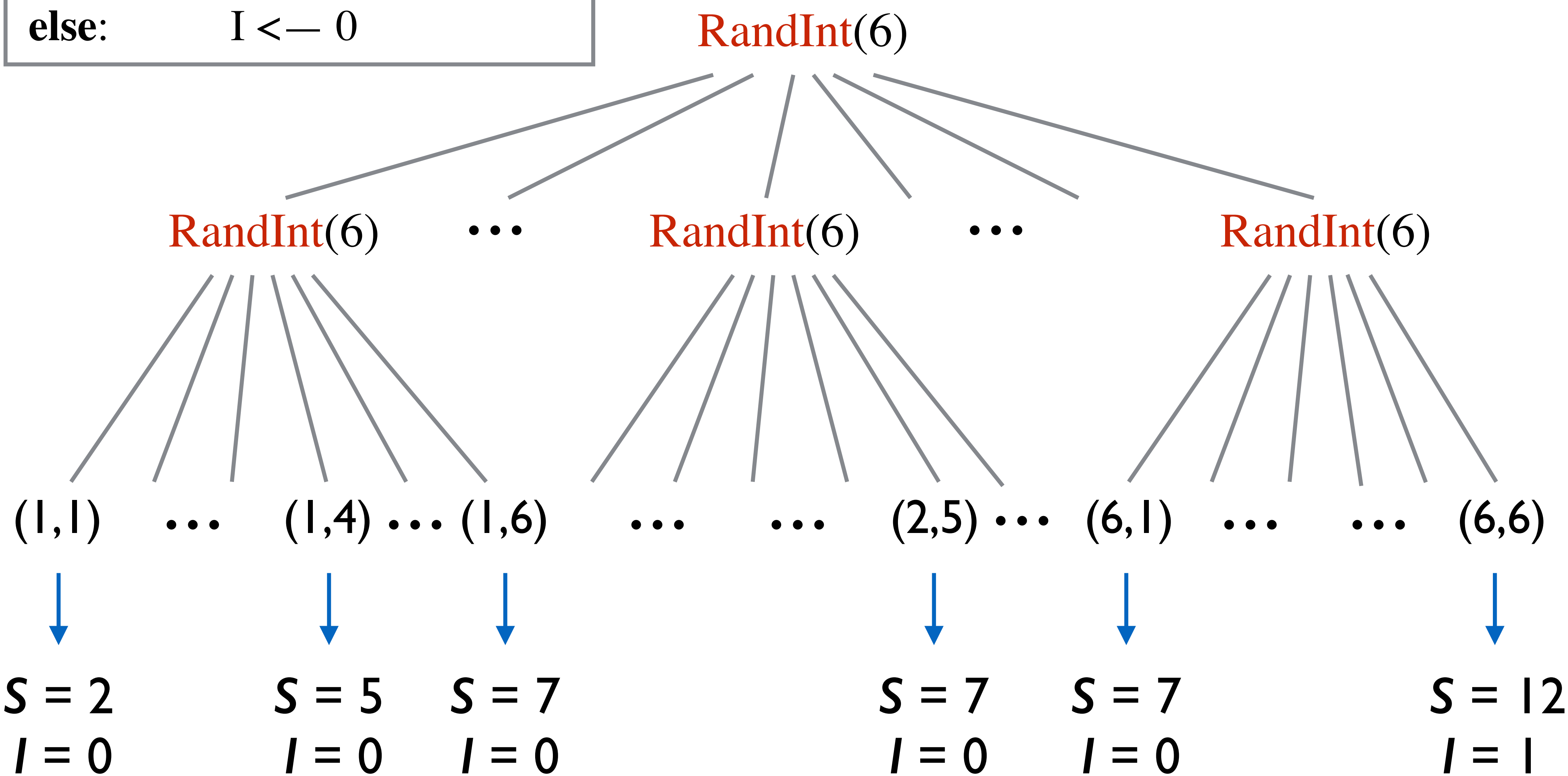
Example:

```
S <— RandInt(6) + RandInt(6)
if S = 12:   I <— 1
else:        I <— 0
```

Random variables: *S* and *I*

What is a Random Variable?

```
S <— RandInt(6) + RandInt(6)
if S = 12:   I <— 1
else:        I <— 0
```



Explicitly Defining the Distribution of a R.V.

$\Omega = \{$(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)$\}$

Distribution:

$$\Pr : \Omega \to [0, 1]$$

for each $\ell \in \Omega$:

$$\Pr[\ell] = 1/36$$

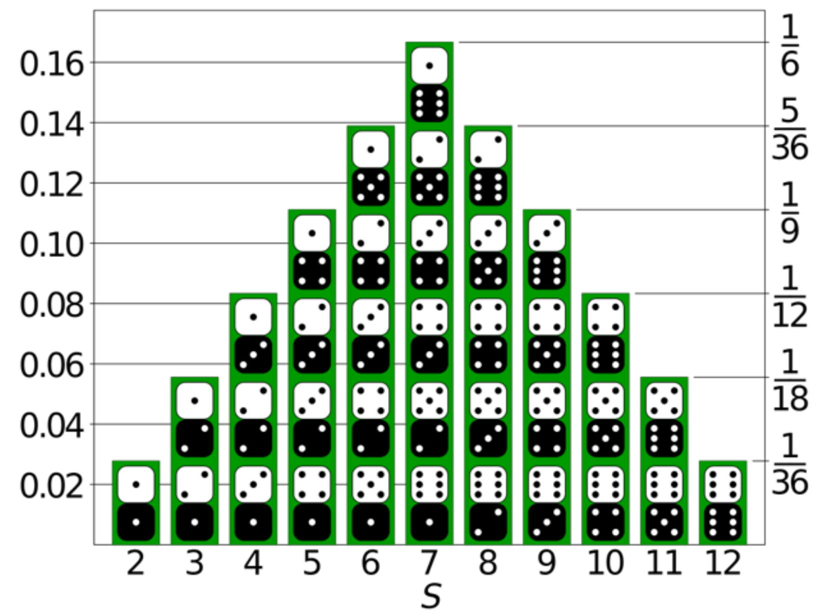
'uniform distribution'

$\boldsymbol{S}$ = sum of two dice

$\Omega' = \text{range}(\boldsymbol{S}) =$

{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}

Distribution:

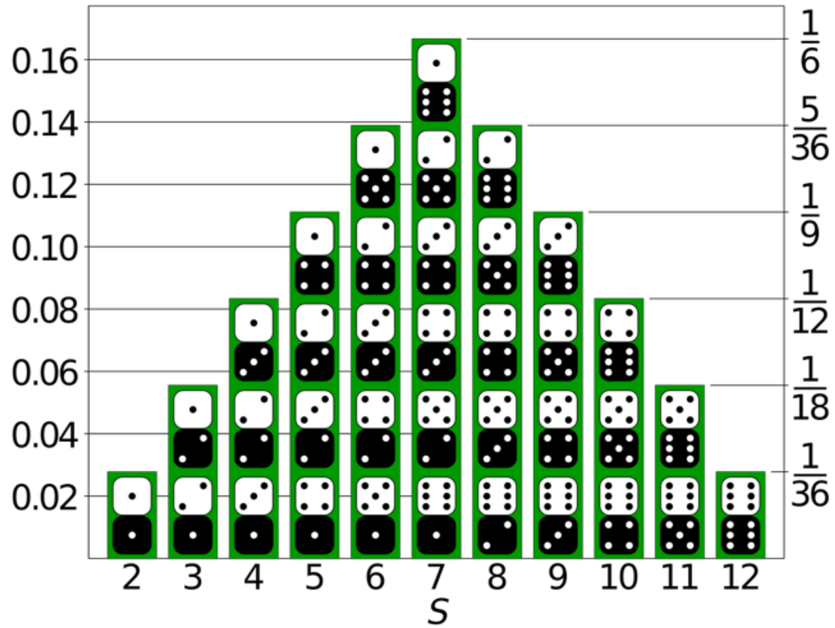


$$\Pr_{\boldsymbol{S}} : \text{range}(\boldsymbol{S}) \to [0, 1]$$

for each $s \in \text{range}(\boldsymbol{S})$: $\Pr_{\boldsymbol{S}}[s] =$

Notation: $\Pr[\boldsymbol{S} = s] = \Pr_{\boldsymbol{S}}[s]$

Explicitly Defining the Distribution of a R.V.



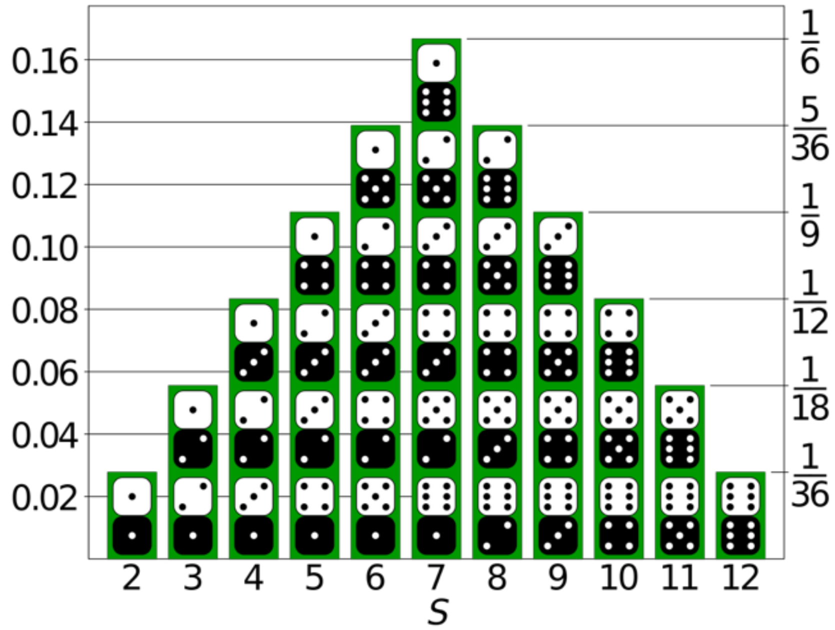
So $\boldsymbol{S} = x$ is a shorthand for the event $\{\ell \in \Omega : \boldsymbol{S}(\ell) = x\}$

$$\Pr[\boldsymbol{S} = x] = \Pr[\ell \in \Omega : \boldsymbol{S}(\ell) = x]$$

Example:

$$\Pr[\boldsymbol{S} = 3] = \Pr[\ell \in \Omega : \boldsymbol{S}(\ell) = 3] = \Pr[\{(1,2),(2,1)\}] = 1/18$$

Explicitly Defining the Distribution of a R.V.



Similarly $\boldsymbol{S} \geq x$ is a shorthand for the event $\{\ell \in \Omega : \boldsymbol{S}(\ell) \geq x\}$

$$\Pr[\boldsymbol{S} \geq x] = \Pr[\ell \in \Omega : \boldsymbol{S}(\ell) \geq x]$$

etc...

Random Variables: How they are introduced

1. Retroactively

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

$$D((1,1)) = 0$$
$$D((2,1)) = -1$$

$$\dots$$

Random Variables: How they are introduced

2. In terms of other random variables

"Let $Y = S^2 + D$"

$$\boldsymbol{Y}((5,3)) = 62$$

$$\ldots$$

Random Variables: How they are introduced

3. Without bothering to give an "experiment"

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

"Let *T* be a random variable that is distributed uniformly over the set {0, 2, 4, 6, 8}."

Describe the **probability mass function (PMF)**.

i.e., the values $\Pr[\boldsymbol{X} = x]$ for all $x \in \mathrm{range}(\boldsymbol{X})$.

(Don't need to think about the "original" Ω.)

Independent Random Variables

Random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ are independent if

for all $x \in \text{range}(\boldsymbol{X}), \quad y \in \text{range}(\boldsymbol{Y})$

the events $\boldsymbol{X} = x$ and $\boldsymbol{Y} = y$ are independent.

i.e. $\Pr[\boldsymbol{X} = x \text{ and } \boldsymbol{Y} = y] = \Pr[\boldsymbol{X} = x] \cdot \Pr[\boldsymbol{Y} = y]$

(similarly for more than 2 random variables)

Expectation of a Random Variable

Expected Value = Mean = (Weighted) Average

Example:

Weight	Value
30% Final	85
20% Midterm	75
50% Homework	82

Weighted Average = $0.3 \cdot 85 + 0.2 \cdot 75 + 0.5 \cdot 82 = 81.5$

Weighted Average = $\sum_{\text{elements } e} \text{value}(e) \cdot \text{weight}(e)$

Expectation of a Random Variable

Expected value of a random variable $\boldsymbol{X}$:

$$\mathbf{E}[\boldsymbol{X}] \stackrel{\text{def}}{=} \sum_{x \in \mathrm{range}(\boldsymbol{X})} x \cdot \Pr[\boldsymbol{X} = x]$$

Expectation of a Random Variable

Example

Let $\boldsymbol{X}$ be the outcome of the roll of a 6-sided die.

$$\mathbf{E}[\boldsymbol{X}]$$

$$= 1 \cdot \Pr[\boldsymbol{X} = 1] + 2 \cdot \Pr[\boldsymbol{X} = 2] + \cdots + 6 \cdot \Pr[\boldsymbol{X} = 6]$$

$$= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6}$$

$$= 3.5$$

What is $\Pr[\boldsymbol{X} = 3.5]$?

(Don't always expect the expected!)

Expectation of a Random Variable

Example

Let $\boldsymbol{X}$ = RandInt(6), $\boldsymbol{Y}$ = RandInt(6), $\boldsymbol{Z}$ = RandInt(6)

Let $\boldsymbol{S} = \boldsymbol{X} + \boldsymbol{Y} + \boldsymbol{Z}$

$\mathbf{E}[\boldsymbol{S}]$

$= 3 \cdot \Pr[\boldsymbol{S} = 3] + 4 \cdot \Pr[\boldsymbol{S} = 4] + \cdots + 18 \cdot \Pr[\boldsymbol{S} = 18]$

lot's of arithmetic :-(

$= 10.5$

Most Useful Equality in Probability Theory:

Linearity of Expectation

Linearity of Expectation

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

(X and Y need not be independent!)

($\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$ not always true!)

Linearity of Expectation

Example

Let X = RandInt(6), Y = RandInt(6), Z = RandInt(6)

Let $S = X + Y + Z$

$$
\begin{aligned}
\mathbf{E}[S] &= \mathbf{E}[X + Y + Z] \\
&= \mathbf{E}[X] + \mathbf{E}[Y + Z] \\
&= \mathbf{E}[X] + \mathbf{E}[Y] + \mathbf{E}[Z] \\
&= 3.5 + 3.5 + 3.5 \\
&= 10.5
\end{aligned}
$$

Most Useful Type of Random Variable:

Indicator Random Variable

Indicator Random Variable

Event —> Random Variable

Let A be an event.

The indicator r.v. for A is:

$$\boldsymbol{I}_A(\ell) = \begin{cases} 1 & \text{if } \ell \in A \\ 0 & \text{if } \ell \notin A \end{cases}$$

$\boldsymbol{I}_A$ is 1 if A happens

$\boldsymbol{I}_A$ is 0 if A does not happen

$$\Pr[\boldsymbol{I}_A = 1] = \Pr[A]$$

$$\Pr[\boldsymbol{I}_A = 0] = 1 - \Pr[A]$$

$$\mathbf{E}[\boldsymbol{I}_A] =$$

$$0 \cdot \Pr[\boldsymbol{I}_A = 0] + 1 \cdot \Pr[\boldsymbol{I}_A = 1]$$

$$= \Pr[\boldsymbol{I}_A = 1]$$

$$= \Pr[A]$$

Most Useful Equality in Probability Theory:

Linearity of Expectation

Most Useful Type of Random Variable:

Indicator Random Variable

magic happens when you put them together.

High Level Idea

Want to compute $\mathbf{E}[\boldsymbol{X}]$:

Write $\boldsymbol{X} = \boldsymbol{I}_1 + \boldsymbol{I}_2 + \cdots + \boldsymbol{I}_n$. (sum of indicator r.v.'s)

Then
$$\begin{aligned}
\mathbf{E}[\boldsymbol{X}] &= \mathbf{E}[\boldsymbol{I}_1 + \boldsymbol{I}_2 + \cdots + \boldsymbol{I}_n] \\
&= \mathbf{E}[\boldsymbol{I}_1] + \mathbf{E}[\boldsymbol{I}_2] + \cdots + \mathbf{E}[\boldsymbol{I}_n] \\
\text{(usually)} &= n \cdot \mathbf{E}[\boldsymbol{I}_1] \\
&= n \cdot \Pr[\boldsymbol{I}_1 = 1]
\end{aligned}$$

Awesome!

(probability that the corresponding event happens)

Example

There are 150 students in 15-251 this semester.

After Midterm 2, we randomly permute the midterms before handing them back.

$\boldsymbol{X}$ = number of students who get their own midterm back.

What is $\mathbf{E}[\boldsymbol{X}]$?

Most Common 3 Random Variables

Bernoulli Random Variable

Introducing via Probability Mass Function (PMF)

$\boldsymbol{X} \sim \text{Bernoulli}(p)$ means:

"$\boldsymbol{X}$ is a Bernoulli random variable with success probability p."

$$\Pr[\boldsymbol{X} = 1] = p$$

$$\Pr[\boldsymbol{X} = 0] = 1 - p$$

So $\text{range}(\boldsymbol{X}) = \{0, 1\}$

<u>Check:</u>

$$\mathbf{E}[\boldsymbol{X}] = p$$

Binomial Random Variable

Introducing via other random variables

$\boldsymbol{X} \sim \text{Binomial}(n, p)$ **means:**

$$\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + \cdots + \boldsymbol{X}_n$$

where $\boldsymbol{X}_i \sim \text{Bernoulli}(p)$ for all $i \in \{1, 2, \ldots, n\}$,

and the $\boldsymbol{X}_i$'s are independent.

So $\text{range}(\boldsymbol{X}) = \{0, 1, 2, \ldots, n\}$

<u>Check:</u>

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

$$\mathbf{E}[\boldsymbol{X}] = np$$

Geometric Random Variable

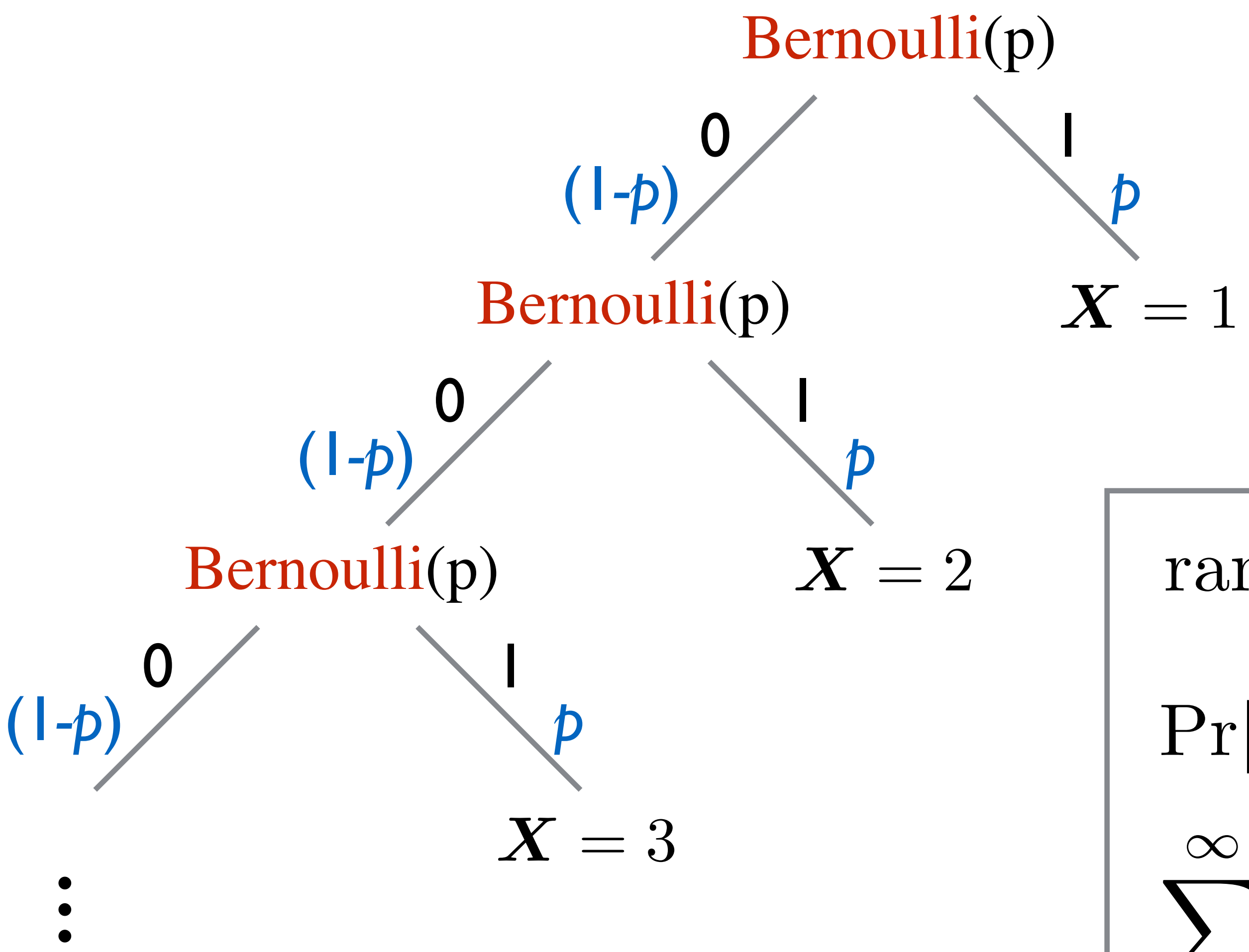
Introducing via code

$\boldsymbol{X} \sim \text{Geometric}(p)$ **means:**

```
X <— 1
while Bernoulli(p) = 0:
   X <— X+1
```

"number of p-biased coin flips until we see **H** for the first time."

Geometric Random Variable



$$\text{range}(\boldsymbol{X}) = \{1, 2, 3, \ldots\}$$

$$\Pr[\boldsymbol{X} = i] = (1-p)^{i-1} p$$

$$\sum_{i=1}^{\infty} \Pr[\boldsymbol{X} = i] = 1$$

(geometric sum)

$$\mathbf{E}[\boldsymbol{X}] = 1/p$$

New Topic:

Randomized Algorithms

Randomness and algorithms

How can randomness be used in computation?

Given some algorithm that solves a problem:

(i) the input can be chosen randomly
average-case analysis

(ii) the algorithm can make random choices
randomized algorithm

Which one will we focus on?

Randomness and algorithms

What is a randomized algorithm?

A *randomized algorithm* is an algorithm that is allowed to *flip a coin* (i.e., has access to random bits).

In 15-251:

A randomized algorithm is an algorithm that is allowed to call:

- RandInt(n)
- Bernoulli(p)

(we'll assume these take $O(1)$ time)

Deterministic vs Randomized

Deterministic

```
def A(x):
    y = 1
    if(y == 0):
        while(x > 0):
            x = x - 1
    return x+y
```

Randomized

```
def A(x):
    y = Bernoulli(0.5)
    if(y == 0):
        while(x > 0):
            x = x - 1
    return x+y
```

For any fixed input (e.g. x = 3):

- the **output** is *invariant*
- the **running time** is *invariant*

- the **output** *can vary*
- the **running time** *can vary*

Deterministic vs Randomized

A **deterministic algorithm** A computes $f : \Sigma^* \to \Sigma^*$ in time $T(n)$ means:

- **correctness**: $\forall x \in \Sigma^*, \quad A(x) = f(x)$.
- **running time**: $\forall x \in \Sigma^*, \quad \#\text{ steps } A(x) \text{ takes is } \leq T(|x|)$.

Note: we require **worst-case** guarantees for **correctness** and **run-time**.

Deterministic vs Randomized

A **randomized algorithm** A computes $f : \Sigma^* \to \Sigma^*$ in time $T(n)$ means:

- **correctness**: $\forall x \in \Sigma^*$, ?
- **running time**: $\forall x \in \Sigma^*$, ?

Deterministic vs Randomized

A Try

A **randomized algorithm** A computes $f : \Sigma^* \to \Sigma^*$ in time $T(n)$ means:

- **correctness**: $\forall x \in \Sigma^*$, $\boxed{A(x)} = f(x)$.
- **running time**: $\forall x \in \Sigma^*$, $\boxed{\#\text{ steps } A(x) \text{ takes}}$ is $\leq T(|x|)$.

these are random

Deterministic vs Randomized

A Try

A **randomized algorithm** A computes $f : \Sigma^* \to \Sigma^*$ in time $T(n)$ means:

- **correctness**: $\forall x \in \Sigma^*$, $\mathbf{Pr}[A(x) = f(x)] = 1$.
- **running time**: $\forall x \in \Sigma^*$,

$$\mathbf{Pr}[\# \text{ steps } A(x) \text{ takes is } \leq T(|x|)] = 1 \; .$$

Is this interesting? No.

A randomized algorithm should gamble with either **correctness** or **run-time**.

$$\forall x \in \Sigma^*$$

		Correctness	Run-time
Deterministic		always	always $\leq T(n)$
Randomized	Type 0	always	always $\leq T(n)$
	Type 1	w.h.p.	always $\leq T(n)$
	Type 2	always	w.h.p. $\leq T(n)$
	Type 3	w.h.p.	w.h.p. $\leq T(n)$

Type 0: may as well be deterministic

Type 1: "Monte Carlo algorithm"

Type 2: "Las Vegas algorithm"

Type 3: Can be converted to type 1. (exercise)

Example

Input: An array B with n/4 1's and 3n/4 0's.

Output: An index that contains a 1.

Deterministic

```
for i = 0 to n-1:
    if B[i] = 1:
        return i
```

correct: always

run-time: always $O(n)$

Randomized

Type 1 (Monte Carlo)

```
repeat 500 times:
    i = RandInt(n)
    if B[i] = 1:
        return i
return "Failed"
```

correct: w.h.p.

run-time: always $O(1)$

Type 2 (Las Vegas)

```
repeat:
    i = RandInt(n)
    if B[i] = 1:
        return i
```

correct: always

run-time: w.h.p. $O(1)$

Example

Input: An array B with n/4 1's and 3n/4 0's.
Output: An index that contains a 1.

	Correctness	Run-time
Deterministic	always	always $O(n)$
Monte Carlo	w.h.p.	always $O(1)$
Las Vegas	always	w.h.p. $O(1)$