Complexity and Advanced Algorithms – Assignment 5

Siddharth Bhat (20161105)

November 5, 2018

1 N balls are thrown uniformly at random into N bins

We first define bernoulli random variables:

$$X_i \equiv \text{number of balls in bin } i$$
$$Y_{i,j} \equiv \begin{cases} 1 & \text{ball } j \text{ goes into bin } i \\ 0 & \text{otherwise} \end{cases}$$

Now, we know that

$$X_i = \sum_j Y i, j$$

We first compute the mean of X_i, that is, the expected number of balls per bin:

$$\mathbb{E}[X_i] = \mathbb{E}[\sum_j Y_{i,j}] = \sum_j \mathbb{E}[Y_{i,j}] = \sum_j \frac{1}{N} = N$$

That is, the mean is one ball per bin.

1.1 Expected number of empty bins

We first create random variables to describe a single bin being empty, which we then use to find the expected number of empty bins.

We also define:

$Emp_i \equiv$ probability of bin i being **empty** after N rounds

$\mathbb{E}\big[Emp_i\big] = 0 \times \mathbb{P}\big[Emp_i = 0\big] + 1 \times \mathbb{P}\big[Emp_i = 1\big] = \mathbb{P}\big[Emp_i = 1\big]$

$=$ probability that the ith bin is empty after N balls

$=$ (probability that a ball is not thrown into the ith bin)^N

$$= \left(\frac{N-1}{N}\right)^N$$

Next, we define $Emp = \sum_i Emp_i$. Emp is a random variable whose value is the number of empty bins after n rounds.

$$\mathbb{E}\big[Emp\big] = \mathbb{E}\big[\sum_i Emp_i\big] = \sum_i \mathbb{E}\big[Emp_i\big] = N\left(\frac{N-1}{N}\right)^N$$

1.2 Probability that some bin has more than $6 \log n$ balls

We know that $\mu(X_i) = \mathbb{E}\big[X_i\big] = 1$. We will use Chernoff bounds to calculate the value

$$\mathbb{P}\big[X_i \geq 6 \log n\big] = \mathbb{P}\big[X_i \geq \mu(1 + (6 \log N - 1))\big]$$

since $\delta > 1$, we know that $\mathbb{P}\big[X \geq \mu(1+\delta))\big] \leq e^{\frac{\mu\delta^2}{4}}$

$$\mathbb{P}\big[X_i \geq 6 \log n\big] \leq e^{\frac{(6 \log N - 1)^2}{4}}$$

2 sum of N IID geometric random variables X_i with parameter p

Recall that the definition of a geometric random variable:

$\mathbb{P}\big[X = k\big] \equiv$ probability of bernoulli trials failing till round $(k-1)$, succeeding in round k

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

2.1 $\mathbb{E}\big[X_1\big]$

$$\mathbb{E}\big[X_1\big] = \sum_{i=1}^{\infty} \mathbb{P}\big[X_1 = i\big] \cdot i$$

$$= \sum_{i=1}^{\infty} (1-p)^{i-1} p \cdot i = p\left(\sum_{i=1}^{\infty} (1-p)^{i-1} \cdot i\right)$$

$$= p\left(\sum_{i=1}^{\infty} (1-p)^{i-1} + \sum_{i=2}^{\infty} (1-p)^{i-1} + \dots + \sum_{i=k}^{\infty} (1-p)^{i-1} + \dots\right)$$

(using sum of GP is $a/(1-r)$)

$$= p\left(\sum_{start=1}^{\infty} \sum_{i=start}^{\infty} (1-p)^{i-1}\right)$$

$$= p\left(\sum_{start=1}^{\infty} \frac{(1-p)^{start-1}}{1-(1-p)}\right) = p\left(\sum_{start=1}^{\infty} \frac{(1-p)^{start-1}}{p}\right) = \frac{p}{p}\left(\sum_{start=1}^{\infty} (1-p)^{start-1}\right)$$

$$= \left(\sum_{start=1}^{\infty} (1-p)^{start-1}\right) = \left(\sum_{start=1}^{\infty} (1-p)^{start-1}\right) = \frac{1}{1-p}$$

2.2 $\mathbb{E}\big[X\big]$

$$\mathbb{E}\big[X\big] = \mathbb{E}\big[\sum_{i=1}^{N} X_i\big] = \sum_{i=1}^{N} \mathbb{E}\big[X_i\big] = \frac{N}{1-p}$$

2.3 Chernoff bound for tail of X

We first define $X \equiv \sum_i X_i$ where X_i are IID geometric distributions. We define $Y_i \equiv e^{tX_i}$.

$$
\begin{aligned}
\mathbb{E}\big[Y_i\big] = \mathbb{E}\big[e^{tX_i}\big] &= \sum_{i=1}^{\infty} e^{ti} \cdot (1-p)^{(i-1)} p \\
&= p \sum_{i=1}^{\infty} e^{ti} \times (1-p)^{(i-1)} \\
&= \frac{p}{(1-p)} \sum_{i=1}^{\infty} e^{ti} \cdot (1-p)^{i} \\
&= \frac{p}{(1-p)} \sum_{i=1}^{\infty} (e^{t} \cdot (1-p))^{i} \\
&= \frac{p}{1-p} \cdot \frac{e^{t}(1-p)}{1-e^{t}(1-p)} \\
&= \frac{pe^{t}}{1-e^{t}(1-p)}
\end{aligned}
$$

Let $Y \equiv \prod_i Y_i$. Notice that:

$$
Y = \prod_i Y_i = \prod_i e^{tX_i} = e^{\sum_i tX_i} = e^{t\sum_i X_i} = e^{tX}
$$

And hence, we will use Y when deriving the Chernoff bound in place of e^{tX}

We now compute the expectation of Y:

$$
\mathbb{E}\big[Y\big] = \mathbb{E}\big[\prod_i Y_i\big]
$$

Y_i are IID, expectation of product = product of expectation

$$
= \prod_i \mathbb{E}\big[Y_i\big] = \left(\frac{pe^{t}}{1-e^{t}(1-p)}\right)^{n}
$$

Now that we have the expectation, we can begin to compute the Chernoff bound:

$$
\mathbb{P}\big[X \geq a\big] = \mathbb{P}\big[e^{tX} \geq e^{ta}\big] \leq \frac{\mathbb{E}\big[e^{tX}\big]}{e^{ta}} = \frac{\mathbb{E}\big[Y\big]}{e^{ta}}
$$

3 Problem 3

To go to an event you are required to collect one sticker of n different types. These stickers are given out uniformly at random by the dealer. You can approach the dealer any number of times till you get at leats one sticker of each type. Find the expected number of times you have to visit the dealer that you get at least one sticker of each type

Let T be the time to collect all N stickers. Let t_i be the time the ith sticker has been collected *after* $i-1$ *stickers* have been collected.

The probability of picking up a **new sticker** $p_i = (n-(i-1))/n$, since we have $n-(i-1)$ "new" candidates out of n total coupons.

So, t_i is a geometric distribution (since we can fail many times at picking a new ith coupon till we succeed), with expectation $1/p_i$, (since the expectation of a geometric distribution is $1/p$ where p is the success rate).

Now, we use linearity of expectations to calculate:

$$\begin{aligned}
&\mathbb{E}[T] = \mathbb{E}\big[\sum_i t_i\big] = \sum_i \mathbb{E}[t_i] = \sum_i \frac{1}{p_i} \\
&= \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{1} \\
&= n\left(\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + 1\right) \\
&= nH_n
\end{aligned}$$

Where H_n is the nth harmonic number