CS5314
Randomized Algorithms

Lecture 13: Balls, Bins, Random Graphs
(Balls-and-Bins Model)

Objectives

- Balls-and-Bins Model
  - throwing m balls into n bins
  - can be applied in many practical situations, e.g., assigning jobs to servers
- Bounds on various scenario
  - E.g., maximum load, number of empty bins
- Poisson Distribution

Balls-and-Bins Model

- Suppose we throw m balls to n bins, independently and uniformly at random

Some interesting questions:

1. What will be the distribution of balls?
2. How many bins are empty?
3. How many balls in the fullest bin?

   (We call this number the maximum load)

Maximum Load

Theorem: Suppose we throw n balls into n bins independently and uniformly at random.

Let L = maximum load

Then, for sufficiently large n,

$$\Pr(L \geq (3 \ln n)/(\ln \ln n)) \leq 1/n$$

(Throughout the notes, we use $\ln x$ to denote $\log_e x$)

How to prove?

Maximum Load (Proof)

Let p = Pr(Bin 1 receives at least M balls)

➔ p = Pr(some set of M balls is in Bin 1)

$\leq C^n_M (1/n)^M$ ... (why?)

Then, since

$C^n_M (1/n)^M \leq 1/(M!)$ ... (why?)

$M^M/(M!) \leq \sum_j (M^j)/j! = e^M$ ... (why?)

we have:

$$p \leq (e/M)^M$$

Maximum Load (Proof)

Let P = $\Pr(L \geq M)$ = Pr(some bin has M balls)

➔ $P \leq np \leq n(e/M)^M$ … (why?)

By setting $M = (3\ln n) / (\ln\ln n)$,

$\Pr(L \geq (3\ln n) / (\ln\ln n))$

$\leq n(e/M)^M$

$\leq n((\ln\ln n) / (\ln n))^M$ … (why?)

$= e^{\ln n} (e^{\ln\ln\ln n - \ln\ln n})^{3\ln n/\ln\ln n}$

$= e^{-2\ln n + o(\ln n)} \leq 1/n$ (for large enough n)

Bucket Sort

Suppose we have $n=2^m$ integers to be sorted

We can sort these integers by Bucket Sort:

1. Create n buckets, B_0, B_1, …, B_{n-1}
2. Put the integer into B_j, if its first m bits = binary representation of j
3. Sort each bucket using Bubble-Sort
4. Output the sorted integers in B_0, then those in B_1, then those in B_2, and so on

Remark: Buckets = Bins, Integers = Balls

Bucket Sort

Suppose each integer is drawn independently and uniformly from $[0,2^k)$ for some $k \geq m$

Question:

What is the expected time for the previous Bucket Sort (assume Steps 1 and 2 are done in O(n) time)?

[Note: the expectation is over the random input]

Bucket Sort

Let X_j be the number of integers in B_j

So, $X_j = \text{Bin}(n, 1/n)$

- Suppose the time to bubble-sort the bucket B_j is cX_j^2 for some constant c

Then, expected time

$$= E[\Sigma\, cX_j^2] + O(n) = \Sigma\, E[cX_j^2] + O(n)$$

$$= cn\, E[X_j^2] + O(n)$$

Bucket Sort

Since for X = Bin(n,p), its second moment is

$$E[X^2] = (E[X])^2 + Var[X]$$
$$= (np)^2 + np\,(1-p)$$

So, $E[X_j^2] = (n(1/n))^2 + n(1/n)(1-1/n) < 2$

and we have:

expected time $< 2cn + O(n) = O(n)$

Fraction of Empty Bins

- Next, we consider the fraction of empty bins, when we throw m balls into n bins uniformly and independently

- Since each ball hits Bin 1 with probability $1/n$, we have

$$\Pr(\text{Bin 1 is empty}) = (1 - (1/n))^m$$
$$\approx e^{-m/n}$$

Fraction of Empty Bins

Let $X_j = 1$ if Bin j is empty

$X_j = 0$ otherwise

Let X = total number of empty bins

$= X_1 + X_2 + \ldots + X_n$

Then, $E[X] = E[X_1 + X_2 + \ldots + X_n]$

$\approx n\, e^{-m/n}$

→ expected fraction of empty bins $\approx e^{-m/n}$

Fraction of Bins with r Balls

How about the expected fraction of bins with exactly r balls (for constant r)?

- Using similar approach, we compute Pr(Bin 1 has exactly r balls), which is

$$C_r^m (1/n)^r (1 - (1/n))^{m-r}$$

$$\approx (m^r/r!)\,(1/n)^r\, e^{-m/n} \qquad \text{when } m, n \gg r$$

$$= e^{-m/n} (m/n)^r / r! \quad = \text{desired fraction}$$

Poisson Distribution

This leads to the following definition:

Definition:

A discrete Poisson random variable X with parameter μ is given by the following probability distribution for $r = 0,1,2,\ldots$:

$$\Pr(X = r) = e^{-\mu}\,\mu^r / r!$$

Remark: Poisson RV $\neq$ Poisson trial !!!

Poisson Distribution

Before we proceed, let us verify that for the previous probability distribution,

$$\Pr(X = 0) + \Pr(X = 1) + \Pr(X = 2) + \ldots = 1$$

By definition:

$$\sum_{r=0 \text{ to } \infty} \Pr(X = r)$$

$$= \sum_{r=0 \text{ to } \infty} e^{-\mu} \mu^{r} / r!$$

$$= e^{-\mu} \sum_{r=0 \text{ to } \infty} \mu^{r} / r! \quad = e^{-\mu} e^{\mu} \quad = 1$$

Expectation of Poisson RV

Theorem: Let X be a Poisson random variable with parameter μ. Then,

$$E[X] = \mu$$

Proof: $E[X] = \sum_{r=0 \text{ to } \infty} r \Pr(X = r)$

$= \sum_{r=1 \text{ to } \infty} r \Pr(X = r)$

$= \sum_{r=1 \text{ to } \infty} r\, e^{-\mu}\, \mu^r / r!$

$= \mu \sum_{r=1 \text{ to } \infty} e^{-\mu}\, \mu^{r-1} / (r-1)! \;\; = \mu$ …(why?)

Sum of Independent Poisson RV

Theorem: Let $X_1, X_2, \ldots, X_n$ be independent Poisson random variables with parameters $\mu_1, \mu_2, \ldots, \mu_n$.

Let $X = X_1 + X_2 + \ldots + X_n$

Then, X is a Poisson random variable with parameter $\mu = \mu_1 + \mu_2 + \ldots + \mu_n$.

How to prove? First prove two RVs. Then general case follows by induction

Sum of Independent Poisson RV

Proof: Consider $X = X_1 + X_2$

Then, $\Pr(X = r) = \Pr(X_1 + X_2 = r)$

$= \sum_{k=0 \text{ to } r} \Pr((X_1 = k) \cap (X_2 = r-k))$

$= \sum_{k=0 \text{ to } r} (e^{-\mu_1} \mu_1^{k} / k!)\, (e^{-\mu_2} \mu_2^{r-k} / (r-k)!)$ ... (why?)

$= (e^{-(\mu_1+\mu_2)} / r!) \sum_{k=0 \text{ to } r} C_k^r \mu_1^{k} \mu_2^{r-k}$ ... (why?)

$= (e^{-(\mu_1+\mu_2)} / r!)\, (\mu_1 + \mu_2)^r$

$= e^{-(\mu_1+\mu_2)} (\mu_1 + \mu_2)^r / r!$

MGF of Poisson RV

Theorem: Let X be a Poisson random variables with parameter μ

Then, the MGF for X is

$$M_X(t) = e^{\mu(e^t-1)}$$

How to prove?

Proof

For any t,

$$M_X(t) = E[e^{tX}]$$

$$= \sum_{r=0 \text{ to } \infty} e^{tr} \Pr(X = r)$$

$$= \sum_{r=0 \text{ to } \infty} e^{tr} (e^{-\mu} \mu^r / r!)$$

$$= e^{-\mu} \sum_{r=0 \text{ to } \infty} (e^t \mu)^r / r!$$

$$= e^{-\mu} e^{e^t \mu}$$

$$= e^{\mu(e^t - 1)}$$

Chernoff Bound for Poisson RV

Theorem: Let Y be a Poisson random variables with parameter μ

Then,

(1) If $x > \mu$, $\Pr(Y \geq x) \leq e^{-\mu} (e\mu)^x / x^x$

(2) If $x < \mu$, $\Pr(Y \leq x) \leq e^{-\mu} (e\mu)^x / x^x$

How to prove?

Poisson RV vs Binomial RV

- When throwing m balls to n bins, the number of balls in a certain bin is a Binomial RV Bin(m, $1/n$)

- However, we see that Bin(m, $1/n$) is close to a Poisson RV, with parameter m/n

- In fact, there is a very strong relation between the Poisson and the Binomial distribution …

Limit of Binomial Distribution

Theorem: Let X_n be a Binomial random variable with parameters n and p, where p is a function of n with $\lim_{n\to\infty} np = \lambda$ for some constant λ

Then, for any fixed r

$$\lim_{n\to\infty} \Pr(X_n = r) = e^{-\lambda} \lambda^r / r!$$

How to prove?

Proof

- We will assume the following inequality:

For $|x| \leq 1$, $\quad e^x(1-x^2) \leq 1+x \leq e^x$

(The proof is left as an exercise)

Firstly,

$$\Pr(X_n = r) = C_r^n\, p^r\, (1-p)^{n-r}$$
$$\leq (n^r/r!)\, p^r\, (1-p)^n\, (1-p)^{-r}$$
$$\leq \left((np)^r/r! \right) e^{-np}\, (1-p)^{-r}$$

Proof

Also,

$$\Pr(X_n = r) \geq ((n-r+1)^r / r!)\, p^r\, (1-p)^{n-r}$$
$$\geq \left(((n-r+1)p)^r / r! \right) (1-p)^n$$
$$\geq \left(((n-r+1)p)^r / r! \right) e^{-np}\, (1-p^2)^n$$
$$\geq \left(((n-r+1)p)^r / r! \right) e^{-np}\, (1-np^2)$$

Now, by taking limits on the two inequalities, we get the desired bound