S.KIMMEL

Goals

Calculate expectation values using

- Indicator random variable
- Linearity of expectation

<u>def</u>: Given a sample space S, a random variable X is a function $X: S \to \mathbb{R}$.

ex: Let S be the sample space consisting of all possible outcomes of 4 coin tosses. Let X be the number of heads that occur.

Q: What is $X(T,H,H,H)$? What is $X(T,T,H,T)$?

A: 1, 3   B: 2, 2 $\boxed{\text{C: 3, 1}}$ D: 4, 4

<u>def</u>: The <u>expected</u> or <u>average</u> value of a random variable X is

$$\mathbb{E}[X] = \sum_{i \in S} \Pr(i)\, X(i).$$

Q: From previous example, what is $\mathbb{E}[X]$? (The averge number of heads in 4 coin flips.)

A) 1 $\boxed{\text{B) 2}}$ C) 2.5   D) 4

S. KIMMEL

- I'm guessing you didn't do the following:

$$\mathbb{E}[X] = \sum_{i \in S} \Pr(i) X(i)$$ (2^4 elements of sample space!)

$$\mathbb{E}[X] = \sum_{\substack{i \in S: \\ X(i)=0}} \Pr(i) \cdot 0 + \sum_{\substack{i \in S \\ X(i)=1}} \Pr(i) + \sum_{\substack{i \in S \\ X(i)=2}} \Pr(i) \cdot 2$$

$$+ \sum_{\substack{i \in S \\ X(i)=3}} \Pr(i) 3 + \sum_{\substack{i \in S \\ X(i)=4}} \Pr(i) 4$$

... good practice to finish on your own!

$\Pr(i) = \frac{1}{2^4} = \frac{1}{16}$ in all cases.

$|\{i \in S : X(i) = 0\}| = 1$ $\quad |\{i \in S : X(i) = 2\}| = \binom{4}{2} = 6$

$|\{i \in S : X(i) = 1\}| = \binom{4}{1} = 4$ $\quad |\{i \in S : X(i) = 3\}| = \binom{4}{3} = 4$

$|\{i \in S : X(i) = 4\}| = 1$

$$\mathbb{E}[X] = \frac{1}{16}\left(1 \cdot 0 + 4 \cdot 1 + 6 \cdot 2 + 4 \cdot 3 + 1 \cdot 4\right)$$
$$= \frac{1}{16}(32) = 2$$

- Instead you used indicator random variables + linearity of expectation (without knowing!)

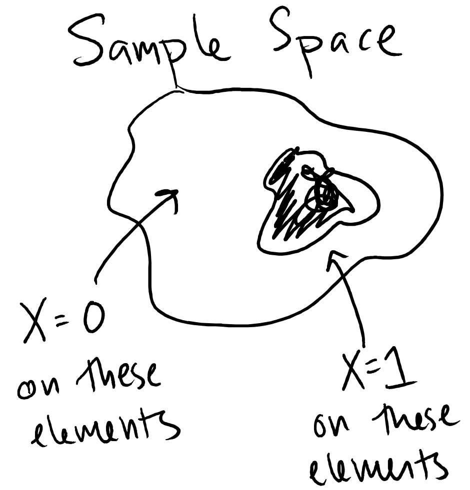
S. KIMMEL

Indicator Random Variable:

def: An <u>indicator random variable</u> X is a random variable such that $X: S \to \{0,1\}$.

An indicator random variable is associated with an event $E \subseteq S$

$$E = \{ i \in S : X(i) = 1 \}$$



$E \subseteq S$ is the event where $X = 1$

Normally write as X_E where

$$X_E(s) = \begin{cases} 1 & \text{if } s \in E \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\mathbb{E}[X_E] = \sum_{i \in S} \Pr(i)\, X_E(i)$$

$$= \sum_{\substack{i \in S: \\ X_E(i)=0}} \Pr(i) \cdot 0 + \sum_{\substack{i \in S \\ X_E(i)=1}} \Pr(i)$$

$$= \sum_{i \in E} \Pr(i) = \Pr(E)$$

$$\boxed{\mathbb{E}[X_E] = \Pr(E)}$$

S.KIMMEL

Linearity of Expectation

Let $Y_1, Y_2, \dots, Y_n$ be random variables on a sample space S. Let $a_1, a_2, \dots a_n \in \mathbb{R}$. Let Y be a random variable s.t.

$$Y = \sum_{k=1}^{n} a_k \cdot Y_k \qquad \left(\text{That is } \forall i \in S,\ Y(i) = \sum_{k=1}^{n} a_k \cdot Y_k(s)\right)$$

Then

$$\mathbb{E}[Y] = \sum_{k=1}^{n} a_k\, \mathbb{E}[Y_k]$$

Ex: Let X_k be the indicator random variable that takes value 1 if k^{th} coin flip is Heads.

$$X = \#\text{ of heads in 4 coin tosses}$$

Then $\quad X = \sum_{k=1}^{4} X_k$

$$\text{e.g.}\quad X(H,T,T,H) = \overset{1}{X_1(H,T,T,H)} + \overset{0}{X_2(H,T,T,H)} + \underset{0}{X_3(H,T,T,H)} + \underset{1}{X_4(H,T,T,H)} = 2$$

$$\mathbb{E}[X] = \sum_{k=1}^{4} \mathbb{E}[X_k] = \sum_{k=1}^{4} \Pr(k^{th} \text{ flip is Heads})$$

linearity of expectation

indicator random variable property

S. KIMMEL

What is average number of heads in 4 coin flips?

GENERAL STRATEGY

read off problem statement $\Downarrow$

1. See average. Need to figure out sample space, random variable of interest

$\{H, T\}^4$ $\qquad$ $X(i) = \#$ heads

2. Write X as a weighted sum of indicator random variables (can do in several steps)

a) Think about what events causes X to increase

Event: 1^{st} flip is H $\rightarrow$ increases X by 1

Event: 2^{nd} flip is H $\rightarrow$ "

" 3^{rd} "

" 4^{th}

b) Create indicator random variables for each event & write X as sum, based on how much each increases

$$X = 1 \cdot X_1 + 1 \cdot X_2 + 1 \cdot X_3 + 1 \cdot X_4 = \sum_{k=1}^{4} X_k$$

3. Use linearity of expectation:

$$\mathbb{E}[X] = \sum_{k=1}^{4} \mathbb{E}[X_k]$$

S.KIMMEL

5. Use $\mathbb{E}[X_E] = \Pr(E)$ for indicator random variables.

$$\mathbb{E}[X] = \sum_{k=1}^{4} \Pr(k^{th} \text{ flip is heads}) = \sum_{k=1}^{4} \frac{1}{2} = 2$$

Super powerful!

ex: Time complexity of random binary search is $O(\log n)$

S.KIMMEL

Q: Consider strings of $\{1,2,4\}^n$, where

$\Pr(k^{th} \text{ digit is } 1) = 1/k$

$\Pr(k^{th} \text{ digit is } 2) = (1-1/k)/2$

$\Pr(k^{th} \text{ digit is } 4) = (1-1/k)/2$

What is average sum of digits?

1. See <u>average</u>. Need sample space, random variable

$\{1,2,4\}^n$ $\quad$ $X(i) = \text{sum of digits in } i$

2. Write X as a weighted sum of indicator random variables (can do in several steps)

$$X = 1 \cdot \underset{\uparrow \atop \#\,1\text{'s}}{X_1} + 2 \cdot \underset{\uparrow \atop \#\,2\text{'s}}{X_2} + 4 \cdot \underset{\uparrow \atop \#\,4\text{'s}}{X_4}$$

$$X_1 = \sum_{k=1}^{n} X_{1,k} \qquad X_{1,k} = \begin{cases} 1 & \text{if } k^{th} \text{ position is } 1 \\ 0 & \text{otherwise} \end{cases}$$

$$X_2 = \sum_{k=1}^{n} X_{2,k} \qquad X_{2,k} = \begin{cases} 1 & \text{if } k^{th} \text{ position is } 2 \\ 0 & \text{otherwise} \end{cases}$$

$$X_4 = \sum_{k=1}^{n} X_{4,k} \qquad X_{4,k} = \begin{cases} 1 & \text{if } k^{th} \text{ position is } 4 \\ 0 & \text{otherwise} \end{cases}$$

$$X = \sum_{k=1}^{n} X_{1,k} + 2\sum_{k=1}^{n} X_{2,k} + 4\sum_{k=1}^{n} X_{4,k}$$

3. Use linearity of expectation:

$$\mathbb{E}[X] = \sum_{k=1}^{n} \mathbb{E}[X_{1,k}] + 2\sum_{k=1}^{n} \mathbb{E}[X_{2,k}] + 4\sum_{k=1}^{n} \mathbb{E}[X_{4,k}]$$

S.KIMMEL

5. Use $\mathbb{E}[X_E] = \Pr(E)$ for indicator random variables.

$$\mathbb{E}[X] = \sum_{k=1}^{n} \underbrace{\mathbb{E}[X_{1,k}]}_{\frac{1}{k}} + \sum_{k=1}^{n} 2\underbrace{\mathbb{E}[X_{2,k}]}_{(1-1/k)/2} + \sum_{k=1}^{n} \underbrace{\mathbb{E}[X_{4,k}]}_{(1-1/k)/2} 4$$

$$= \sum_{k=1}^{n} \frac{1}{k} + 1 - \frac{1}{k} + 2 - 2/k$$

$$= \sum_{k=1}^{n} 3 - \frac{2}{k} = 3n - O(\log n)$$

$$\sum_{k=1}^{n} \frac{1}{k} \approx \ln(n)$$