CSE 312

Foundations of Computing II

Lecture 7: More on independence; start random variables

slido.com/4171468

for polls and anonymous questions

Agenda

- Recap
- Independence As An Assumption
- Conditional Independence
- New Topic: Random Variables

Bayes Theorem with Law of Total Probability

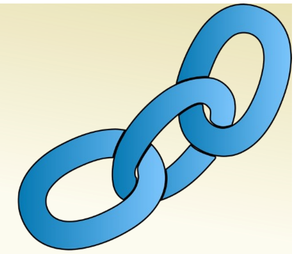
Bayes Theorem with LTP: Let $E_1, E_2, \dots, E_n$ be a partition of the sample space, and F, G events. Then,

$$P(G|F) = \frac{P(F|G)P(G)}{P(F)} = \frac{P(F|G)P(G)}{\sum_{i=1}^{n} P(F|E_i)P(E_i)}$$

Simple Partition: In particular, if E is an event with non-zero probability, then

$$P(G|F) = \frac{P(F|G)P(G)}{P(F|E)P(E) + P(F|E^C)P(E^C)}$$

Chain Rule



$$\mathbb{P}(\mathcal{B}|\mathcal{A}) = \frac{\mathbb{P}(\mathcal{A} \cap \mathcal{B})}{\mathbb{P}(\mathcal{A})} \Rightarrow \mathbb{P}(\mathcal{A})\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{A} \cap \mathcal{B})$$

Theorem. (Chain Rule) For events $\mathcal{A}_1, \mathcal{A}_2, \dots, \mathcal{A}_n$,

$$\mathbb{P}(\mathcal{A}_1 \cap \dots \cap \mathcal{A}_n) = \mathbb{P}(\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_2|\mathcal{A}_1) \cdot \mathbb{P}(\mathcal{A}_3|\mathcal{A}_1 \cap \mathcal{A}_2)$$

$$\cdots \mathbb{P}(\mathcal{A}_n|\mathcal{A}_1 \cap \mathcal{A}_2 \cap \dots \cap \mathcal{A}_{n-1})$$

An easy way to remember: We have n events and we can evaluate their probabilities sequentially, conditioning on the occurrence of previous events.

Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are (statistically) **independent** if

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A}) \cdot \mathbb{P}(\mathcal{B}).$$

Alternatively,

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(B)$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$

"The probability that $\mathcal{B}$ occurs after observing $\mathcal{A}$" -- Posterior
= "The probability that $\mathcal{B}$ occurs" -- Prior

Agenda

- Recap
- Independence As An Assumption
- Conditional Independence
- New Topic: Random Variables

Independence as an assumption

- People often assume it **without justification.**
- Example: A sky diver has two chutes

A : event that the main chute doesn't open $\qquad \mathbb{P}(A) = 0.02$

B : event that the backup doesn't open $\qquad \mathbb{P}(B) = 0.1$

- What is the chance that at least one opens assuming independence?

Independence as an assumption

- People often assume it **without justification.**
- Example: A sky diver has two chutes

A : event that the main chute doesn't open $\mathbb{P}(A) = 0.02$

B : event that the backup doesn't open $\mathbb{P}(B) = 0.1$

- What is the chance that at least one opens assuming independence?

Assuming independence doesn't justify the assumption! Both chutes could fail because of the same rare event e.g., freezing rain.

Corollaries of independence of two events

- Example: A sky diver has two chutes

  A : event that the main chute doesn't open $\mathbb{P}(A) = 0.02$
  B : event that the backup doesn't open $\mathbb{P}(B) = 0.1$

- What is the chance that **both open** assuming independence?

Agenda

- Recap
- Sometimes Independence Occurs for Nonobvious Reasons
- Independence As An Assumption
- Conditional Independence
- New Topic: Random Variables

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on $\mathcal{C}$ if $\mathbb{P}(\mathcal{C}) \neq 0$ and $\mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid \mathcal{C}) = \mathbb{P}(\mathcal{A} \mid \mathcal{C}) \cdot \mathbb{P}(\mathcal{B} \mid \mathcal{C})$.

Plain Independence. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** if

$$\mathbb{P}(\mathcal{A} \cap \mathcal{B}) = \mathbb{P}(\mathcal{A}) \cdot \mathbb{P}(\mathcal{B}).$$

Equivalence:

- If $\mathbb{P}(\mathcal{A}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B}|\mathcal{A}) = \mathbb{P}(\mathcal{B})$
- If $\mathbb{P}(\mathcal{B}) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A}|\mathcal{B}) = \mathbb{P}(\mathcal{A})$

Conditional Independence

Definition. Two events $\mathcal{A}$ and $\mathcal{B}$ are **independent** conditioned on C if $\mathbb{P}(C) \neq 0$ and $\mathbb{P}(\mathcal{A} \cap \mathcal{B} \mid C) = \mathbb{P}(\mathcal{A} \mid C) \cdot \mathbb{P}(\mathcal{B} \mid C)$.

Equivalence:

- If $\mathbb{P}(\mathcal{A} \cap C) \neq 0$, equivalent to $\mathbb{P}(\mathcal{B} | \mathcal{A} \cap C) = \mathbb{P}(B \mid C)$
- If $\mathbb{P}(\mathcal{B} \cap C) \neq 0$, equivalent to $\mathbb{P}(\mathcal{A} | \mathcal{B} \cap C) = \mathbb{P}(\mathcal{A} \mid C)$

Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin twice independently. What is the probability both tosses heads?

$$\Pr(HH) = \Pr(HH \mid C1)\Pr(C1) + \Pr(HH \mid C2)\Pr(C2) \qquad \text{LTP}$$

Example – More coin tossing

Suppose there is a coin C1 with Pr(Head) = 0.3 and a coin C2 with Pr(Head) = 0.9. We pick one randomly with equal probability and flip that coin 2 times independently. What is the probability we get all heads?

$$\begin{aligned}\Pr(HH) &= \Pr(HH \mid C1)\Pr(C1) + \Pr(HH \mid C2)\Pr(C2) \\ &= \Pr(H \mid C1)^2\Pr(C1) + \Pr(H \mid C2)^2\Pr(C2) \\ &= 0.3^2 \cdot 0.5 + 0.9^2 \cdot 0.5 = 0.45\end{aligned}$$

LTP

Conditional Independence

Is $\Pr(H_1 H_2) = 0.45 = \Pr(H_1)\Pr(H_2)$?

$$\Pr(H) = \Pr(H \mid C1)\Pr(C1) + \Pr(H \mid C2)\Pr(C2) = 0.6$$

$$\Pr(H_1)\Pr(H_2) = 0.36 \neq \Pr(H_1 H_2)$$

New topic: random variables

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)

Random Variables (Idea)

Often: We want to **capture quantitative properties** of the outcome of a random experiment, e.g.:

- *What is the total of two dice rolls?*
- *What is the number of coin tosses needed to see the first head?*
- *What is the number of heads among 5 coin tosses?*

Random Variables

Definition. A **random variable (RV)** for a probability space $(\Omega, \mathbb{P})$ is a function $X: \Omega \to \mathbb{R}$.

The set of values that X can take on is called its range/support Ω_X

Example. Number of heads in 2 independent coin flips $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

RV Example

20 balls labeled 1, 2, … , 20 in an urn

- Draw a subset of 3 uniformly at random
- Let $X = \text{maximum of the 3 numbers on the balls}$
  - Example: $X(2, 7, 5) = 7$
  - Example: $X(15, 3, 8) = 15$

RV Example

20 balls labeled 1, 2, …, 20 in an urn

- Draw a subset of 3 uniformly at random
- Let $X = \text{maximum of the 3 numbers on the balls}$
  - Example: $X(2, 7, 5) = 7$
  - Example: $X(15, 3, 8) = 15$
- What is $|\Omega_X|$?

Poll: **slido.com/4171468**

A. 20^3
B. 20
C. 18
D. $\binom{20}{3}$

Agenda

- Random Variables
- Probability Mass Function (pmf)
- Cumulative Distribution Function (CDF)

Example: Returning Homeworks

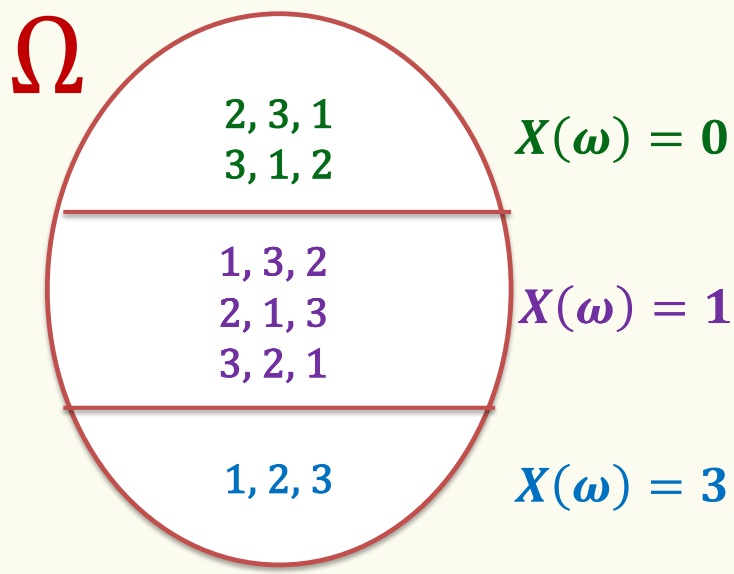
- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	
$1/6$	$1, 3, 2$	
$1/6$	$2, 1, 3$	
$1/6$	$2, 3, 1$	
$1/6$	$3, 1, 2$	
$1/6$	$3, 2, 1$	

Example: Returning Homeworks

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Pr(\omega)$	ω	$X(\omega)$
$1/6$	$1, 2, 3$	3
$1/6$	$1, 3, 2$	1
$1/6$	$2, 1, 3$	1
$1/6$	$2, 3, 1$	0
$1/6$	$3, 1, 2$	0
$1/6$	$3, 2, 1$	1

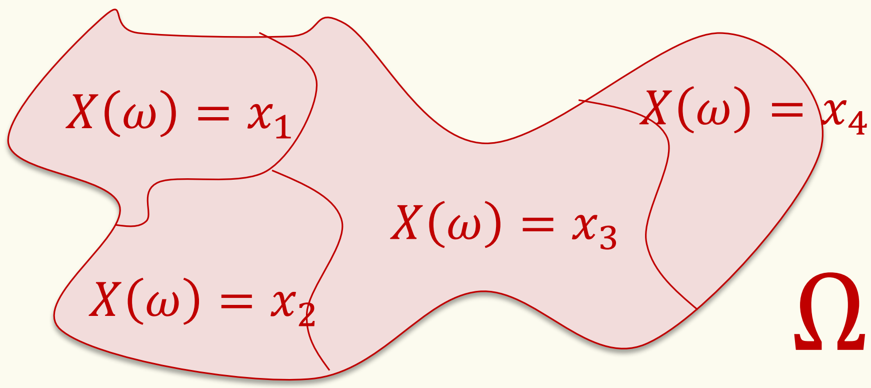


Probability Mass Function (PMF)

Definition. A **random variable (RV)** for a probability space $(\Omega, \mathbb{P})$ is a function $X: \Omega \to \mathbb{R}$.

The set of values that X can take on is called its range/support Ω_X

Random variables partition the sample space.



Definition. For a RV $X: \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

Probability Mass Function (PMF)

Definition. For a RV $X : \Omega \to \mathbb{R}$, we define the event

$$\{X = x\} \stackrel{\text{def}}{=} \{\omega \in \Omega \mid X(\omega) = x\}$$

The **probability mass function** (PMF) of X tells us the probabilities of these events, i.e., the probability that X takes each value in Ω_X

We use the notation

$$p_X(x) = \mathbb{P}(X = x) = \mathbb{P}(\{\omega \in \Omega \mid X(\omega) = x\})$$

For the probability mass function

$$\sum_{x \in \Omega_X} \mathbb{P}(X = x) = 1$$

Probability Mass Function

Flipping two independent coins $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

$X = \text{number of heads in the two flips}$

$X(HH) = 2 \qquad X(HT) = 1 \qquad X(TH) = 1 \qquad X(TT) = 0$

$\Omega_{\mathrm{X}} = \{0, 1, 2\}$

What is the pmf of X?

Probability Mass Function

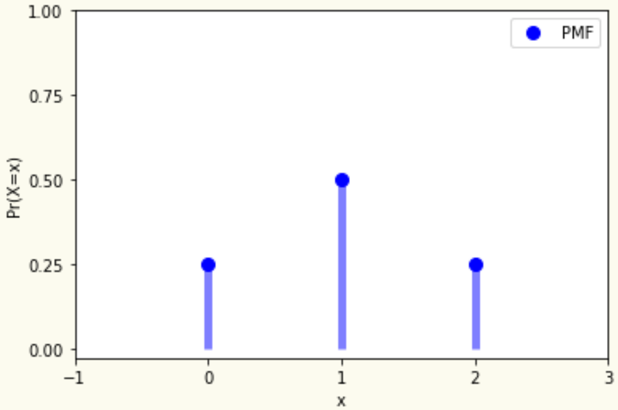
Flipping two independent coins $\Omega = \{\mathrm{HH}, \mathrm{HT}, \mathrm{TH}, \mathrm{TT}\}$

$X = \text{number of heads in the two flips}$

$$X(HH) = 2 \qquad X(HT) = 1 \qquad X(TH) = 1 \qquad X(TT) = 0$$

$$\Omega_{\mathrm{X}} = \{0, 1, 2\}$$

$$\Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$



RV Example

20 balls labeled 1, 2, …, 20 in a bin

- Draw a subset of 3 uniformly at random
- Let X = maximum of the 3 numbers on the balls

Poll: **slido.com/4171468**

What is $p_X(20) = P(X = 20)$?

A. $\binom{20}{2} / \binom{20}{3}$

B. $\binom{19}{2} / \binom{20}{3}$

C. $19^2 / \binom{20}{3}$

D. $19 \cdot 18 / \binom{20}{3}$

Agenda

- Random Variables
- Probability Mass Function (PMF)
- Cumulative Distribution Function (CDF)

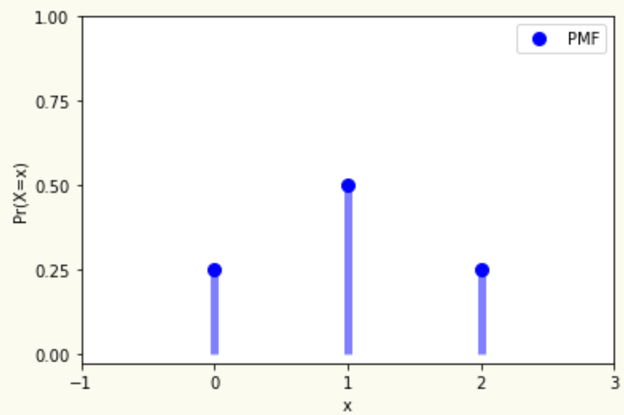
Cumulative Distribution Function (CDF)

Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of X specifies for any real number x, the probability that $X \leq x$.

$$\mathrm{F}_X(x) = \Pr(X \leq x)$$

Go back to 2 coin flips, where X is the number of heads

$$\Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \\ 0, & o.w. \end{cases}$$



Cumulative Distribution Function (CDF)

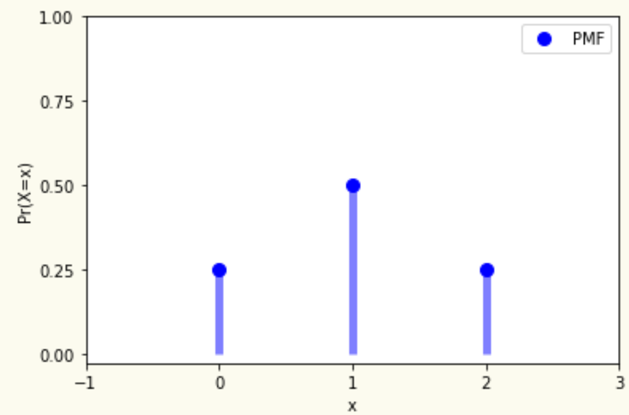
Definition. For a RV $X: \Omega \to \mathbb{R}$, the cumulative distribution function of X specifies for any real number x, the probability that $X \leq x$.

$$\mathrm{F}_X(x) = \Pr(X \leq x)$$

Go back to 2 coin clips, where X is the number of heads

$$\Pr(X = x) = \begin{cases} \frac{1}{4}, & x = 0 \\ \frac{1}{2}, & x = 1 \\ \frac{1}{4}, & x = 2 \end{cases}$$

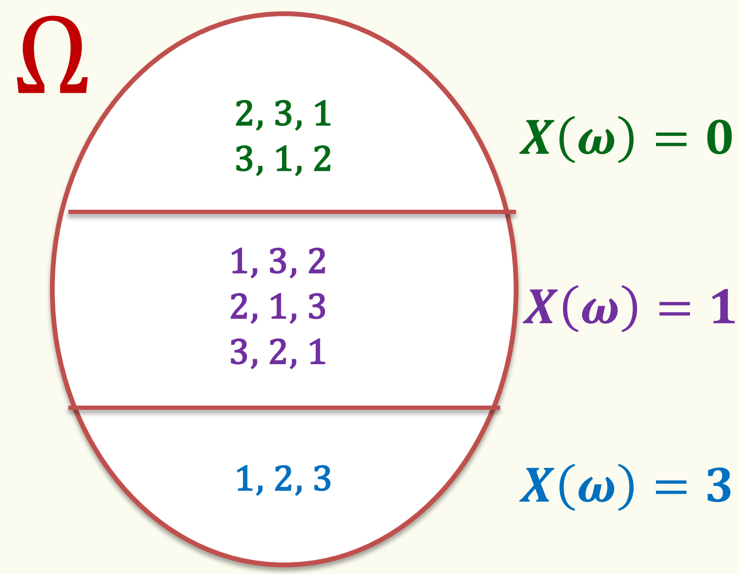
$$F_X(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{4}, & 0 \leq x < 1 \\ \frac{3}{4}, & 1 \leq x < 2 \\ 1, & 2 \leq x \end{cases}$$



Example: Returning Homeworks

- Class with 3 students, randomly hand back homeworks. All permutations equally likely.
- Let X be the number of students who get their own HW

$\Pr(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	3
1/6	1, 3, 2	1
1/6	2, 1, 3	1
1/6	2, 3, 1	0
1/6	3, 1, 2	0
1/6	3, 2, 1	1



Example: Returning Homeworks

$Pr(\omega)$	ω	$X(\omega)$
1/6	1, 2, 3	3
1/6	1, 3, 2	1
1/6	2, 1, 3	1
1/6	2, 3, 1	0
1/6	3, 1, 2	0
1/6	3, 2, 1	1

- Class with 3 students, randomly hand back homeworks.

All permutations equally likely.

- **Let X be the number of students who get their own HW**

Probability Mass Fn

PMF p_X

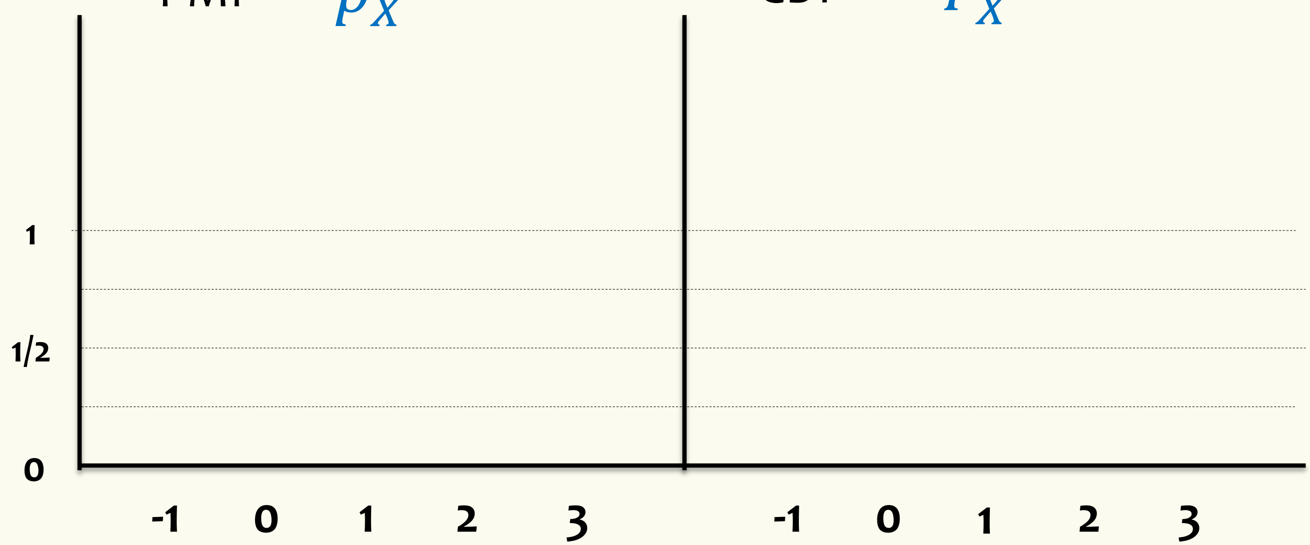
Cumulative Distribution Fn

CDF F_X

$$\mathbf{p_X(0) = P(X = 0) = 1/3}$$

$$\mathbf{p_X(1) = P(X = 1) = 1/2}$$

$$\mathbf{p_X(2) = P(X = 3) = 1/6}$$



Example – Number of Heads

We flip n coins, independently, each heads with probability p

$\Omega = \{\mathrm{HH} \cdots \mathrm{HH}, \mathrm{HH} \cdots \mathrm{HT}, \mathrm{HH} \cdots \mathrm{TH}, \dots, \mathrm{TT} \cdots \mathrm{TT}\}$

$X =$ # of heads

$$p_X(k) = P(X = k) =$$

Example – Number of Heads

We flip n coins, independently, each heads with probability p

$$\Omega = \{\mathrm{HH} \cdots \mathrm{HH}, \mathrm{HH} \cdots \mathrm{HT}, \mathrm{HH} \cdots \mathrm{TH}, \dots, \mathrm{TT} \cdots \mathrm{TT}\}$$

$X =$ # of heads

$$p_X(k) = P(X = k) = \binom{n}{k} \cdot p^k \cdot (1-p)^{n-k}$$

# of sequences with k heads

Prob of sequence w/ k heads


gifak-net
h_2
h_1
b
a
S=a*h₁
S=b*h₂