# Big Data Computing Homework 2

Fabio Fraschetti 1834942

November 2023

[Fabio Fraschetti]

# 1 Exercise 2

***a)***
I calculated the probability that two vectors don't have the same sign as $\frac{\phi}{\pi}$. Then we can apply the technique studied and see that it return 0 or 1 so we can say that it's a Bernoulli distribution and so the estimator of a Bernoulli is:

$$\frac{\sum_{i=1}^{m} Ber_i(\frac{\phi}{\pi})}{m} = \frac{\hat{\phi}}{\pi} \Longrightarrow \hat{\phi} = \frac{\pi \sum_{i=1}^{m} Ber_i(\frac{\phi}{\pi})}{m}$$

The hash functions that we consider are the one composed by all combinations of +1 and -1.

***b)***
For this exercise I will use Chernoff Bound like this:

$$\Pr[|\hat{\phi}(x,y)-\phi(x,y)| > \epsilon\phi(x,y)] \le \delta \Longrightarrow \Pr[\hat{\phi}(x,y) \ge (1+\epsilon)\phi(x,y)]+\Pr[\hat{\phi}(x,y) \le (1-\epsilon)\phi(x,y)] \le \delta$$

here i have decomposed the inequality into the sum of the probability that m $\hat{\phi}(x,y)$ is in the range $(1-\epsilon)\phi$ and $(1+\epsilon)\phi$. The result of this probabilities are:

$$e^{\frac{\epsilon^2}{3}\phi\frac{m}{\pi}} + e^{\frac{\epsilon^2}{2}\phi\frac{m}{\pi}} \le 2e^{\frac{\epsilon^2}{3}\phi\frac{m}{\pi}} = \delta$$

in the inequality I have applied an increase of the left part. Then we find m as:

$$m = -ln\Big(\frac{\delta}{2}\Big)\Big(\frac{3}{\epsilon^2}\frac{\pi}{\phi}\Big)$$

Where the angle $\theta = \frac{\pi}{m}$ where $m$ represent the approximation and $\pi$ to transform it in an angle.

***c)***
The pairs of vectors are $\binom{n}{2} = \frac{n(n-1)}{2}$. I can say that starting from this inequality:

$$\Pr[|\hat{\phi}(x_i,x_j) - \phi(x_i,x_j)| > \epsilon\phi(x_i,x_j)] \le \delta$$

I can compute this probability for every pair by adding the exponential $\binom{n}{2}$

$$1 - (1 - \Pr[|\hat{\phi}(x_i,x_j) - \phi(x_i,x_j)| > \epsilon\phi(x_i,x_j)])^{\frac{n(n-1)}{2}} \le \delta$$

$$(1 - \Pr[|\hat{\phi}(x_i,x_j) - \phi(x_i,x_j)| > \epsilon\phi(x_i,x_j)]) \ge (1-\delta)^{\frac{1}{\frac{n(n-1)}{2}}}$$

Where the first (1-) indicate that at least one element exist and the second because we need to find the minimum value.

$$\Pr[|\hat{\phi}(x_i,x_j) - \phi(x_i,x_j)| > \epsilon\phi(x_i,x_j)] \le 1 - (1-\delta)^{\frac{1}{\frac{n(n-1)}{2}}}$$

Here we can apply the result obtained in the exercise above and write:

$$2e^{\frac{\epsilon^2}{3}\phi\frac{m}{\pi}} = 1 - (1-\delta)^{\frac{1}{\frac{n(n-1)}{2}}}$$

from this I compute m as:

$$m = -ln\Big(\frac{1-(1-\delta)^{\frac{1}{\frac{n(n-1)}{2}}}}{2}\Big)\Big(\frac{3}{\epsilon^2}\frac{\pi}{\phi}\Big)$$

We can see that the result of the previous exercise only differ by the numerator of the logarithm that is represented from the probability wrote in the second line.