Probability Review

Part I

Design and Analysis of Algorithms I

Topics Covered

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

See also:

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) >= 0

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

Example #1 : Rolling 2 dice. Ω = {(1,1), (2,1), (3,1),...,(5,6),(6,6) }
Example #2 : Choosing a random pivot in outer QuickSort call.
Ω = {1,2,3,...,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?

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

S = {(n/4+1)th smallest element,.., (3n/4)th smallest element

Pr[S] = (n/2)/n = 1/2

Concept #2 – Events

An event is a subset

The probability of an event S is

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

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

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

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

Pr[S] = (n/2)/n = 1/2

Concept #3 - Random Variables

A Random Variable X is a real-valued function

$$X : \Omega \rightarrow \Re$$

Ex#1 : Sum of the two dice

Ex#2 : Size of subarray passed to 1^{st} recursive call.

Concept #4 - Expectation

Let $X : \Omega \rightarrow \Re$ be a random variable.

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

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

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] = (1/n)*0 + (1/n)*2 + ... + (1/n)*(n-1)

= (n-1)/2

Concept #4 - Expectation

Let $X : \Omega \to \Re$ 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 : Sum of the two dice, E[X] = 7

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

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 :

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

CRUCIALLY:
HOLDS EVEN WHEN X_j's ARE NOT INDEPENDENT!
[WOULD FAIL IF REPLACE SUMS WITH PRODUCTS]

Ex#1 : if X_1,X_2 = the two dice, then
$E[X_j] = (1/6)(1+2+3+4+5+6) = 3.5$

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

Linearity of Expectation (Proof)

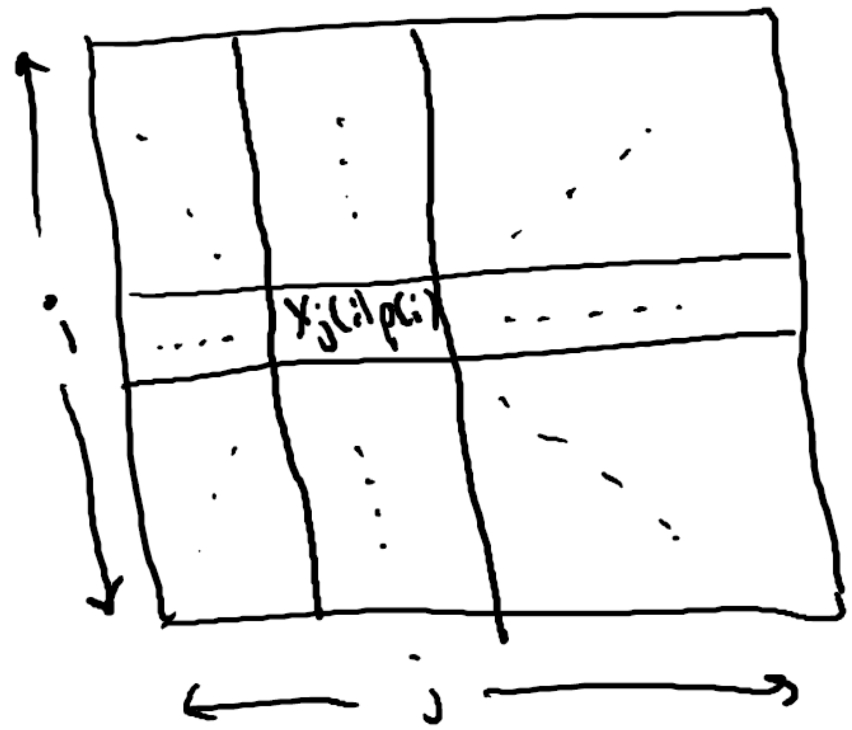
$$\sum_{j=1}^{n} E[X_j] = \sum_{j=1}^{n} \sum_{i \in \Omega} X_j(i) p(i)$$

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

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

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

Q.E.D.



Example: Load Balancing

Problem : need to assign n processes to n servers.

Proposed Solution : assign each process to a random server

Question : what is the 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.

<u>Goal</u> : compute E[Y]

“indicator random variable”

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

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

Load Balancing Solution (con'd)

We have

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

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

$$= \sum_{j=1}^{n} (\underbrace{Pr[X_j = 0] \cdot 0}_{0} + \underbrace{Pr[X_j = 1]}_{= 1/n \text{ (servers chosen uniformly at random)}} \cdot 1)$$

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