5.2 Indicator random variables

A <u>sample space</u> is a set consisting of events:

$$S = \{A_1, A_2, \ldots, A_n\}$$

together with a probability function

$$P_r : S \to [0,1] \qquad (\text{i.e., } P_r\{A_i\} \in [0,1])$$

such that $\sum_{i=1}^{n} P_r\{A_i\} = 1$

e.g. Coin toss : $S = \{H, T\}$

$$Pr\{H\} = \frac{1}{2}$$
$$Pr\{T\} = \frac{1}{2}$$

e.g. Dice throw : $S = \{1, 2, 3, 4, 5, 6\}$

$$Pr\{1\} = Pr\{2\} = Pr\{3\} = Pr\{4\} = Pr\{5\} = Pr\{6\} = \frac{1}{6}$$

Indicator random variable $I\{A\}$

associated with A is

$$I\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

e.g. Coin toss $I\{H\} = \begin{cases} 1 & \text{if } H \text{ occurs} \\ 0 & \text{if } T \text{ occurs.} \end{cases}$

With each event A we define an
<u>indicator random variable</u> X_A

$$X_A = I\{A\} = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{if } A \text{ does not occur} \end{cases}$$

<u>Lemma</u>: Given sample space S, event $A \in S$
let $X_A = I\{A\}$.

Then $E[X_A] = \Pr\{A\}$

using indicator random variables:

e.g. Consider the coin toss.

we have $E[X_H] = P_r\{H\} = \frac{1}{2}$

$$E[X_T] = P_r\{T\} = \frac{1}{2}$$

what is the Expected number of H's in n coin tosses?

Let X be random variable denoting number of H's in n coin tosses.

Let X be random variable denoting number of H's in n coin tosses.

Then $X = X_1 + X_2 + \cdots + X_i + \cdots + X_n$

where X_i is the indicator random variable that the i^{th} toss is H.

Then $X = \sum_{i=1}^{n} X_i$

so $E[X] = E[\sum_{i=1}^{n} X_i]$

$= \sum_{i=1}^{n} E[X_i]$ (by independence of events)

$$= \sum_{i=1}^{n} P_r\{X_i\}$$

$$= \sum_{i=1}^{n} \frac{1}{2}$$

$$= \frac{n}{2}$$

Hiring problem

Say we have n candidates in random order.

Let X be the random variable for the number of times a candidate is hired..

$$\text{let } X_i = I\{\text{candidate } i \text{ is hired}\}$$
$$= \begin{cases} 1 & \text{if candidate } i \text{ is hired} \\ 0 & \text{otherwise} \end{cases}$$

then

$$X = X_1 + X_2 + \dots + X_i + \dots + X_n$$

so $E[X] = E[X_1 + X_2 + \dots + X_n]$

$= E[X_1] + E[X_2] + \dots + E[X_i] + \dots + E[X_n]$.

$= P_r\{X_1\} + P_r\{X_2\} + \dots + P_r\{X_i\} + \dots + P_r\{X_n\}$

Note: $P_r\{X_1\} = 1$

$P_r\{X_2\} = 1/2$ $- \left\{ \begin{array}{l} \text{second candidate} \\ \text{must be best of 2} \end{array} \right.$

$P_r\{X_3\} = \frac{1}{3}$ $- \left\{ \begin{array}{l} \text{third candidate} \\ \text{must be best of 3} \end{array} \right.$

$\vdots$

$Pr\{X_i\} = \frac{1}{i}$ - $\begin{cases} i^{th} \text{ candidate} \\ \text{must be best of } i \end{cases}$

$\vdots$

$Pr\{X_n\} = \frac{1}{n}$

Thus $E[X] = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{i} + \dots + \frac{1}{n}$

$$= \sum_{i=1}^{n} \frac{1}{i}$$

$$= \ln(n) + O(1)$$

So the expected (or average) number of hires for n candidates is $\approx \ln(n)$

thus HIRE-ASSISTANT has average-case running cost $O(c_h \ln(n))$

↖ under the assumption that the order of candidates is a uniform random permutation.

That's why we first randomly shuffle the candidates.