Design and Analysis of Algorithms I

Probability Review

Part I

Topics Covered

- Sample spaces
- Events
- Random variables
- Expectation
- Linearity of Expectation

See also:

- Lehman-Leighton notes (free PDF)
- Wikibook on Discrete Probability

Concept #1 – Sample Spaces

Sample space Ω = "all possible outcomes"

[in algorithms, Ω is usually finite]

Also: each outcome $i \in \Omega$ has a **probability** $p(i) \geq 0$.

Constraint: $\sum_{i \in \Omega} p(i) = 1$.

Example #1: Rolling 2 dice. $\Omega = \underbrace{\{(1,1),(2,1),(3,1),\ldots,(5,6),(6,6)\}}_{\text{36 ordered pairs}}$

and $p(i) = 1/36$ for all $i \in \Omega$.

Example #2: choosing a random pivot in outer QuickSort call.

$\Omega = \{1,2,\ldots,n\}$ (index of pivot) and $p(i) = 1/n$ for all $i \in \Omega$.

Concept #2 – Events

An event is a subset $S \subseteq \Omega$.

The probability of an event S is $\sum_{i \in S} p(i)$.

Consider the event (i.e., the subset of outcomes for which) "the sum of the two dice is 7". What is the probability of this event?

- $^1/_{36}$
- $^1/_{12}$
- $^1/_6$
- $^1/_2$

$S = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$

$Pr[S] = 6/36 = 1/6$

Consider the event (i.e., the subset of outcomes for which) "the chosen pivot gives a 25-75 split of better". What is the probability of this event?

$S = \{ (\frac{n}{4}+1)^{th}$ smallest element, ..., $(\frac{3n}{4})^{th}$ smallest element $\}$

- ○ $1/n$
- ○ $1/4$
- ○ $1/2$
- ○ $3/4$

$Pr[S] = \frac{n/2}{n} = \frac{1}{2}$

Concept #2 – Events

An event is a subset $S \subseteq \Omega$.

The probability of an event S is $\sum_{i \in S} p(i)$.

Ex#1: sum of dice = 7. $S = \{(1,6),(2,5),(3,4),\ldots,(6,1)\}$.

$Pr[S] = 6/36 = 1/6$.

Ex#2: pivot gives 25-75 split or better.

$S = \{(\frac{n}{4}+1)^{th}$ smallest element, ..., $(\frac{3n}{4})^{th}$ smallest element$\}$.

$Pr[S] = \frac{n/2}{n} = 1/2$.

Tim Roughgarden

Concept #3 - Random Variables

A random variable X is a real-valued function

$X : \Omega \to \mathbb{R}$.

Ex#1: sum of the two dice.

Ex#2: size of subarray passed to 1st recursive call.

Concept #4 - Expectation

Let $X : \Omega \to \mathbb{R}$ be a random variable.

The expectation $E[X]$ of X = average value of X

$$= \sum_{i \in \Omega} X(i) \cdot p(i)$$

Tim Roughgarden

What is the expectation of the sum of two dice?

- 6.5
- 7
- 7.5
- 8

Which of the following is closest to the expectation of the size of the subarray passed to the first recursive call in QuickSort?

- $n/4$
- $n/3$
- $n/2$
- $3n/4$

let X = subarray size

Then $E[X] = \frac{1}{n}\cdot 0 + \frac{1}{n}\cdot 1 + \frac{1}{n}\cdot 2 + \cdots + \frac{1}{n}\cdot(n-1)$

$= \frac{(n-1)}{2}$

Concept #4 - Expectation

Let $X : \Omega \to \mathbb{R}$ be a random variable.

The expectation $E[X]$ of X = average value of X

$$= \sum_{i \in \Omega} X(i) \cdot p(i)$$

Ex#1: if X = sum of 2 dice, $E[X] = 7$.

Ex#2: if X = 1st subarray size in QuickSort, $E[X] = (n-1)/2$.

Concept #5 – Linearity of Expectation

Claim [LIN EXP]: let $X_1, \ldots, X_n$ be random variables defined on Ω. Then:

$$E\left[\sum_{j=1}^{n} X_j\right] = \sum_{j=1}^{n} E[X_j]$$

CRUCIALLY: HOLDS EVEN WHEN X_j's ARE NOT INDEPENDENT!

[WOULD FAIL IF REPLACE SUMS WITH PRODUCTS]

Ex: if X_1, X_2 = the two dice, then

$E[X_j] = \frac{1}{6}(1+2+3+4+5+6) = 3.5$

By LIN EXP: $E[X_1 + X_2] = E[X_1] + E[X_2] = 3.5 + 3.5 = 7.$

Tim Roughgarden

Linearity of Expectation (Proof)

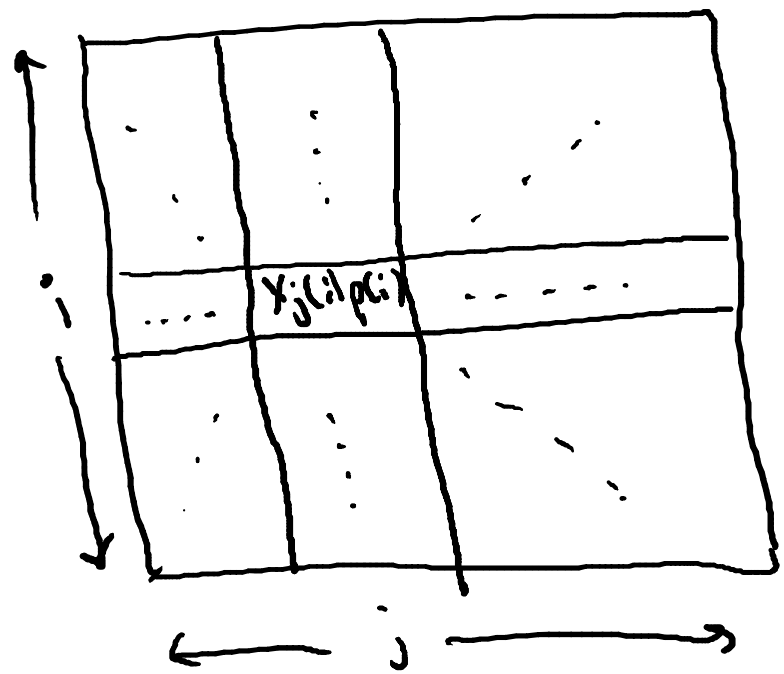
$$\sum_{j=1}^{n} E[X_j] = \sum_{j=1}^{n} \sum_{i \in \Omega} X_j(i)p(i)$$

$$= \sum_{i \in \Omega} \sum_{j=1}^{n} X_j(i)\,p(i)$$

$$= \sum_{i \in \Omega} p(i) \left(\sum_{j=1}^{n} X_j(i) \right)$$

$$= E\left[\sum_{j=1}^{n} X_j \right]$$

QED!



Tim Roughgarden

Example: Load Balancing

Problem: need to assign n processes to n servers.

Proposed Solution: assign each process to a random server.

Question: what is expected number of processes assigned to a server?

Load Balancing Solution

Sample space Ω = all n^n assignments of processes to servers, each equally likely.

Let Y = total number of processes assigned to the first server.

Goal: compute $E[Y]$.

"indicator random variable"

Let $X_j = \begin{cases} 1 & \text{if } j^{th} \text{ process assigned to first server} \\ 0 & \text{otherwise} \end{cases}$

Note: $Y = \sum_{j=1}^{n} X_j$

Tim Roughgarden

Load Balancing Solution (con'd)

We have

$$E[Y] = E\left[\sum_{j=1}^{n} X_j\right] \quad (\text{since } Y = \sum_{j=1}^{n} X_j)$$

$$= \sum_{j=1}^{n} E[X_j] \quad (\text{LIN EXP})$$

$$= \sum_{j=1}^{n} \left(\cancel{Pr[X_j=0]} \cdot 0 + \underbrace{Pr[X_j=1]}_{=\frac{1}{n}} \cdot 1\right)$$

0

(servers chosen uniformly at random)

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

$$= 1.$$

Tim Roughgarden