Probability And Computing - End Sem

Jerry John Thomas

111901055

15 December 2021

1 Q1

Its better to have an have a quick average case check first then check the worst case which is big oh.

Starting with a random levelling of given set S, take $L_1 = S$. Now $|S| = n$, now each time a new L_i is constructed by taking each element of L_{i-1} with a probability of 1/2.

Let S have n elements namely $S = L_1 = \{a_1, a_2, a_3, ..., a_n\}$ Now constructing L_2, each a_i is taken into L2 with a probability of 1/2. Now lets construct L_j for a_i to be present in it must be in all sequences from $L_2, L_3, ...L_{j-1} and L_j$. For a_i to be present in L_j it must be picked up in consecutive $j-1$ selections, hence this happens with a probability of $(\frac{1}{2})^{j-1}$.

Now Z_i be an indicator random variable, which takes the value of 1 when the element a_i reaches the set L_j, with a probability $p = (\frac{1}{2})^{j-1}$ and takes the value 0 if a_i is not present in L_j.

$$E[Z_i] = 1.p + 0 = p = (\frac{1}{2})^{j-1}$$

Now Z is the cardinality of L_j. $Z = \sum_{i=1}^{n} Z_i$ Now by linearity of expectations of Z we have

$$E[Z] = n.(\frac{1}{2})^{j-1}$$

, Hence the mean is in the order of n

Now we can proceed the to find the big oh notation case which is the worst case scenario, since we start with S which has n elements, after j-1 selections we have L_j whose cardinality is again upperbounded by n as no new elements can be added

$$|L_j| \leq n$$

Hence,

$$\sum_{j=1}^{r} |L_j| \leq r.n = O(n)$$

Hence it is upper bound be order of n , hence proved.

2 Q2

Terminologies revisit - Winning money is the money the casino will give, i.e. winning money in round 1 is 100.

We first try to establish that that there is high probability of eventually winning before getting into how much we will win.
The gambling terminates after the first win, this is very similar to the geometric random variable case. Probability of winning in a single gamble is $p = \frac{1}{2}$. Let Y be the geometric random variable, and to win in the kth round (lossing k-1 round before).

$$P(Y = k) = (1 - p)^{k-1}p$$

Since $p = \frac{1}{2}$ we have

$$P(Y = k) = (p)^k = (\frac{1}{2})^k$$

If we sum the probabilities that we either win in round 1 or round 2 or round 3 ...round n as n $\rightarrow \infty$ we get

$$\begin{aligned} &= \sum_{k=1}^{\infty}(\frac{1}{2})^k \\ &= \frac{\frac{1}{2}}{1 - \frac{1}{2}} \\ &= 1 \end{aligned} \tag{1}$$

Since the probability of this is 1, it means this is a sure event hence it will happen which is we will win in some round eventually.

Another way to see this is to see that the probability to loose everytime without winning is

$$P(\text{lose all of n rounds}) = (\frac{1}{2})^n$$

$$\text{as n} \rightarrow \infty \text{ this probability tends to } 0$$

Hence the probability of loosing every round tends to zero which means the probability of eventually winning atleast one round is $(1 - (\frac{1}{2})^n)$ and tends to 1. Also since we have a non-zero probability of winning in any round, we will eventually win. If it was zero we would have never won.

Since we have established that we will eventually win, lets us look at what profit he shall get if he wins in some round. We initially start in round 1 with 100 and each time we double the money we gamble. In round-i we gamble with 100.2^{i-1} hence if we win, we will get the same money we gambled in the round-i.

if we reach round-i, it means we lost $i - 1$ rounds, hence the total money lost in the $i - 1$ rounds is

$$Loss = \sum_{x=1}^{x=i-1} 2^{x-1}.100$$

$$Loss = \frac{1.2^{i-1} - 1}{2 - 1} 100$$

$$Loss = 100.(2^{i-1} - 1)$$

$\therefore$ if we loose in (i-1) rounds and win in ith round, the net profit is amount won in ith round - losses during (i-1) rounds.

$$Profit = 100.2^{i-1} - 100.(2^{i-1} - 1)$$

$$Profit = 100$$

3 Q3

References - Lemma 4.2 of Class notes and set balancing from Probability and Computing - second edition(by Mitzenmacher and Upfal).

We have a random vector A of length n, $A = [a_1, a_2, ...a_n]$ with

$$\mathrm{a}_i = \begin{cases} 1 & \text{P=}\frac{1}{2} \\ -1 & \text{P=}\frac{1}{2} \end{cases}$$

Now we are randomly choosing a subset, now each element of A could be there with P=$\frac{1}{2}$ and not be there with P=$\frac{1}{2}$ as it is a random subset. Now let the subset B of size k be $B_k = [b_1, b_2....b_k]$

Now for each b_i

$$\mathrm{b}_i = \begin{cases} 1 & \text{P=}\frac{1}{2} \\ -1 & \text{P=}\frac{1}{2} \end{cases}$$

the above is because we have 4 cases for each a_i a. choosen and value is 1, b.chosen and value is -1 , c.not chosen with value 1 , d.chosen with value -1. Each of a,b,c,d will happen with a equal probability of $\frac{1}{4}$, hence there are 2 cases of chosen both are having equal probabilities. hence we have the above b_i conditioning.

Now lets focus on the subset B_k. Let subset sum be $S_k = \sum_{i=1}^{i=k} b_i$,We need Probability of $S_k \geq \sqrt{4nln(n)}$ Applying lemma 4.2 (chernoff bound) which is

$$P(X \geq a) \leq e^{-a^2/2n}$$

we get

$$\begin{aligned} P(S_k \geq \sqrt{4nln(n)}) &\leq e^{\frac{-4nln(n)}{2.k}} \\ &\leq e^{\frac{-2nln(n)}{k}} \\ \text{as k} \leq \text{x we have} & \\ &\leq e^{-2ln(n)} \\ &\leq e^{ln(n^{-2})} \\ &\leq n^{-2} \end{aligned} \tag{2}$$

Now coming to the claim in the question, which is the other way round i.e. S_k takes atmost the value.

$$\therefore P(S_k \geq \sqrt{4nln(n)}) \geq 1 - n^{-2}$$

we can make it more loose (as specified by the inital question)

$$P(S_k \geq \sqrt{4nln(n)}) \geq 1 - \frac{1}{n^2} \geq 1 - \frac{2}{n^2} \tag{3}$$

4 Q4

We will be deriving a lot of small results that will eventually lead to this final answer. References - Theorem 5.10 of Probability and Computing - second edition(by Mitzenmacher and Upfal)

4.1

(this part is same as quiz3 question) Consider Z, a Poisson Random Variable with $\mu = m$ as mean (kindly note, m is used as mean only in this subsection), Now

$$P(Z = a) = \frac{e^{-\mu}.\mu^a}{a!}$$

$$P(Z = m + h) = \frac{e^{-m}.m^{m+h})}{(m + h)!} \tag{4}$$

$$P(Z = m - h - 1) = \frac{e^{-m}.m^{m-h-1})}{(m - h - 1)!} \tag{5}$$

Now we try to find a relationship between 1 and 2.

$$\begin{aligned}
P(Z = m + h) = & \frac{e^{-m}.m^{m-h-1}.m^{2.h+1}}{(m - h - 1)!(m - h).(m - h - 1)...(m + h)} \\
& = \frac{e^{-m}.m^{m-h-1}}{(m - h - 1)!} \times \frac{m^{2.h+1}}{(m - h).(m - h - 1)...(m + h)} \\
& = P(Z = m - h - 1) \times \frac{m^{2.h+1}}{(m - h).(m - h - 1)...(m + h)} \\
& = P(Z = m - h - 1) \times \frac{m}{m} \times \prod_{t=1}^{h} \frac{m}{m - t} \times \frac{m}{m + t} \\
& = P(Z = m - h - 1) \times \prod_{t=1}^{h} \frac{m^2}{m^2 - t^2} \\
& Hencethefractionintheproductterm \geq 1 \\
& \geq P(Z = m - h - 1)
\end{aligned} \tag{6}$$

Hence we have the result as $P(Z = m + h) \geq P(Z = m - h - 1)$

4.2

Using 4.1 we try to prove $P(Z \geq \mu) \geq \frac{1}{2}$. We are using m instead of μ here.

$$\begin{aligned} P(Z \geq m) &= \sum_{h=m}^{\infty} P(Z = h) \\ &= \sum_{h=0}^{\infty} P(Z = m + h) \\ &\geq \sum_{h=0}^{m-1} P(Z = m + h) \\ &\geq \sum_{h=0}^{m-1} P(Z = m - h - 1) \quad from 4.1 \\ &= P(Z < m) \end{aligned} \tag{7}$$

Hence we have $P(Z \geq m) \geq P(Z < m)$. Since $P(Z \geq m) + P(Z < m) = 1$, we have $P(Z \geq m) \geq \frac{1}{2}$

4.3

In this section, we know that $E[f(Y_1^m, ...Y_n^m)] \geq E[f(X_1^m, ...X_n^m)].Pr(\sum Y_i = m)$

Now for a $E[f(X_1^m, ...X_n^m)]$ that is monotone increasing function on m, we try to prove $E[f(Y_1^m, ...Y_n^m)] \geq E[f(X_1^m, ...X_n^m)].Pr(\sum Y_i \geq m)$

$$
\begin{aligned}
E[f(Y_1^m, ...Y_n^m)] &= \sum_{k=0}^{\infty} E[f(Y_1^m, ...Y_n^m) | \sum Y_i^m = k].P(\sum Y_i^m = k) \\
&\geq \sum_{k=m}^{\infty} E[f(Y_1^m, ...Y_n^m) | \sum Y_i^m = k].P(\sum Y_i^m = k) (\text{f is non-negative}) \\
&\text{By theorem 5.6 we have The distribution of } (Y_1^m, ...Y_n^m) \text{ conditioned on } \sum Y_i^m = k \\
&\text{is same as } (X_1^m, ...X_n^m) \text{ regardless of the value of m.} \\
&= \sum_{k=m}^{\infty} E[f(X_1^k, ...X_n^k)].P(\sum Y_i^m = k) \\
&\text{f is monotonous increasing and } X_a^b \text{ represents b balls into a bins} \\
&\text{since k goes from m to } \infty \text{ and more the balls higher the function} \\
&\text{hence } X_i^m \leq X_i^k \\
&\geq \sum_{k=m}^{\infty} E[f(X_1^m, ...X_n^m)].P(\sum Y_i^m = k) \\
&= E[f(X_1^m, ...X_n^m)].P(\sum Y_i^m \geq m)
\end{aligned}
\tag{8}
$$

4.4

Now to prove the actual claim. Using 4.2 and 4.3 we have

$$
\begin{aligned}
P(Y_i^m \geq m) &\geq \frac{1}{2} && \text{From 4.2} \\
E[f(X_1^m, ...X_n^m)] &\leq \frac{E[f(Y_1^m, ...Y_n^m)]}{P(Y_i^m \geq m)} && \text{from 4.3}
\end{aligned}
\tag{9}
$$

Combining both we get

$$E[f(X_1^m, ...X_n^m)] \leq 2.E[f(Y_1^m, ...Y_n^m)]$$

Hence our claim is proved.