Aashsih Kumar pcr902

December 3, 2019

Contents

1 Illustration of Hoeffding's Inequality

1.1 Plot for Empirical Frequency, Markov bound, chebychev bound and hoeffding bound

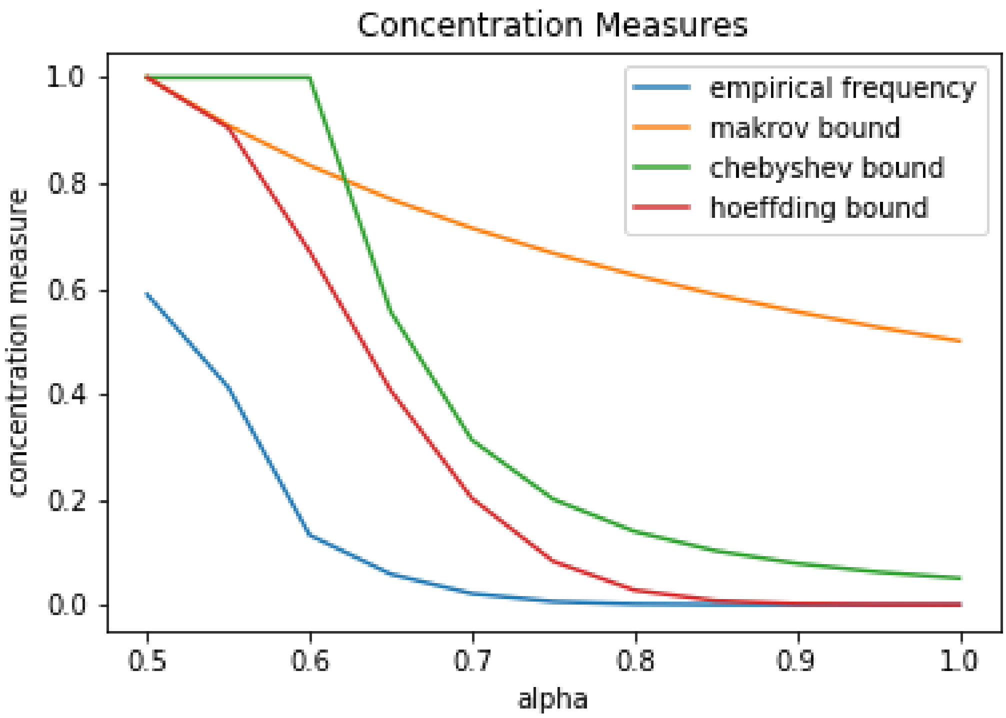

Figure 1: Empirical Frequency, Markov bound, chebychev bound and hoeffding bound

1.2 Comparison of Hoeffding bound with other plots

Hoeffding plot gives a better and tighter bound than both markov and chebychev bounds. For $\alpha \geq 0.85$, it gives a very tight bound over the empirical frequency and coincides with the empirical frequency plot.

1.3 Comparison of Exact Probability for $\alpha = 0.95, 1.0$ with Hoeffding bound

$$P(\frac{1}{20}\sum_{i=1}^{20} X_i \geq 0.95) = 21 * (\frac{1}{20})^{20} = 2.0027161e - 25 \tag{1}$$

$$P(\frac{1}{20}\sum_{i=1}^{20} X_i \geq 0.1) = (\frac{1}{20})^{20} = 9.5367432e - 27 \tag{2}$$

Hoeffding bounds for $\alpha = 0.95, 1.0$ are 0.00030353913807886244, $4.539992976248405e - 05$ respectively.
Both are very very small and close to zero.

2 The effect of scale (range) and normalisation of random variables inHoeffding's inequality

Corollary 2.5 : Let$X_1, ..., X_n$ be independent random variables, such that $X_i \in [0,1], E[X_i] = \mu$ for all i, then for every ϵ¿ 0:

$$P(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \epsilon) \leq e^{-2n\epsilon^2} \tag{3}$$

We need to prove above corollary.
Let $P(X_i) = p$ and $X_i \in [0,1]$
So, $\mu = E[X_i] = 0 * p + 1 * p = p$

$E[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} E[X_i] = n * p$
From the Hoeffding's inequality we have

$$P(|\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i]| \geq \epsilon') \leq e^{\frac{-2\epsilon'^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{4}$$

$$P(|\sum_{i=1}^{n} X_i - n * p| \geq \epsilon') \leq e^{\frac{-2\epsilon'^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{5}$$

$$P(\frac{1}{n}|\sum_{i=1}^{n} X_i - n * p| \geq \frac{1}{n} * \epsilon') \leq e^{\frac{-2\epsilon'^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{6}$$

$$P(\frac{1}{n} * \sum_{i=1}^{n} X_i - p \geq \frac{1}{n} * \epsilon') \leq e^{\frac{-2\epsilon'^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{7}$$

$$P(\frac{1}{n} * \sum_{i=1}^{n} X_i - \mu \geq \frac{1}{n} * \epsilon') \leq e^{\frac{-2\epsilon'^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{8}$$

Let $\epsilon' = n * \epsilon$

$$P(\frac{1}{n} * \sum_{i=1}^{n} X_i - \mu \geq \epsilon) \leq e^{\frac{-2*(n*\epsilon)^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{9}$$

$X_i \in [0,1]$
So, $a_i = 0 and 1_i = 1$
$\sum_{i=1}^{n}(b_i - a_i)^2 = n$

$$P(\frac{1}{n} * \sum_{i=1}^{n} X_i - \mu \geq \epsilon) \leq e^{\frac{-2*(n*\epsilon)^2}{n}} \tag{10}$$

$$P(\frac{1}{n} * \sum_{i=1}^{n} X_i - \mu \geq \epsilon) \leq e^{-2*n(*\epsilon)^2} \tag{11}$$

Hence, proved

3 Distribution of Student's Grades

Let $X_1, X_2, ...X_7$ be the scores in the assignment.
$\hat{Z} = \frac{1}{7} * \sum_{i=1}^{n} X_i$
$E[X_i] = p$ for all i
For$p = 50$ and $\delta = 0.05$
We need to find maximal value of z such that $P(\hat{Z} \leq z) \leq \delta$

3.1 Markov's Inequality

Let $Q = 100 - \hat{Z}$
From Markov's Inequality,

$$
\begin{aligned}
P(Q \geq \epsilon) &\leq \frac{E[Q]}{\epsilon} && (12)\\
P(100 - \hat{Z} \geq \epsilon) &\leq \frac{E[100 - \hat{Z}]}{\epsilon} && (13)\\
P(\hat{Z} \leq 100 - \epsilon) &\leq \frac{100 - E[\hat{Z}]}{\epsilon} && (14)\\
P(\hat{Z} \leq 100 - \epsilon) &\leq \frac{100 - E[\frac{1}{7} * \sum_{i=1}^{7} X_i]}{\epsilon} && (15)\\
P(\hat{Z} \leq 100 - \epsilon) &\leq \frac{100 - \frac{1}{7} * \sum_{i=1}^{7} E[X_i]}{\epsilon} && (16)\\
P(\hat{Z} \leq 100 - \epsilon) &\leq \frac{100 - p}{\epsilon} && (17)
\end{aligned}
$$

Let $z = 100 - \epsilon$

$$
P(\hat{Z} \leq z) \leq \frac{100 - p}{\epsilon} \tag{18}
$$

Thus,

$$
\begin{aligned}
\frac{100 - p}{100 - z} &\leq 0.05 && (19)\\
\frac{50}{100 - z} &\leq 0.05 && (20)\\
z &\leq -900 && (21)
\end{aligned}
$$

3.2 Chebychev Inequality

From chebychev Inequality,

$$
\begin{aligned}
P(|Z - E[Z]| \geq \epsilon) &\leq \frac{Var[Z]}{\epsilon^2} && (22)\\
& && (23)
\end{aligned}
$$

We have

$$P(\hat{Z} \leq z) \leq 0.05 \tag{24}$$
$$P(\hat{Z} - p \leq z - p) \leq 0.05 \tag{25}$$
$$P(|\hat{Z} - p| \geq p - z) \leq 0.05 \tag{26}$$

By using chebychev inequality, we can also say that

$$P(|\hat{Z} - p| \geq p - z) \leq \frac{Var[\hat{Z}]}{(p-z)^2} \tag{27}$$

By using last two equations we can conclude that

$$\frac{Var[\hat{Z}]}{(p-z)^2} \leq 0.05 \tag{28}$$
$$\frac{Var[\hat{Z}]}{(p-z)^2} \leq 0.05 \tag{29}$$
$$Var[\hat{Z}] \leq 0.05 * (p-z)^2 \tag{30}$$

The variance of a random variable $X \in [a, b]$ is maximized when X = 1 with probability 1/2 and X = b with probability 1/2. In other words, let Y be a random variable, such that P(Y = a) = 1/2 and P(Y = b) = 1/2, then for any random variable $X \in [a, b]$ we have $Var[X] \leq Var[Y]$.
a = 0, b = 100 , $\mu = 50$
$Var[Y] = \frac{(0-50)^2+(100-50)^2}{2} = 2500$

$$Var[\hat{Z}] = Var[\frac{1}{7}\sum_{i=1}^{7} X_i] = \frac{1}{7^2}\sum_{i=1}^{7} Var[X_i] = \frac{1}{7}Var[X_i] \leq \frac{1}{7}Var[Y_i] \leq \frac{2500}{7} \tag{31}$$

From last two equations we can say that

$$\frac{2500}{7} \leq 0.05 * (p-z)^2 \tag{32}$$
$$\frac{2500}{7 * 0.05} \leq (50-z)^2 \tag{33}$$
$$\sqrt{\frac{2500}{7 * 0.05}} \leq 50 - z \tag{34}$$
$$z \leq 50 + \sqrt{\frac{2500}{7 * 0.05}} \tag{35}$$

3.3 Hoeffding inequality

From Hoeffding inequality, where $X_i \in [a_i, b_i]$

$$P(|\sum_{i=1}^{n} X_i - E[\sum_{i=1}^{n} X_i]| \geq \epsilon) \leq e^{\frac{-2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}} \tag{36}$$

$$\tag{37}$$

We have

$$P(\hat{Z} \leq z) \leq 0.05 \tag{38}$$
$$P(\hat{Z} - p \leq z - p) \leq 0.05 \tag{39}$$
$$P(|\hat{Z} - p| \geq p - z) \leq 0.05 \tag{40}$$

Let $\epsilon = p - z$
From Hoeffding inequality, we can make the following inference
here $[a_i, b_i] = [0, 100]$

$$e^{\frac{-2\epsilon^2}{\sum_{i=1}^{7}(100-0)^2}} \leq 0.05 \tag{41}$$
$$\frac{-2\epsilon^2}{7 * 100^2} \leq \ln 0.05 \tag{42}$$
$$\frac{-2(50-z)^2}{7 * 100^2} \leq \ln 0.05 \tag{43}$$

3.4 non-vacuous value of z

Markrov inequality gives non-vacous value. In this case, it gives bound smaller than 0.

4 The Airline Question

4.1 Bound the probability that the number of people that show up for a flight will be larger than the number of seats

P(the person will not show up) = 0.05 Let X_i be the random variable that the i^{th} person shows up.
$X_i = 1$, when the i^{th} person shows up else 0.
$\mu = E[X_i] = 0.95 * 1 + 0.05 * 0 = 0.95$
Let Z be the total number of people showed up
$Z = \sum_{i=1}^{100} X_i$

$$P(Z \geq 100) = P(\frac{Z}{100} \geq 1) \tag{44}$$
$$= P(\frac{Z}{100} - \mu \geq 1 - \mu) \tag{45}$$

We can compare the above equation with the following hoeffding inequality

$$P(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \epsilon) \leq e^{-2n\epsilon^2} \tag{46}$$

Let $\epsilon = 1 - \mu$
So, we can say that

$$P(\frac{Z}{100} - \mu \geq 1 - \mu) \leq e^{-2*100*(1-\mu)^2} \tag{47}$$
$$\leq e^{-0.5} \tag{48}$$

4.2

Event 1: In the sample of 10000 passengers, where each passenger shows up with probability p, we observe 95 percent shows up
Event 2 : In the sample of 100 passengers, where each passenger shows up with probability p, everybody shows up
P(Event 2) = p^{100}

Let X_i be the random variable that the i^{th} person shows up.
$X_i = 1$, when the i^{th} person shows up else 0.
$\mu = E[X_i] = p * 1 + (1 - p) * 0 = p$
Let Z be the total number of people showed up
$Z = \sum_{i=1}^{10000} X_i$

$$P(Event1) = P(Z \geq 0.95 * 10000) \tag{49}$$
$$= P(\frac{Z}{10000} \geq 0.95) \tag{50}$$
$$= P(\frac{Z}{100} - \mu \geq 0.95 - \mu) \tag{51}$$

We can compare the above equation with the following hoeffding inequality

$$P(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu \geq \epsilon) \leq e^{-2n\epsilon^2} \tag{52}$$

Let $\epsilon = 0.95 - \mu$
So, we can say that

$$P(\frac{Z}{10000} - \mu \geq 0.95 - \mu) \leq e^{-2*10000*(0.95-\mu)^2} \tag{53}$$
$$\leq e^{-2*10000*(0.95-p)^2} \tag{54}$$

P(Event 1 and Event 2 happens simultaneously) = P(Event1) P(Event2)

$$P(Event1)P(Event2) \leq p^{100}e^{-2*10000*(0.95-p)^2} \tag{55}$$

5 Logistic Regression

5.1 Cross-entropy error measure

5.2 Logistic regression loss gradient

5.3 Logistic regression implementation

5.4 Iris flower data