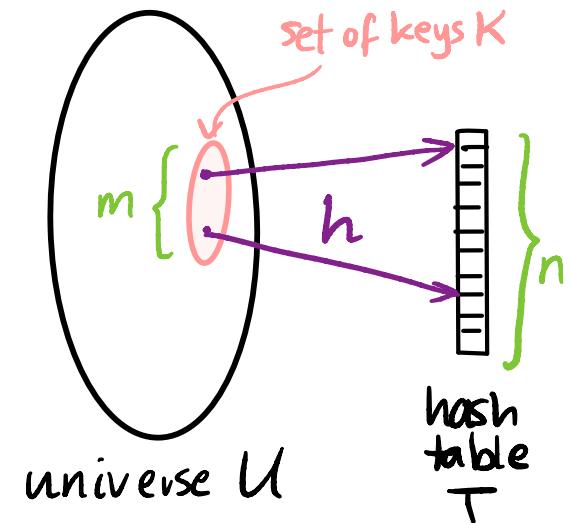


CS70 – Spring 2024

Lecture 18 – March 19

Summary of Last Lecture - all based on balls & bins

- For a random hash function, to avoid collisions (with good prob.) need size of hash table to be
 $\approx (\text{no. of keys stored})^2$



- To collect at least one copy of each of n coupons, need to take about $n \ln n$ random samples
- If we randomly distribute n jobs among n processors, the largest load on any processor is likely to be around $\frac{\ln n}{\ln \ln n}$

Ideas/techniques
are more
important than
calculations!

Today

- Random variables (=functions/measurements on probability spaces)
- Distributions
- Expectation
- The Unreasonable Power of Linearity of Expectation

Random Variables

Measurements on probability spaces

Example: $\Omega = \text{space of 3 fair coin tosses}$

$$= \{\text{HHH}, \text{HHT}, \text{HTH}, \text{THH}, \text{HTT}, \text{THT}, \text{TTH}, \text{TTT}\}$$

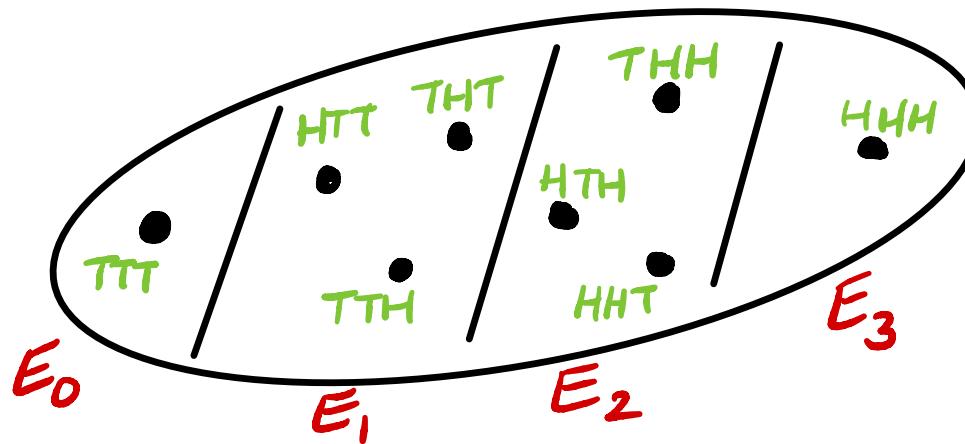
Uniform probabilities: $\Pr[\omega] = \frac{1}{8} \quad \forall \omega \in \Omega$

For any $\omega \in \Omega$, let $X(\omega) := \text{number of Heads in } \omega$

- Note:**
- $X(\omega)$ is a (real) number (actually, a non-neg. integer)
 - $X(\omega) \in \{0, 1, 2, 3\}$

For any $\omega \in \Omega$, let $X(\omega) := \text{number of Heads in } \omega$

- For any $i \in \{0, 1, 2, 3\}$, $E_i := \{\omega : X(\omega) = i\}$ is an event
- The events $\{E_i\}$ partition Ω



The collection $\{(i, \Pr[E_i])\}$ is the distribution of X

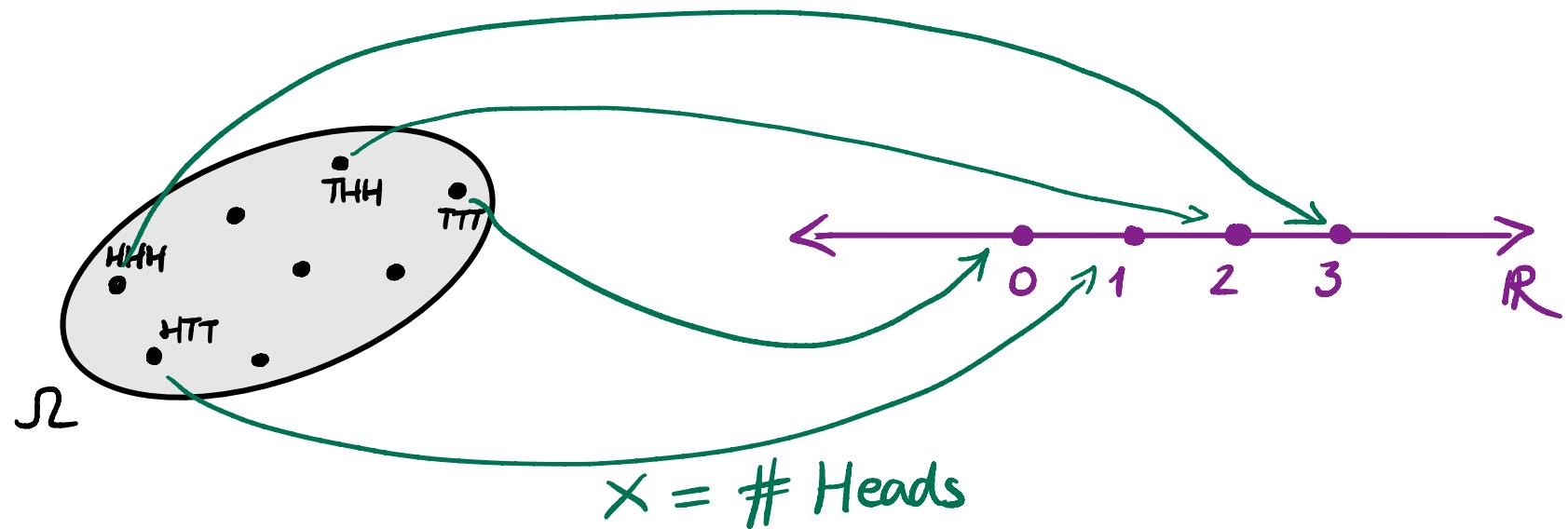
Here:

$$\begin{aligned}\Pr[X=0] &= 1/8 \\ \Pr[X=1] &= 3/8 \\ \Pr[X=2] &= 3/8 \\ \Pr[X=3] &= 1/8\end{aligned}$$

Random Variable : Definition

Defn : A random variable on a prob. space Ω is a function $X: \Omega \rightarrow \mathbb{R}$

I.e., X assigns a real value $X(\omega)$ to each $\omega \in \Omega$



Defn : The distribution of a (discrete) r.v. X is :

- the set of possible values for X
- for each possible value a the probability $\Pr[X=a]$

Check: For any r.v. X with set of possible values \mathcal{A} , we have

$$\sum_{a \in \mathcal{A}} \Pr[X=a] = 1$$

Proof : $\sum_{a \in \mathcal{A}} \Pr[X=a] = \sum_{a \in \mathcal{A}} \left(\sum_{\omega: X(\omega)=a} \Pr[\omega] \right)$

$$= \sum_{\omega \in \Omega} \Pr[\omega] = 1$$

since $\forall \omega \in \Omega$
 \exists unique $a \in \mathcal{A}$
s.t. $X(\omega) = a$

□

Examples

1. Roll 2 fair dice

$X = \text{sum of scores on dice}$

$$\Omega = \{(i,j) : 1 \leq i, j \leq 6\} \quad |\Omega| = 36$$

$$X(i,j) = i+j$$

$$X(\omega) \in \{2, 3, \dots, 11, 12\}$$



Distribution of X

$$\Pr[X=2] = 1/36$$

$$\Pr[X=3] = 2/36 = 1/18$$

$$\Pr[X=4] = 3/36 = 1/12$$

:

$$\Pr[X=7] = 6/36 = 1/6$$

:

$$\Pr[X=12] = 1/36$$

6
5
4
3
2
1
1	2	3	4	5	6	

2. Random Permutations

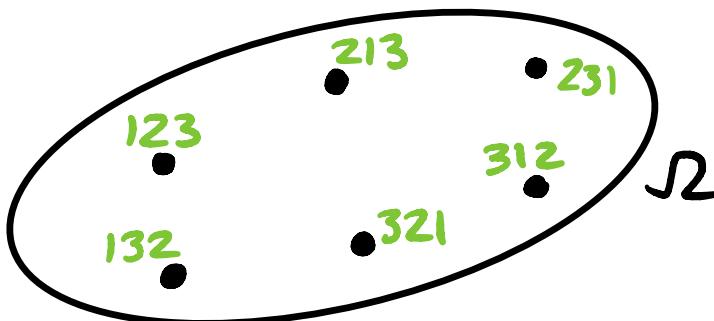
Collect the IDs of n students

Redistribute them randomly (one per student)

$\Omega = \text{set of } \underline{\text{permutations}} \text{ of } n \text{ items}$

$$|\Omega| = n!$$

E.g. for $n=3$

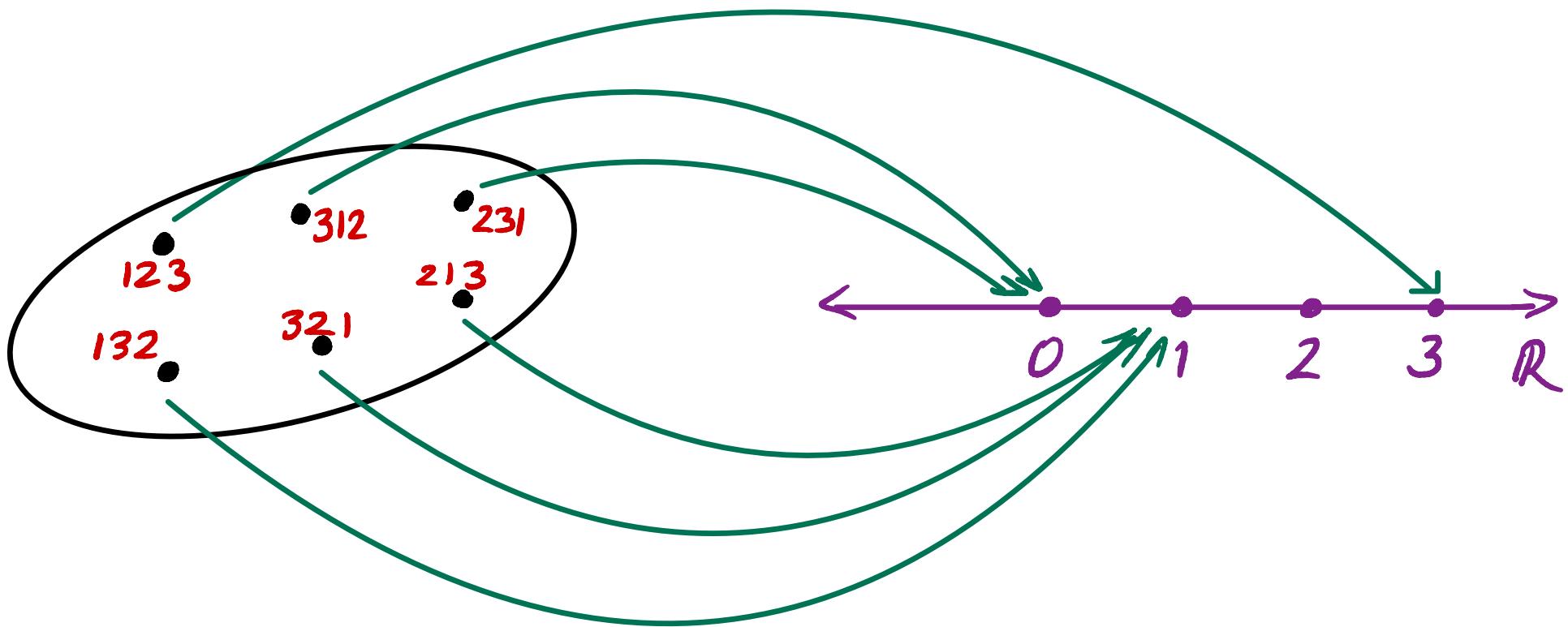


$$|\Omega|=3! = 6$$

Uniform probability space: $\Pr[\omega] = \frac{1}{n!} \quad \forall \omega$

Random variable $X = \text{no. of students who get their own ID}$
a.k.a. "fixed points"

X = no. of fixed points in a random permutation



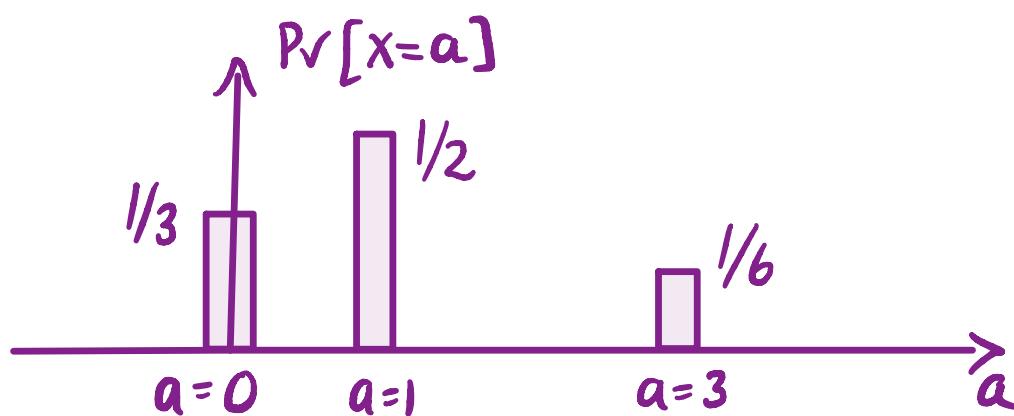
Distribution of X

$$\Pr[X=0] = 2/6 = 1/3$$

$$\Pr[X=1] = 3/6 = 1/2$$

$$\Pr[X=3] = 1/6 = 1/6$$

Histogram



3. Binomial Distribution

Toss n biased coins, each having Heads prob. p

$\mathcal{S} = \{H, T\}^n$ (= all strings of length n over alphabet $\{H, T\}$)

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

where $i = \text{no. of Heads}$
in ω

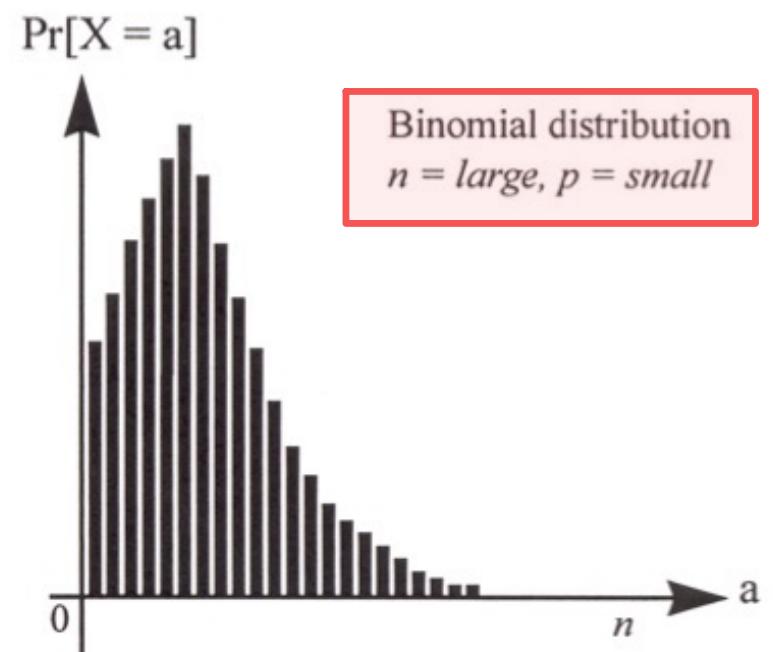
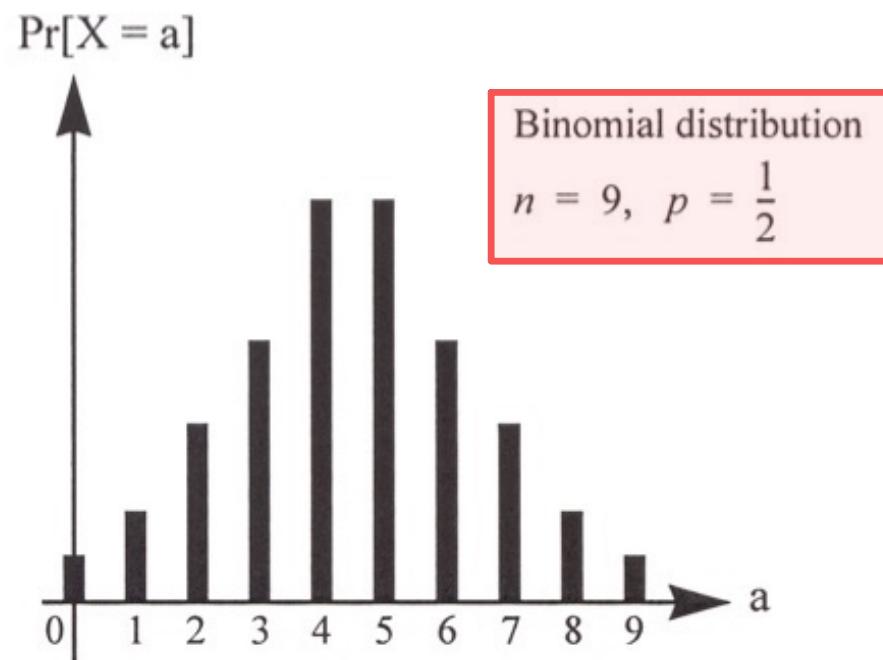
Random variable $X = \text{no. of Heads}$ $X \in \{0, 1, \dots, n\}$
What is the distribution of X ?

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

We say X has binomial distribution with parameters n, p

$$X \sim \text{Bin}(n, p)$$

Pictures of $\text{Bin}(n, p)$



4. Hypergeometric Distribution

Deal a 5-card poker hand $|H2| = \binom{52}{5}$

R.V. $X = \text{no. of hearts in your hand}$

$$X \in \{0, 1, 2, 3, 4, 5\}$$

Distribution :

$$\Pr[X=0] = \frac{\binom{39}{5}}{\binom{52}{5}}$$

$$\Pr[X=1] = \frac{\binom{13}{1} \times \binom{39}{4}}{\binom{52}{5}}$$

$$\Pr[X=k] = \frac{\binom{13}{k} \binom{39}{5-k}}{\binom{52}{5}}$$

4. Hypergeometric Distribution

Deal a 5-card poker hand

$$|H2| = \binom{52}{5}$$

R.V. $X = \text{no. of hearts in your hand}$

$$X \in \{0, 1, 2, 3, 4, 5\}$$

Distribution:

$$\Pr[X=0] = \frac{\binom{39}{5}}{\binom{52}{5}}$$

$$\Pr[X=1] = \frac{\binom{13}{1} \times \binom{39}{4}}{\binom{52}{5}}$$

.

$$\Pr[X=k] = \frac{\binom{13}{k} \binom{39}{5-k}}{\binom{52}{5}}$$

Note: $\sum_{k=0}^n \binom{B}{k} \binom{N-B}{n-k} = \binom{N}{n}$!

More generally:

- box of N balls, B black, rest white
- draw n balls w.o. replacement
- $X = \# \text{ of black balls drawn}$

$$\Pr[X=k] = \frac{\binom{B}{k} \binom{N-B}{n-k}}{\binom{N}{n}}$$

Hypergeometric distribution,
parameters (N, B, n)

Joint Distributions

Defn: The joint distribution of two r.v.'s X, Y on the same prob. space is the set

$$\{(a, b, \Pr[X=a, Y=b]) : a \in A, b \in B\}$$

where A, B are the possible values of X, Y resp.

The marginal distribution of X is given by

$$\Pr[X=a] = \sum_{b \in B} \Pr[X=a, Y=b]$$

X, Y are independent if

$$\Pr[X=a, Y=b] = \Pr[X=a] \times \Pr[Y=b] \quad \forall a, b$$

Joint Distributions

Example : Throw two fair dice

Random variables :

X = score on first die

Y = —— second ——

Z = sum of scores

$$\Pr[X=3, Y=5] = 1/36$$

$$\Pr[X=3, Z=9] = 1/36$$

X, Y independent ?

X, Z independent ?

Expectation (= mean/average)

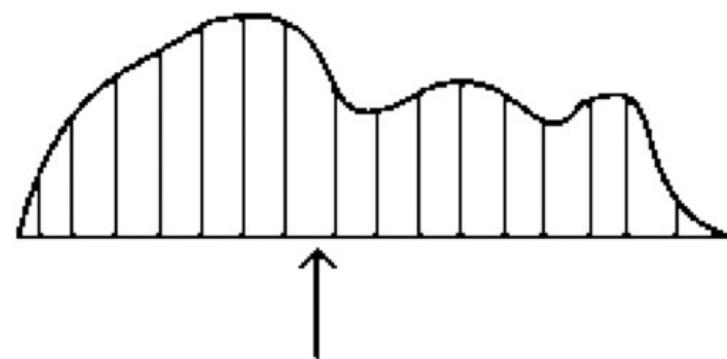
Simplest quantity that summarises the distribution of a r.v.

Defn: The expectation of a (discrete) r.v. X is

$$E[X] := \sum_{a \in A} a \times \Pr[X=a]$$

where A is the set of possible values of X

$E[X]$ measures the "center of mass" of the distribution



Expectation: Examples

1. $X = \text{score on one fair die}$

$$\begin{aligned} E[X] &= \left(\frac{1}{6} \times 1\right) + \left(\frac{1}{6} \times 2\right) + \dots + \left(\frac{1}{6} \times 6\right) \\ &= \frac{1}{6} \times (1+2+\dots+6) = \boxed{\frac{7}{2}} \end{aligned}$$

1'. $Y = \text{sum of scores on two fair dice}$

$$E[Y] = \left(\frac{1}{36} \times 2\right) + \left(\frac{2}{36} \times 3\right) + \left(\frac{3}{36} \times 4\right) +$$

$$\dots + \left(\frac{1}{36} \times 12\right)$$

$$= \dots$$

$$= \boxed{7}$$

6
5
4
3
2
1
1	2	3	4	5	6		

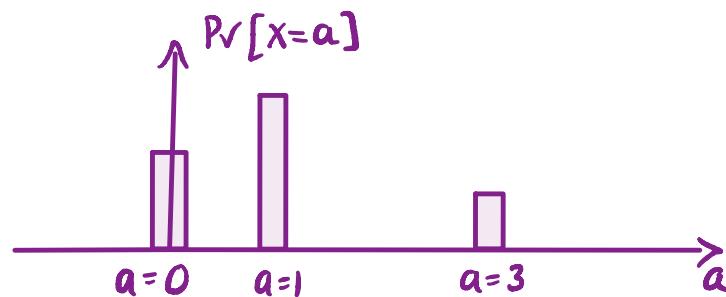
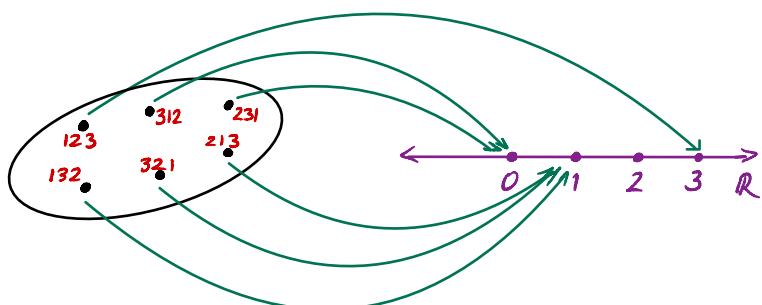
2. X = no. of fixed points in a random permutation
($n=3$)

Distribution of X

$$\Pr[X=0] = 2/6 = 1/3$$

$$\Pr[X=1] = 3/6 = 1/2$$

$$\Pr[X=3] = 1/6 = 1/6$$



$$E[X] = \left(\frac{1}{3} \times 0\right) + \left(\frac{1}{2} \times 1\right) + \left(\frac{1}{6} \times 3\right) = 0 + \frac{1}{2} + \frac{1}{2} = \boxed{1}$$

3. Roulette

Roulette wheel : 36 numbers
(18 black / 18 red) plus 0, 00

Bet \$1 on red : win \$1 if red,
lose \$1 if black or green

X = amount won lost

$$\left. \begin{array}{l} \Pr[X=+1] = \frac{18}{38} \\ \Pr[X=-1] = \frac{20}{38} \end{array} \right\} E[X] = \left(1 \times \frac{18}{38}\right) + \left(-1 \times \frac{20}{38}\right) = -\frac{1}{19}$$



Note: presence of 0, 00 make this game unfair

Linearity of Expectation

Thm: For any random variables X, Y on prob. space Ω ,

$$(i) \quad E[X+Y] = E[X] + E[Y]$$

$$(ii) \quad E[aX] = aE[X] \quad \text{for constant } a$$

Proof: Note that $E[X] = \sum_{\omega \in \Omega} X(\omega) \times \Pr[\omega]$

$$\begin{aligned} (i) \quad E[X+Y] &= \sum_{\omega \in \Omega} (X+Y)(\omega) \times \Pr[\omega] \\ &= \sum_{\omega \in \Omega} X(\omega) \times \Pr[\omega] + \sum_{\omega \in \Omega} Y(\omega) \times \Pr[\omega] \end{aligned}$$

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

(ii) Easy exercise

Crucial: Does not assume X, Y are independent!!!

Does not say that $E[XY] = E[X]E[Y]$ or $E[1/X] = 1/E[X]$

Linearity of Expectation : Examples

1. Two fair dice

$X = \text{sum of dice rolls}$

Then $\underbrace{X = X_1 + X_2}_{\text{where } X_1 = \text{score on first die}}$ $X_2 = \text{--- - second ---}$

$$\forall \omega \in \Omega: X(\omega) = X_1(\omega) + X_2(\omega)$$

So by linearity

$$E[X] = E[X_1 + X_2] = E[X_1] + E[X_2] = \frac{7}{2} + \frac{7}{2} = \boxed{7}$$

Linearity of Expectation : Examples

2. Multiple roulette games

Play roulette 100 times (\$1 stake each time)

X = amount won/lost

$X = X_1 + X_2 + \dots + X_{100}$ where X_i = amt. won/lost in i th game

Recall: $E[X_i] = -\frac{1}{19} \quad \forall i$

Linearity: $E[X] = \sum_{i=1}^{100} E[X_i] = 100 \times \left(-\frac{1}{19}\right) \approx -5.26$

3. Multiple coin tosses

Toss a biased coin (Heads prob. p) n times

$$X = \# \text{Heads}$$

$$X \sim \text{Binomial}(n, p)$$

$$X = X_1 + \dots + X_n$$

where $X_i = \begin{cases} 1 & \text{if } i\text{th toss Heads} \\ 0 & \dots \dots \text{Tails} \end{cases}$

Note that $E(X_i) = (1 \times p) + (0 \times (1-p)) = p$

Linearity: $E[X] = \sum_{i=1}^n E[X_i] = \boxed{n \times p}$

indicator
r. v.

expectation of
a $\text{Bin}(n, p)$
distribution

4. Balls & Bins

Recall: toss m balls u.a.r. into n bins

R.V. $X = \# \text{ of empty bins}$

$$X = \sum_{i=1}^n X_i \quad \text{where } X_i = \begin{cases} 1 & \text{if bin } i \text{ empty} \\ 0 & \dots \text{not empty} \end{cases}$$

$$\begin{aligned} \text{Then } E[X_i] &= (1 \times \Pr[\text{bin } i \text{ empty}]) + (0 \times \Pr[\text{bin } i \text{ not empty}]) \\ &= \Pr[\text{bin } i \text{ empty}] \\ &= \left(1 - \frac{1}{n}\right)^m \quad \text{all } m \text{ balls must choose a different bin} \end{aligned}$$

Linearity:

$$E[X] = \sum_{i=1}^n E[X_i] = n \left(1 - \frac{1}{n}\right)^m \approx n e^{-m/n}$$

$$\text{E.g. if } m=n, E[X] \approx n e^{-1} \approx 0.37n$$

5. Fixed points in a random permutation

General case: n items

$X_n = \# \text{ of fixed points}$ [Recall: $E[X_3] = 1$]

$X = \sum_{i=1}^n X_i$ where $X_i = \begin{cases} 1 & \text{if } i \text{ is a fixed point} \\ 0 & \text{otherwise} \end{cases}$

$E[X_i] = \Pr[i \text{ is a fixed point}] = 1/n$

Linearity: $E[X] = \sum_{i=1}^n E[X_i] = n \times \frac{1}{n} = \boxed{1}$

Bottom line: If we collect & redistribute IDs of n people, the expected # who get their own ID is always 1 (indep. of n)