# Lecture - 28

Thursday, 20 October 2016 (17:10-18:00)

Bucket Sort, Poisson Distribution

### 1 Bucket Sort

Bucket sort is a smart sorting algorithm which sorts n elements in O(n) time, based on certain assumption.

All of you sitting in this class have some phone numbers. Assume we want to sort these phone numbers. What we do is, we make  $9 \times 10 = 90$  buckets to fit the numbers in. The buckets ate labelled as 10, 11, 12, .....99. If a number starts with AB, we put it in the corresponding bucket. For example- Phone number 9652435172 goes to bucket 96, phone number 8723913854 goes to bucket 87 and so on. Please note that the buckets are arranged in the sorted order. Let i and j be two buckets such that i < j, then all the elements in bucket i will be less than all the elements in bucket j as shown in the Figure 1.



Fig. 1: Bucketing step in bucket sort

The elements in each bucket are not sorted yet. So, one needs to go every bucket and sort the elements there. When one sorts the individual buckets and concatenate all these buckets, we get a sorted array as shown in Figure 2.

#### 1.1 Analysis

Assume a bucket has  $\alpha$  elements and we use bubble sort to sort the elements inside a bucket. Time taken in sorting a bucket =  $O(\alpha^2)$ 



Fig. 2: Sort the individual buckets

Number of buckets is assumed to be equal to the number of elements = n. We associate a random variable  $X_i$  with  $i^{th}$  bucket where  $1 \le i \le n$ .  $X_i$  represents the number of elements present in the  $i^{th}$  bucket. Complexity of sorting the  $i^{th}$  bucket  $= \sum_{i=1}^{n} cX_i^2$ . Our aim is to find  $E[\sum_{i=1}^{n} cX_i^2]$ .

#### We assume that all the elements are distributed uniformly at random<sup>1</sup>.

From the linearity of expectations, we can write:

$$E[\sum_{i=1}^{n} cX_{i}^{2}] = cE[\sum_{i=1}^{n} X_{i}^{2}]$$
  
=  $cE[\sum_{i=1}^{n} X^{2}]$  (We replace  $X_{i}$  with X because of symmetry between the buckets.)

 $= cE[\sum_{i=1}^{n} n(n-1)p^2 + np]$  (Probability of a ball falling in a bin = 1/n. p = 1/n. E[X] = np. Similarly, one can find  $E[X^2]$ .)

 $= c \times n \times n(n-1)(\frac{1}{n})^2 + n\frac{1}{n}$  (Substituting p = 1/n)  $= c \times n \times (1 - \frac{1}{n} + 1)$ = 2cn - c= O(n)

<sup>&</sup>lt;sup>1</sup> Kindly note that bucket sort works only under this assumption

### 2 Birthday Paradox

Question: How many people should be there in a room so that there is at least one pair of people having a birthday on the same day.

This can be considered as a balls and bins problem. Let there are n balls and m bins. We have to find the expected number of balls which should be uniformly at random thrown in these bins such that there is at least one bin having more than one ball. In the case of birthday paradox m = 365, since there are 365 days in an year.

First ball comes and falls in some bin.

 $\Pr(\text{Next ball falls in come other bin}) = 1 - \frac{1}{m}$ 

 $Pr(Next ball falls in a new bin (not in any of the previous two)) = 1 - \frac{2}{m}$ 

Likewise, we throw n balls and no collision is achieved.

Pr ( $n^{th}$  ball is thrown in a new bin)=  $1 - \frac{n-1}{m}$ 

Pr (n balls fall in the bin without any collision=)  $(1 - \frac{1}{m}) \times (1 - \frac{2}{m}) \times \dots \times (1 - \frac{n-1}{m})$ .

$$\approx \left(e^{-\frac{1}{m}}\right) \times \left(e^{-\frac{2}{m}}\right) \times \dots \times \left(e^{-\frac{n-1}{m}}\right)$$
$$= e^{-\frac{1}{m} + \left(-\frac{2}{m}\right) + \dots + \left(-\frac{n-1}{m}\right)}$$
$$= e^{-\frac{n(n-1)}{2m}}$$

$$\approx e^{-\frac{n^2}{2m}}$$

Pr (collision)= 1- pr(no collision) =  $1 - e^{-\frac{n^2}{2m}}$ 

To get a 50% chance of collision, put  $1 - e^{-\frac{n^2}{2m}} = \frac{1}{2}$ .

$$n = \sqrt{2ln2m} = O(\sqrt{m})$$

## 3 Poisson Process

Let  $\lambda$  be the expected number of cars which enter our institute in an hour. We assume this distribution to be uniform. What is the probability that given an hour, the number of cars which entered the institute is k?

Assume an hour is divided into a very large number of very small time frames (assume in timeframes of 1 microsecond). These time frames are so small, that no two cars can enter the institute in the same time frame. Assume the number of time frames to be n.

Now, one can model this with the help of binomial distribution.

X is a random variable. There are n trials. This has been shown in Figure 3.



Fig. 3: One hour divided into n very small time slots

Each slot represents a trial. If a car enters in a slot, this is termed as success, else failure. Number of cars entering in n slot represents the number of cars entering in an hour.

$$Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

The mean of the distribution =  $\lambda$ . So,  $np = \lambda$ , or  $p = \frac{\lambda}{n}$ .

Hence, 
$$Pr(X = k) = {\binom{n}{k}} \frac{\lambda}{n}^{k} (1 - \frac{\lambda}{n})^{n-k}$$
  

$$= \frac{n!}{k! \times (n-k)!} \times \frac{\lambda^{k}}{n^{k}} \times (1 - \frac{1}{\frac{n}{\lambda}})^{\frac{n}{\lambda} \times \frac{\lambda}{n} \times (n-k)}$$

$$= \frac{n!}{(n-k)! \times n^{k}} \times \frac{\lambda^{k}}{k!} (1 - \frac{1}{\frac{n}{\lambda}})^{\frac{n}{\lambda} \times \frac{\lambda}{n} \times (n-k)}$$

$$= \frac{n(n-1)(n-2)...(n-k+1)}{n^{k}} \times \frac{\lambda^{k}}{k!} (1 - \frac{1}{\frac{n}{\lambda}})^{\frac{n}{\lambda} \times \frac{\lambda}{n} \times (n-k)}$$