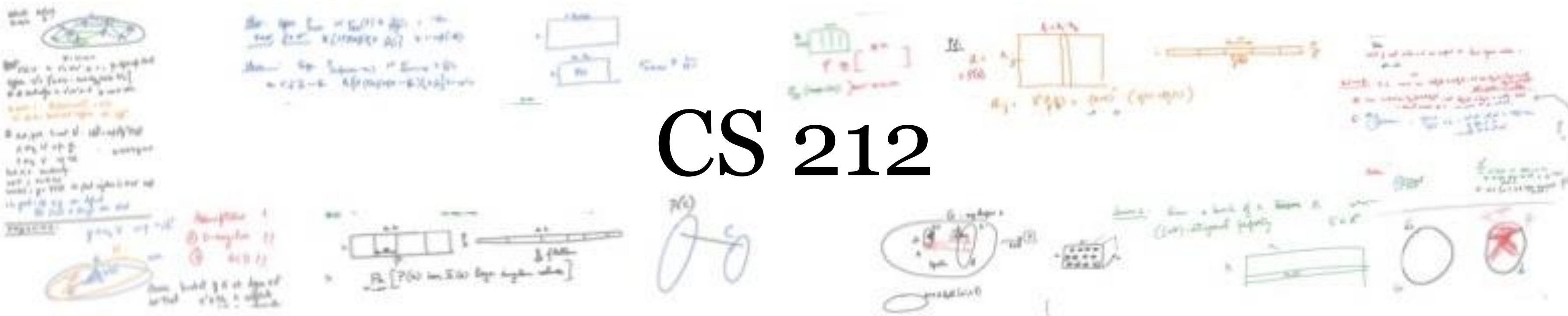


CS 212

Mathematical Foundations of Computer Science

Lecture 14: Random Variables

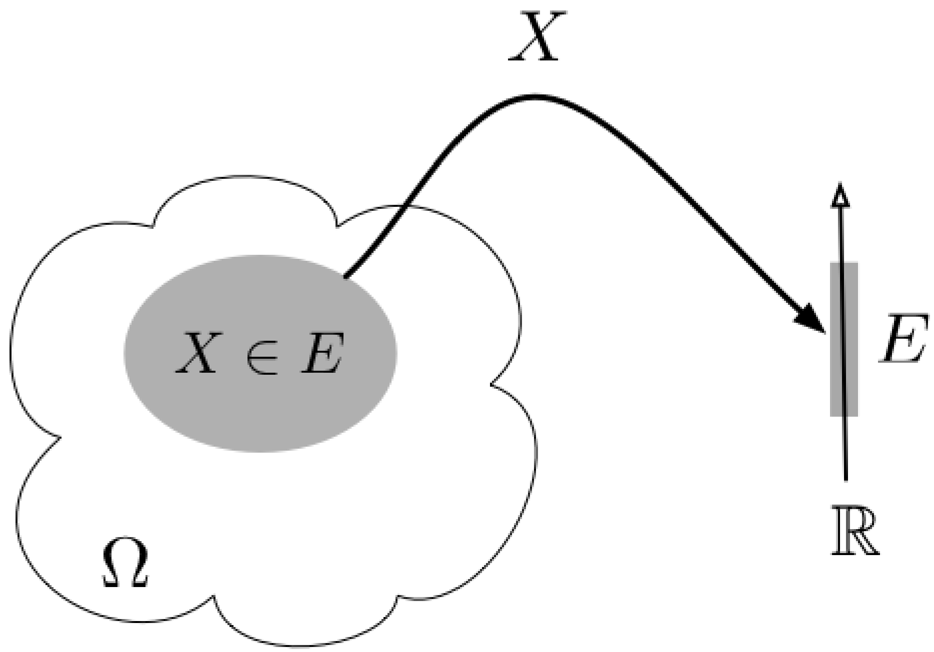
Announcements

- Midterm on Wed, October 26th in class
- Not open book. One "cheat" sheet i.e., two sides where you can write down anything.
- Midterm portions: everything up to what is covered today.
- PS4 is out. Due on Tuesday as usual.

Questions we'd like to answer

- What is the typical number of collisions in a hash table?
- How do I guess the average height of a class without going over each student's value?
- Average running time of a randomized algorithm

Pretty messy with direct counting…

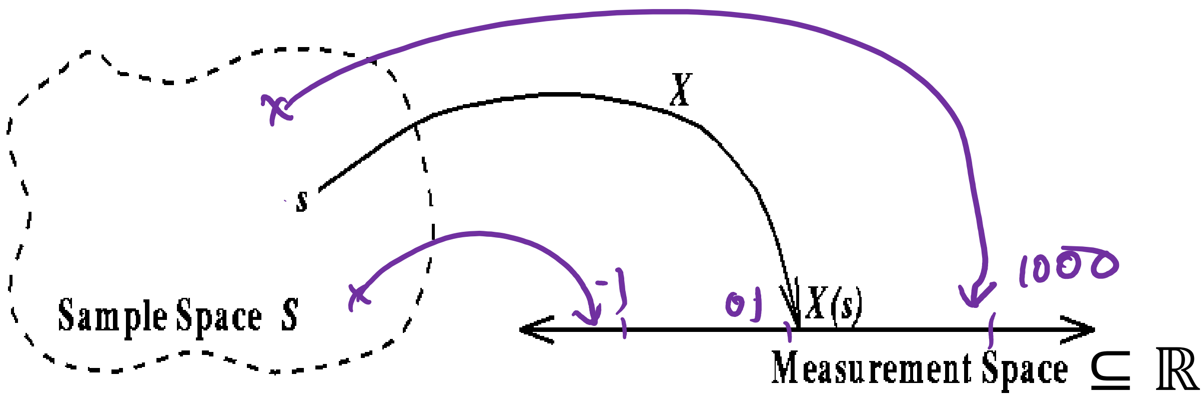


Random Variables, Expectations

Random Variables (r.v.)

Let S be sample space in a probability distribution

A Random Variable is a real-valued function on S



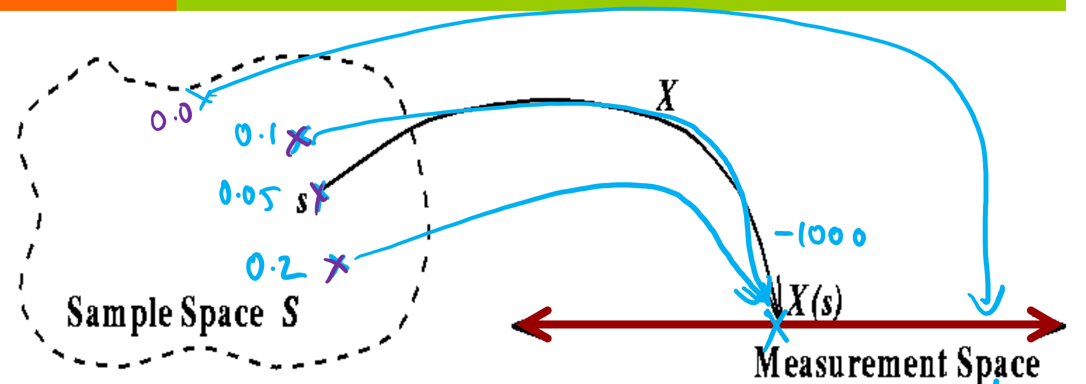
Examples:

X = value of 1st die in a two-dice roll X(3,4) = 3, X(1,6) = 1

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

Two views of Random Variables

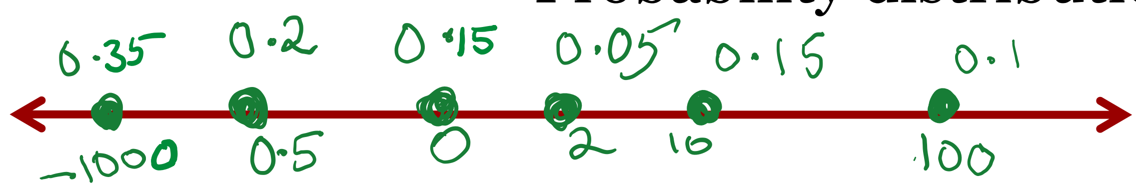
1. Think of a R.V. as function from S to the reals ℝ (input to the function is random)



prob of 0.35 assigned to
X = -1000

2. Or think of the induced distribution on ℝ, randomness is "pushed" to the values of the function.

Probability distribution on ℝ



Measurement space ℝ

Expectation

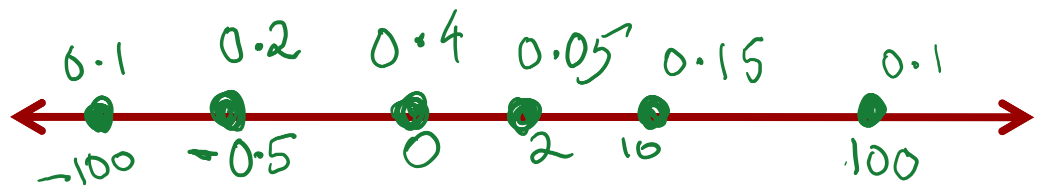
Average/mean value of the random variable X

The expectation, or expected value of a random variable X is

$$\mathbb{E}[X] = \sum_{t \in S} Pr(t) \times \underline{X(t)} = \sum_{k \in \mathbb{R}} k \times \Pr[X = k]$$

Sample space

X has a prob. distribution on its values



Measurement space $\mathbb{R}$

Examples of Random Variables

Use letters like A, B, E for events

R.V. = random variable

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

0/1 R.V. or Bernoulli R.V. or Indicator R.V: X takes value 0 or 1.

E.g., Coin toss: X=1 when Heads, 0 when Tails

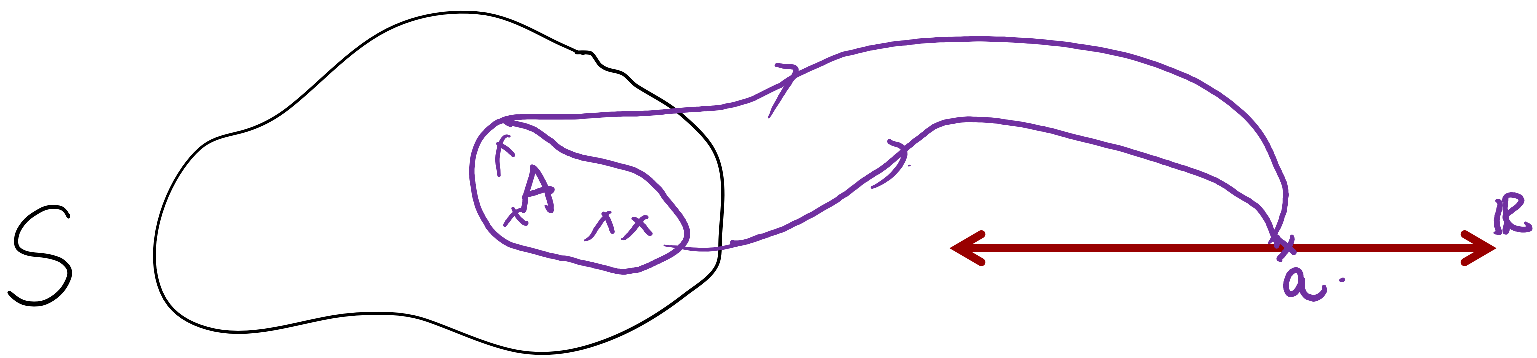
$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

$$\mathbb{E}[X] = \sum_{k \in \mathbb{R}} \Pr[X=k] \times k = 1 \times p + 0 \times (1-p) = p$$

If Y is the number when fair die is rolled,

$$\mathbb{E}[Y] = \frac{1}{6} * 1 + \frac{1}{6} * 2 + \frac{1}{6} * 3 + \cdots + \frac{1}{6} * 6$$
$$= 3.5$$

From R.V. to Events



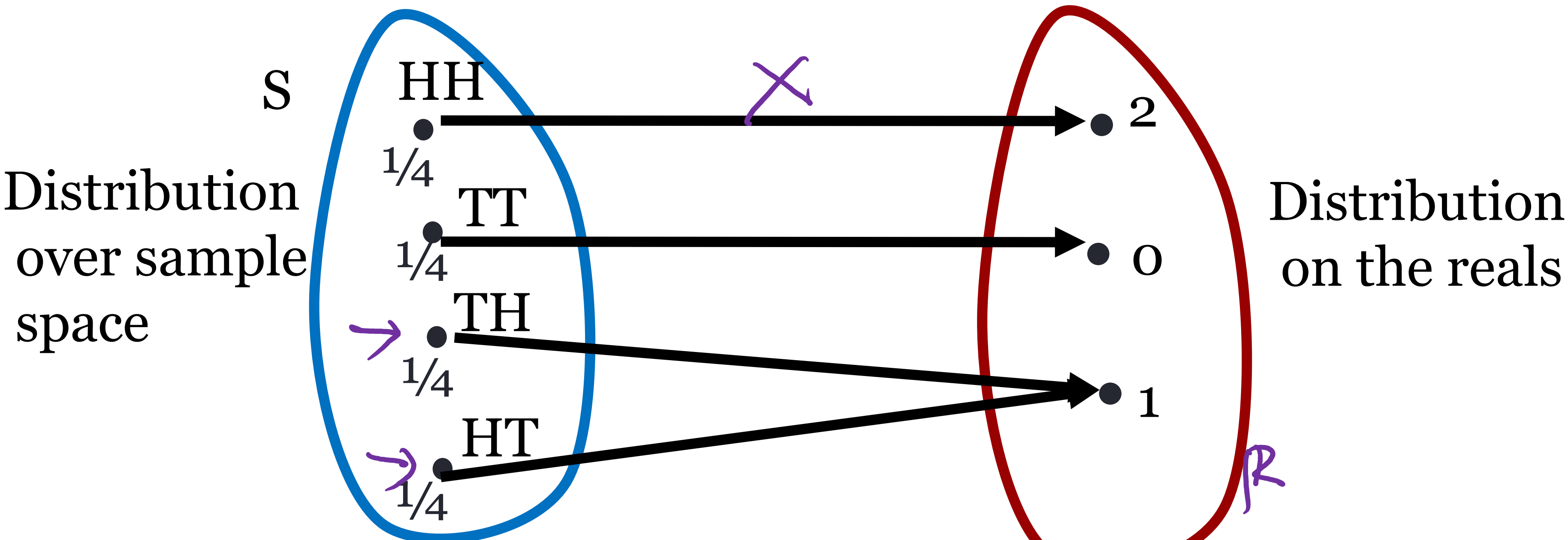
For any random variable X and value a,
we can define the event A that X = a

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

Coin Tosses

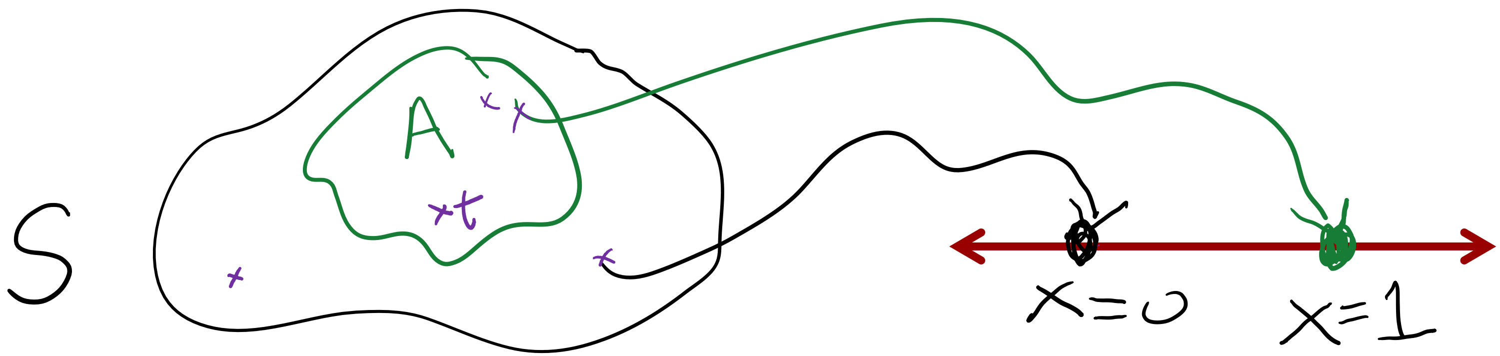
Two coin tosses. Sample space S={TT, TH, HT, HH}

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



Qn: What is Pr[X=1]? $= \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$ $\Pr[X=100] = 0$

From Events to RVs



For any event A, can define the indicator random variable for A: 0/1 - RVs.

$$X_A(t) = \begin{cases} 1 & \text{if } t \in A \\ 0 & \text{if } t \notin A \end{cases}$$

$t \in$ Sample space S

Expectation $\mathbb{E}[X_A] = 1 * \Pr(X_A = 1) + 0 * \Pr(X_A = 0) = \Pr(A)$

Conditional Expectation

Expectation: $\mathbb{E}[X] = \sum_{t \in S} Pr(t) \times X(t) = \underline{\sum_{k \in \mathbb{R}} k \times \Pr[X = k]}$

Conditional expectation: Average value of the random variable X conditioned on event A

$$\mathbb{E}[\underset{\text{r.v.}}{X} | \underset{\text{event}}{A}] = \sum_{k \in \mathbb{R}} k \times \Pr[\underbrace{X = k}_{\text{event}} \mid A] = \sum_{t \in A} \frac{Pr(t)}{\Pr[A]} \times X(t)$$

Thm. Given any random variable X, event A

$$\mathbb{E}[X] = \mathbb{E}[X|A]\Pr[A] + \mathbb{E}\left[X\middle|\bar{A}\right]\Pr[\bar{A}]$$

Linearity of Expectation

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

Thm. If X and Y are random variables , and if Z = X+Y, then

$$\mathbb{E}[Z] = \mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

**even if X and Y are not independent !*

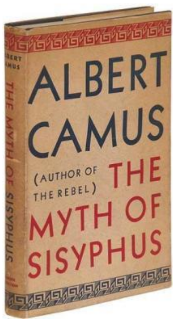
Thm. If $X_1, X_2, \ldots, X_n$ are R.Vs , if Z $= \overbrace{a_1X_1 + a_2X_2 + \cdots + a_nX_n}^{\text{linear function}}$

$$\mathbb{E}[Z] = a_1\mathbb{E}[X_1] + a_2\mathbb{E}[X_2] + \cdots + a_n\mathbb{E}[X_n]$$ $\forall\, a_1, a_2, a_3, \ldots, a_n \in \mathbb{R}$

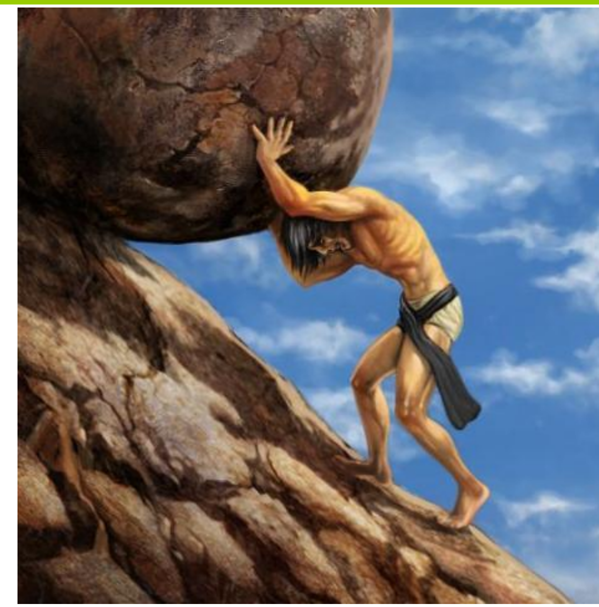
e.g. Two rolled dice. X_1= number on first die. X_2= number on 2nd die.
X=sum of numbers of two dice= $X_1 + X_2$.

$$\mathbb{E}[X_1 + X_2] = \mathbb{E}[X_1] + \mathbb{E}[X_2] = 3.5 + 3.5 = 7$$

Geometric Random Variables



Sisyphus tries to roll a rock up a hill everyday. Every day he tries, he succeeds with probability $p = \frac{1}{100}$ (success independent of other days). He stops when he succeeds. What is his expected number of attempts ?



Let X be the R.V. representing #attempts

Pr[X=1] = p ? Pr[X=2] = $(1-p)p$? Pr[X=k]= $\underline{(1-p)^{k-1} \times p}$?

$$\mathbb{E}[X] = \Pr[X=1] \times 1 + \Pr[X=2] \times 2 + \cdots + \Pr[X=k] \times k + \cdots$$

$$\mathbb{E}[X] = \underbrace{p + (1-p)p \times 2 + \cdots + \underline{(1-p)^{k-1}p} \times k + \cdots}_{\text{Arithmetico-Geometric Series}} = \frac{1}{p} = 100$$

$a + ar + ar^2 + ar^3 + \cdots + ar^{k-1} + \cdots = \frac{a}{1-r}$

$\forall a$
$|r| < 1$

Expectation of Geometric RV

$\Pr[X \geq k] = (1-p)^{k-1}$

$\underbrace{\Pr[X = k] = (1-p)^{k-1} p}$

Let indicator r.v. Y_i =cost (time) spent on day i

(Y_i: did Sisyphys *try* to climb up on day i)?

Number of days = X $= Y_1 + Y_2 + Y_3 + \cdots + Y_k + \cdots$

Ans= $E[X] = E[Y_1] + E[Y_2] + E[Y_3] + \cdots + E[Y_k] + \cdots$

$E[Y_1] =$ 1 ? $E[Y_2] =$ $(1-p) = \Pr[X \geq 2]$?

$E[Y_k] = E[Y_k | X < k]\Pr[X < k] + E[Y_k | X \geq k]\Pr[X \geq k]$

$=$ $1 \times (1-p)^{k-1} + 0 \times (\ )$? $= (1-p)^{k-1}$

$a=1$ $r=1-p$

$E[X] = 1 + (1-p) + \cdots + (1-p)^{k-1} + \cdots = \frac{1}{1-(1-p)} = \frac{1}{p}$ $=100$

Independent R.V.s

If events A and event B are independent:

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

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

$\forall a,b \in \mathbb{R}$ $\quad \Pr[X=a \wedge Y=b] = \Pr[X=a] \times \Pr[X=b]$

Thm: If random variables X and Y are independent:

(converse is not necessarily true)

$$\mathbb{E}[X \cdot Y] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

Aside: $\mathrm{Cov}(X, Y) = \mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y]$ is measure of dependence