Homework5

Qinyun Song

1 Markov's inequality

For a random variable X and $X = -9$ with probability 0.5 while $X = 11$ with probability 0.5. Then we can see that

$$E[X] = 0.5 \times -9 + 0.5 \times 11 = 1 \tag{1}$$

And the probability

$$Pr\left[X > 10\right] = 0.5 = \frac{1}{2} \tag{2}$$

2 Comparing concentration inequalities

1. The expectation of X is

$$E\left[X\right] = 1 \times \frac{1}{3} + 0 \times \frac{2}{3} = \frac{1}{3} \tag{3}$$

   Then from the *Markov's inequality*, we can see that

$$Pr\left[X > \frac{t}{3}\right] \leq \frac{1}{t} \tag{4}$$

2. We can calculate the variance of X as

$$var(X) = N \times \frac{1}{3} \times \frac{2}{3} = \frac{2N}{9} \tag{5}$$

3. Define variable $y_i = x_i - \frac{1}{3}$. Then we can see that $E\left[y_i\right] = 0$. We can see that

$$y_i = \frac{2}{3} wp \frac{1}{3} and y_i = -\frac{1}{3} wp \frac{2}{3} \tag{6}$$

   Thus we know that

$$var(Y) = E\left[Y^2\right] = \frac{4}{9} \times \frac{1}{3} + \frac{1}{9} \times \frac{2}{3} = \frac{6}{27} = \frac{2}{9} \tag{7}$$

   So for the std-dev, we have

$$std - dev(Y) = \sqrt{var(Y)} = \frac{\sqrt{2}}{3} \tag{8}$$

   If we want $X > \frac{2N}{3}$, then we need $Y > \frac{2N+1}{3}$. Then by using the *Chebychev's inequality*, we know that

$$Pr\left[X > \frac{2N}{3}\right] = Pr\left[|Y| > \frac{2N+1}{3}\right] = Pr\left[|Y| > \frac{\sqrt{2}}{3} \times \frac{2N+1}{\sqrt{2}}\right] \leq \frac{2}{(2N+1)^2} \tag{9}$$

4. .

3 Hashing

1. .
2. .
3. .
4. .

4 Election prediction

1. Define variable x_i as the $i-th$ person vote 1 or 0. It satisfies the following function

$$x_i = 1 if vote for 1 = -1 if vote for 0 \tag{10}$$

Then we know that

$$E[X] = 1 \times p - 1 \times (1-p) = 2p - 1 \tag{11}$$

Define another variable $x_i' = x_i - (2p-1)$. Then we know that $E[X'] = 0$. We can also know that

$$E[X'] = 4 \times (p(1-p)^2 + (1-p)p^2) = 4p(1-p) \tag{12}$$

So by using the *Chernoff inequality*, we know that

$$Pr\left[\sum_i x_i' > t\right] \leq 2e^{\frac{-t^2/2}{4np(1-p)+t}} \tag{13}$$

Since

$$\sum_i x_i' = \sum_i x_i - n(2p-1) \tag{14}$$

we know that

$$Pr\left[\sum_i x_i' > t\right] = Pr\left[\sum_i x_i > t + n(2p-1)\right] \tag{15}$$

Since we want to know if $\sum_i x_i = 0$ or not, we then need

$$t = -n(2p-1) \tag{16}$$

So we know that

$$Pr\left[\sum_i x_i > 0\right] \leq 2e^{\frac{-(-n(2p-1))^2/2}{4np(1-p)-n(2p-1)}} = 2e^{\frac{-n(2p-1)^2}{4p-4p^2+1}} \tag{17}$$

Since $p = 0.75$, we know that

$$Pr\left[\sum_i x_i > 0\right] \leq 2e^{\frac{-0.25n}{1.75}} = 2e^{-\frac{n}{7}} \tag{18}$$

If we ned at least 99% confidence, we will require the right part of the equation be no less than 99%. So we want that

$$2e^{-\frac{n}{7}} \geq 99\% \tag{19}$$

Finally we can see that, to ensure the confidene no less than 99%, we need

$$n \geq 5 \tag{20}$$

So the bound of n is no less than 5.

2. If the probability now is $p = 0.501$, then we know that

$$Pr\left[\sum_i x_i > 0\right] \leq 2e^{\frac{-0.000004n}{1.999996}} = 2e^{-\frac{n}{499999}} \tag{21}$$

Then we need

$$2e^{-\frac{n}{499999}} \geq 99\% \tag{22}$$

So we need that

$$n \geq 351599 \tag{23}$$

5 Estimating mean and median

1. For the case $n = 3$ and $k = 2$, define the empirical average as

$$u'_{i_1,i_2} = \frac{1}{2}(A[i_1] + A[i_2]) \tag{24}$$

So we have

$$E[u'] = \frac{1}{3}u'_{1,2} + \frac{1}{3}u'_{2,3} + \frac{1}{3}u'_{1,3} \tag{25}$$

$$= \frac{1}{3}\sum_{i=1}^{3} A[i] \tag{26}$$

$$= u \tag{27}$$

Then if $k = 2$, the variance can now be calculated as

$$var = E\left[(u' - u)^2\right] \tag{28}$$

$$= \frac{1}{3}(u'_{1,2} - u)^2 + \frac{1}{3}(u'_{2,3} - u)^2 + \frac{1}{3}(u'_{1,3} - u)^2 \tag{29}$$

$$= \frac{1}{6}(\sum_{i=1}^{3} A[i]^2 - A[1]A[2] - A[2]A[3] - A[1]A[3]) \tag{30}$$

Compare with the replacement case, in this case, we disregard the case that we choose one indice twice.

2. .

6 Randomized Min-Cut

1. .
2. .
3. .
4. .