DS Assignment 1

Rohit Patil — 17110126

January 31, 2020

Colab Notebook Link : https://colab.research.google.com/drive/1xi0gvQdZeVXDJFlkEA-TSmcOFd8YXQSY

1 Question 1

(a) probability of each vote being independently misrecorded is p = 0.02.
Let X be a random variable which corresponds to the number of votes misrecorded.
let X_i be the random variable which takes value 1 if i'th vote is misrecorded, else 0.
All X_i are independent. Hence $\sum_1^n X_i$ is the number of misrecorded votes.
i.e $X = \sum_1^n X_i$

To find: Upper bound on the probability that more than 4% of the votes are misrecorded in an election of 1,000,000 ballots.
n = 1000000, $\frac{4}{100} * n = 40000$

i.e we want to find an upper bound on $P(X \geq 400000)$
now, $\mu = np = 1000000 * (0.02) = 20000$

We have, Chernoff bound :

$$P(|X - \mu| > \delta\mu) \leq exp(-\frac{\delta^2\mu}{3})$$

$$P(|X - \mu| > 40000 - \mu) \leq exp(-\frac{\delta^2 * 20000}{3})$$

$$40000 - \mu = \delta\mu$$

$$40000 - 20000 = \delta * 20000$$

$$\delta = 1$$

$$Hence, P(X \geq 400000) = exp(-\frac{20000}{3})$$

(b) Let X and Y be the random variables for number of misrecorded votes for A and B respectively.
let Z = X-Y, as X and Y are independent,
$E[Z] = E[X] - E[Y]$
$E[Z] = E[X] - E[Y]$
$\mu = n_1 * p - n_2 * p$
$\mu = 20000 * 0.02 = 400$

Given : candidate A received 510,000 votes and that candidate B received 490,000 votes.

now, net votes added to B considering the misrecorded ones are X-Y. for B to win the elections, $X - Y \geq 10000$.

Hence, we have to find $P(X - Y \geq 10000)$, i.e. $P(Z \geq 10000)$

$$P[Z \geq 10000] = P[Z - \mu \geq 10000 - \mu]$$
$$\mu = 400,$$
$$P[Z - 400 \geq 10000 - 400] \leq exp(-\frac{\delta^2 * 400}{3})$$
$$P[Z - 4000 \geq 9600] \leq exp(-\frac{\delta^2 * 400}{3})$$
$$\therefore \delta\mu = 9600$$
$$\therefore \delta = 24$$
$$Hence, P[Z - 4000 \geq 9600] \leq exp(-\frac{24^2 * 400}{3})$$
$$\therefore P[Z \geq 10000] \leq exp(-76800)$$

2 Question 3

Verify whether the following is a distance function.

$$d(x, y) = min_i|x_i - y_i|$$

let $X = \begin{bmatrix}1\\2\\3\\4\end{bmatrix}$ and $Y = \begin{bmatrix}1\\4\\5\\6\end{bmatrix}$

Here, $X_0 = Y_0 = 1$, hence $min_i|X_i - Y_i| = 0$
i.e $d(X, Y) = 0$ and $X \neq Y$
but for any distance function, if $d(X, Y) = 0$ the $X = Y$
$\therefore$ d is not a distance function
[1]

[1]Collaborator: Anubhav Jain