# CS648 : Randomized Algorithms
# Semester I, 2011-12, CSE, IIT Kanpur

*Lecture 5,6 : Linearity of Expectation*

In the previous lecture, we introduced three interesting randomized experiments and introduced some random variables for them. We failed to calculate expected value of these random variables by just applying the definition of expected value. In this lecture, we took a *microscopic* view of each of these experiments with the objective of finding the expected value of the corresponding random variables. This approach indeed worked. An interesting outcome of this approach is that we discover a very powerful tool for analysing a randomized experiment. This tool is called the *linearity of expectation*.

## 1 Microscopic view of the randomized experiment

### 1.1 Ball-bin experiment:

There are $m$ balls and $n$ bins. The bins are arranged along a line and labeled 1 to $n$ from left to right. The experiment goes like this: each ball selects a bin randomly uniformly and independent of other balls, and falls into that bin. Let $X$ be the random variable for the number of empty bins. We are interested in finding out $\mathbf{E}[X]$.

Instead of considering all bins at a time, let us focus on $i$th bin for some $i \leq m$. Let us associate a random variable $X_i$ for $i$th bin which is defined as follows.

$$X_i = \begin{cases} 1 & \text{if } i\text{th bin is empty} \\ 0 & \text{otherwise} \end{cases}$$

Let us try to find out the expected value of $X_i$. Since $X_i$ takes 0-1 value only, it can be observed that

$$\mathbf{E}[X_i] = \mathbf{P}[i\text{th bin is empty}]$$

To calculate the probability that $i$th bin is empty, we view the entire ball-bin experiment from the perspective of $i$th bin. Contrast it with the case when we viewed it from the perspective of all bins simultaneously in our *macroscopic* approach. Doesn't the new *microscopic* look easier ? For $i$th bin to be empty, none of the balls should select $i$th bin. Let $\mathcal{E}_i^j$ be the event that $j$th ball does not select $i$th bin. So it can be observed that

$$\mathbf{P}[i\text{th bin is empty}] = \mathbf{P}[\cap_{j=1}^{n} \mathcal{E}_i^j]$$

Now note that each event $\mathcal{E}_i^j$ is independent of others events $\mathcal{E}_i^k, k \neq j$ owing to the fact that each ball selects its destination independent of other balls. Furthermore $\mathbf{P}[\mathcal{E}_i^j] = 1 - 1/n$ since each ball selects its destination uniformly randomly. Hence

$$\mathbf{P}[\cap_{j=1}^{n} \mathcal{E}_i^j] = \Pi_{j=1}^{n} \mathbf{P}[\mathcal{E}_i^j] = \left(1 - \frac{1}{m}\right)^n$$

Hence

$$\mathbf{E}[X_i] = \mathbf{P}[i\text{th bin is empty}] = \left(1 - \frac{1}{m}\right)^n$$

So we observe that finding out $\mathbf{E}[X_i]$ turns out to be much easier than finding out $\mathbf{E}[X]$. You are encouraged to *find out if there is any relationship between $X$ and $X_i$'s*. The intention behind this exploration is quite natural : can we use $\mathbf{E}[X_i]$'s to calculate $\mathbf{E}[X]$ ?

## 1.2 Red-blue balls out of bag:

There are $r$ red balls and $b$ blue balls inside a bag. We take out balls uniformly randomly out of the bag and arrange them along a line in the left to right order. Let $Y$ be the random variable for the number of red balls preceding all blue balls. We are interested in finding out $\mathbf{E}[Y]$.

So let us label the red balls arbitrarily but uniquely using integers from 1 to $r$, and define random variable $Y_i$ for $i$th red ball as follows.

$$Y_i = \begin{cases} 1 & \text{if } i\text{th red ball precedes all the blue balls} \\ 0 & \text{otherwise} \end{cases}$$

Observe that since $Y_i$ takes 0-1 value only, so $\mathbf{E}[Y_i] = \mathbf{P}[Y_i = 1]$. So we need to calculate the probability for the event "$Y_i = 1$". For this purpose, let us consider some special case to develop some familiarity and insight into the problem.

If there is just one red ball in the bin, that is $r = 1$, then what is the probability that this red ball will precede all the blue balls ? For this event to happen, the red ball must be removed from the bin in the first step. However, since we select balls from the bin uniformly randomly, each ball is equally likely to be removed in the first step. So the probability is $1/(b+1)$. A careful reader may feel that the same analysis should hold for each red ball when the number of red balls is more than 1. This is because the real *competitors* of a red ball are only the blue balls. In other words, the order in which the remaining red balls are removed from the bin does not influence the probability of $i$th red ball to precede all blue balls. However, it is quite common that a intuitively appealing idea leads to a wrong argument/solution for a probability problem. So we shall try to provide a formal proof based on this intuition. We shall use the following theorem, which we introduced in the first week of this course.

**Theorem 1.1** *Let $(\Omega, \mathbf{P})$ be a given probability space, and let $S = \{\mathcal{E}_1, \ldots, \mathcal{E}_n\}$ be a partition of the sample space $\Omega$. Then for any event $\mathcal{E}$,*

$$\mathbf{P}[\mathcal{E}] = \sum_{i=1}^{n} \mathbf{P}[\mathcal{E} \cap \mathcal{E}_i] = \sum_{i=1}^{n} \mathbf{P}[\mathcal{E}|\mathcal{E}_i]\ \mathbf{P}[\mathcal{E}_i]$$

In order to effectively calculate $\mathbf{P}[Y_i = 1]$, we focus our attention on the all the red ball excluding the $i$th red ball. For the sake of analysis, it is helpful to consider that the balls are placed from left to right in the order they are removed : the ball which is removed in the beginning is placed at the first place and the ball which is removed at last is placed at $(r+b)$th place. Our set of $r-1$ red balls will occupy exactly $r-1$ distinct places out of the possible $r+b$ places. Let $\mathcal{A}$ be the set of all distinct ways (combinations) of selecting $r-1$ places out of $r+b$ places. For any $s \in \mathcal{A}$, we define an event $\mathcal{E}_s$ as follows: $\mathcal{E}_s$ is the event that the $r-1$ red balls occupy the positions corresponding to the elements of set $s$. What can we say about the set of events $\{\mathcal{E}_s | s \in \mathcal{A}\}$? Well, these events define a partition of the sample space (Do it as a homework). So applying Theorem 1.1, we observe that

$$\mathbf{P}[Y_i = 1] = \sum_{s \in \mathcal{A}} \mathbf{P}[Y_i = 1|\mathcal{E}_s]\ \mathbf{P}[\mathcal{E}_s]$$

Let us try to find out $\mathbf{P}[Y_i = 1|\mathcal{E}_s]$ now. This can be formulated in plain English as follows. Conditioned on an event $\mathcal{E}_s$, that is, given the location of all red balls excluding the $i$th red ball in the output, what is the probability that $i$th red ball precedes all blue balls ? There are total $b+1$ places to be occupied by the $i$th red ball and $b$ blue balls. On the basis of uniformly random manner in which the balls are being removed from the bin, the first place among these $b+1$ places can be occupied by any of these $b+1$ balls with equal probability. So conditioned on $\mathcal{E}_s$, $i$th red ball has $1/(b+1)$ chances of occupying the first vacant place. Hence $\mathbf{P}[Y_i = 1|\mathcal{E}_s] = \frac{1}{b+1}$. (Note that this conditional probability does not depend upon $s$). Hence,

$$\mathbf{P}[Y_i = 1] = \sum_{s \in \mathcal{A}} \frac{1}{b+1} \mathbf{P}[\mathcal{E}_s] = \frac{1}{b+1} \sum_{s \in \mathcal{A}} \mathbf{P}[\mathcal{E}_s] = \frac{1}{b+1}$$

The last equality uses the fact that $\{\mathcal{E}_s | s \in \mathcal{A}\}$ defines a partition of the sample space.

We have thus found that $\mathbf{E}[Y_i] = \frac{1}{b+1}$. So the question we should ask is the following : *what is the relationship between $Y_i$ and $Y$ ?* The intention behind this exploration is again quite natural : can we use $\mathbf{E}[Y_i]$ for various values of $i$ to calculate $\mathbf{E}[Y]$ ?

## 1.3 Randomized quick sort:

The randomized experiment is the execution of randomized quick sort on a set $S$ of $n$ elements. As discussed earlier, the sample space consists of all rooted binary trees on $n$ nodes. Let $Z$ be the random variable for the number of comparisons performed during randomized quick sort on set $S$. We are interested in finding out $\mathbf{E}[Z]$.

Let us develop some familiarity with the randomized quick sort. Firstly, it is a *comparison* based sorting algorithm. [1] Therefore, it is not the absolute values of the elements of $S$ which governs the execution of the algorithm; instead, it is their *relative* values. This suggests us to view the given set $S$ as $\{e_1, e_2, \cdots, e_n\}$, where $e_i$ is the $i$th smallest element. The randomized quick sort is a recursive algorithm based on divide and conquer strategy. In the first step we select a pivot element, say $x$, randomly uniformly from $S$. This pivot element is compared with all the other elements of $S$ to compute two sets : the set $S_{<x}$ of elements smaller than $x$ and the set $S_{>x}$ of elements greater than $x$. We recursively sort $S_{<x}$ and $S_{>x}$. Notice the following observation.

**Observation 1.1** *No element from $S_{<x}$ will be compared with any element from $S_{>x}$ during the algorithm.*

As observed earlier, it is not easy, if not impossible, to calculate $\mathbf{E}[Z]$ from definition. So taking microscopic view, let us focus on just a pair of elements $e_i$ and $e_j$ for any $1 \le i < j \le n$. Note that $e_i$ and $e_j$ will either be compared once or never during quick sort. So let us introduce a random variable $Z_{ij}$ as follows.

$$Z_{ij} = \begin{cases} 1 & \text{if } e_i \text{ is compared with } e_j \\ 0 & \text{otherwise} \end{cases}$$

We are interested in finding out $\mathbf{E}[Z_{ij}]$.

$$\mathbf{E}[Z_{ij}] = \mathbf{P}[Z_{ij} = 1] = \mathbf{P}[e_i \text{ is compared with } e_j \text{ during randomized quick sort}] \tag{1}$$

Let us analyse the execution of the randomized quick sort on $S$. If the first pivot element selected is either smaller than $e_i$ or larger than $e_j$, then $e_i$ and $e_j$ continue to remain in the same recursive call at the next level (of the recursion tree). However, in some recursive call, sooner or later, some pivot element must be selected from the subset $\{e_i, \ldots, e_j\}$. Let us denote this subset by $S_{ij}$. What happens in that recursive call ? If the pivot element in that call is neither $e_i$ nor $e_j$, then $e_i$ and $e_j$ will end up belonging to different recursive calls henceforth, and using Observation 1.1, they will never be compared during the algorithm. On the other hand, if the pivot element is either $e_i$ or $e_j$, then $e_i$ and $e_j$ get compared. We can thus state our second observation.

**Observation 1.2** *In the execution of randomized quick sort on set $S$, the elements $e_i$ and $e_j$ will be compared if and only if the element from the set $S_{ij}$ which is selected first as a pivot element is either $e_i$ or $e_j$.*

Observation 1.2 and Equation 1 implies the following insight. $\mathbf{P}[Z_{ij} = 1]$ is equal to the probability that the element which is selected first as a pivot element from $S_{ij}$ is either $e_i$ or $e_j$. To calculate this probability, let us define events $\mathcal{E}_k$ for $i \le k \le j$ as follows. We say that event $\mathcal{E}_k$ takes place if $e_k$ is selected as a pivot element before any other element from $S_{ij}$. Due to uniformly random way underlying the selection of pivot elements during randomized quick sort, it can be seen that

$$\mathbf{P}[\mathcal{E}_k] = \frac{1}{|S_{ij}|} = \frac{1}{j-i+1}$$

(ponder over the above statement and the equation.)

Hence

$$\mathbf{P}[Z_{ij} = 1] = \mathbf{P}[\mathcal{E}_i \cup \mathcal{E}_j] = \mathbf{P}[\mathcal{E}_i] + \mathbf{P}[\mathcal{E}_j] = \frac{2}{j-i+1}$$

We have used the fact that the events $\mathcal{E}_i$ and $\mathcal{E}_j$ are mutually disjoint in the above derivation.

You are encouraged to *enquire about the relationship between the random variable $Z$ which is the total number of comparisons and the random variables $Z_{ij}$'s*. The intention behind this exploration is again quite natural : can we use $\mathbf{E}[Z_{ij}]$ for various values of $i, j$ to calculate $\mathbf{E}[Z]$ ?

[1] Other examples of *comparison* based sorting algorithms are merge sort, selection sort, heap sort, insertion sort. Examples of sorting algorithm which are not comparison based includes integer sorting algorithms, like : bucket sort, radix sort.

# 2 The usefulness of taking the microscopic view

For each of the three randomized experiments discussed in the previous section, we asked three similar questions. For example, for the ball-bin experiment, what is the relationship between $X_i$'s and $X$ ? The reader may observe easily that for any $\omega \in \Omega$, the value taken by $X$ is the same as the sum of the values taken by $X_i$'s. Same fact holds for the case of $Y$ and $Y_i$'s, and $Z$ and $Z_{ij}$'s. This leads us in a natural way to define the concept of *sum of* random variables.

**Definition 2.1** *(*Sum of random variables*) Given a probability space $(\Omega, \mathbf{P})$, let there be random variable $U$, $V$ and $W$. We call $U$ as sum of random variables $V$ and $W$ if for each $\omega \in \Omega$,*

$$U(\omega) = V(\omega) + W(\omega)$$

*We represent it by the notation "$U = V + W$".*

The above definition can be extended to sum of more than 2 random variable in a straight forward manner.

So what might be the relationship between expected value of $U$ and expected value of $V$ and $W$ if "$U = V + W$"? We state a very important theorem which relates expected value of a sum of random variables with the sum of expected values of each constituent random variable.

**Theorem 2.1** *(***Linearity of expectation***) Given a probability space $(\Omega, \mathbf{P})$, let $U_1, \cdots, U_k$ be any $k$ random variables. Let $U$ be another random variable such that $U = \sum_i U_i$. Then*

$$\mathbf{E}[U] = \sum_{i=1}^{k} \mathbf{E}[U_i]$$

The reader is strongly encouraged to derive a proof of the above theorem. The extent to which you have developed the understanding of the above theorem and its proof can be judged by the preciseness and clarity with which you can provide a very satisfactory answer to the following question. *Why the constituent random variables do not have to independent for linearity of expectation to hold ?*

Attentive reader might have realized the usefulness of Theorem 2.1 for the three examples mentioned above.

- For ball-bin problem, the expected number of empty bins is

  $$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = n\left(1 - \frac{1}{n}\right)^m$$

  Note that it is much more compact and informative than the complicated expression you might have obtained if you had tried other ways.

- For red-blue balls problem, the expected number of red balls preceding all blue balls, is

  $$\mathbf{E}[Y] = \sum_{i=1}^{r} \mathbf{E}[Y_i] = \sum_{i=1}^{r} \frac{1}{b+1} = \frac{r}{b+1}$$

- For randomized quick sort, it implies that the expected number of comparisons is

  $$\mathbf{E}[Z] = \sum_{1 \le i < j \le n} \mathbf{E}[Z_{ij}] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{2}{j-i+1}$$

  This is the exact expression. Try to see how does it look like asymptotically. It is not difficult...

## 3 Usefulness of Linearity of Expectation

The usefulness of linearity of expectation can be summarized in the following few sentences.

Suppose there is a random variable $X$ defined over a probability space $(\Omega, \mathbf{P})$, and one aims to compute expected value of $X$. However, it turns out to be quite difficult to calculate $\mathbf{E}[X]$ directly by its definition, that is, by using equation

$$\mathbf{E}[X] = \sum_{\omega} X(\omega) \cdot \mathbf{P}[\omega]$$

In this case, try to express $X$ as sum of random variables $X_1, \ldots, X_j$ such that it is easy to calculate $\mathbf{E}[X_i]$ for each $i < j$. There may be various ways to express $X$ as sum of random variables. However, finding out the constituent random variables $X_1, ..., X_j$ whose expected value is easy to compute requires ingenuity and experience. This is the most enjoyable as well as challenging part of the analysis.

## 4 Practice problems

1. In the ball-bin experiment analysed in the chapter, what is the expected number of bins with exactly 2 balls ?

2. $n$ persons, all of different heights are standing in a queue facing the window of ticket distributor in a cinema hall. What is the expected number of persons visible to the person at the ticket window if each permutation of $n$ persons is equally likely ?

3. There are $n$ weak joints is a straight stick. The stick is dropped on the floor from a certain height. As a consequence each weak joint can break with probability $p$ independent of other joints. What is the expected number of parts into which the stick will break ? What is the expected number of parts having one unbroken joint ?

4. There are $n$ different letters and $n$ envelopes with unique address written on each envelope. Each letter has to be placed into the right envelope (with right address). But we assign them randomly : We mix the pile of letters and the pile of envelopes separately thoroughly and then pick a letter and an envelope, place the letter in the envelope, and post it. We do so until no letter is left. What is the expected number of recipients who get the correct letter ?

5. What about product of random variables ? Can we still say that $\mathbf{E}[Y \cdot Z] = \mathbf{E}[Y] \times \mathbf{E}[Z]$ ? If not, then are there some conditions under which it holds ?

With an open mind and scientific spirit, keep pondering over the solution of the above three problems.