Assignment-2

November 6, 2017

Team 32

Instructor
Prashanth L.A.

Vamsi Dikkala CS17M048
Vivek Kumar Agrawal CS17M049

Contents

1 Question-I

Expectation:

$$\mathrm{E}(\overline{X}) = E\left(\frac{X_1+X_2+.....+X_n}{N}\right)$$

$E(aX) = aE(X)$

$E(\overline{X}) = \frac{1}{N}\left(E(X_1) + E(X_2) + ... + E(X_n)\right)$

$E(\overline{X}) = \frac{1}{N}\left(\mu + \mu + ... + \mu\right)$

$E(\overline{X}) = \frac{1}{N}\left(N\mu\right)$

$E(\overline{X}) = \mu$

Variance: $\mathrm{X}_1, X_2, ..., X_n$ *are i.i.d random variables*

$Var(X_1+X_2+X3+......+X_n) = Var(X_1)+Var(X_2)+Var(X_3)+....+Var(X_n)$

$$Var(\overline{X}) = Var\left(\frac{X_1 + X_2 + ... + X_n}{N}\right)$$

$$Var(aX) = a^2 Var(X)$$

$$Var(\overline{X}) = \frac{1}{N^2} Var\left(X_1 + X_2 + ... + X_n\right)$$

$$Var(\overline{X}) = \frac{1}{N^2}\left(Var(X_1) + Var(X_2) + ... + Var(X_n)\right)$$

$$Var(\overline{X}) = \frac{1}{N^2}\left(\sigma^2 + \sigma^2 + ... + \sigma^2\right)$$

$$Var(\overline{X}) = \frac{1}{N^2}\left(N\sigma^2\right)$$

$$Var(\overline{X}) = \frac{\sigma^2}{N}$$

2 Question-2

Theorem 1: Let $X_1, X_2, ..., X_n denote\ a\ sequence\ of\ i.i.d\ random\ variables\ with\ X_i \in [a, b],\ for\ all\ i,\ where\ -\infty < a <= b < \infty.\ Letting\ \overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i\ and\ \mu\ denote\ E[X_i]\ for\ all\ i,\ we\ have$

$$P(\overline{X}_N - \mu \geq \epsilon) \leq exp\Big(-\frac{2N\epsilon^2}{(b-a)^2}\Big)\ and\ P(\overline{X}_N - \mu \leq -\epsilon) \leq exp\Big(-\frac{2N\epsilon^2}{(b-a)^2}\Big) \tag{1}$$

We can rewrite equation 1 as

$$P(\overline{X}_N - \epsilon \geq \mu) \leq exp\Big(-\frac{2N\epsilon^2}{(b-a)^2}\Big)\ and\ P(\overline{X}_N + \epsilon \leq \mu) \leq exp\Big(-\frac{2N\epsilon^2}{(b-a)^2}\Big) \tag{2}$$

$$P\Big(\mu \in [\overline{X}_N - \epsilon^{'}, \overline{X}_N + \epsilon^{'}]\Big) \geq 1 - \delta \tag{3}$$

from the given hint, Complement of L.H.S of equation 3 is atmost δ.

$$P\Big(\mu \notin [\overline{X}_N - \epsilon^{'}, \overline{X}_N + \epsilon^{'}]\Big) \leq \delta \tag{4}$$

Substitute $\epsilon^{'}\ in\ equation$ 2

$$P(\overline{X}_N - \epsilon^{'} \geq \mu) \leq exp\Bigg(-\frac{2N\epsilon^{'2}}{(b-a)^2}\Bigg)\ and\ P(\overline{X}_N + \epsilon^{'} \leq \mu) \leq exp\Bigg(-\frac{2N\epsilon^{'2}}{(b-a)^2}\Bigg) \tag{5}$$

Using equation (4) and (5) we can write

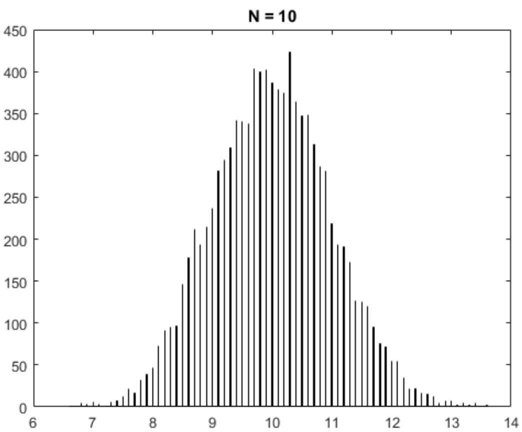
$$P(\overline{X}_N - \epsilon^{'} \geq \mu) + P(\overline{X}_N + \epsilon^{'} \leq \mu) \leq \delta \tag{6}$$

$$
\begin{aligned}
exp\left(-\frac{2N\epsilon^{'2}}{(b-a)^2}\right)+exp\left(-\frac{2N\epsilon^2}{(b-a)^2}\right) &\leq \delta\\
2\ exp\left(-\frac{2N\epsilon^{'2}}{(b-a)^2}\right) &\leq \delta\\
Take\ Natural\ logarithm\ on\ both\ sides\\
-\frac{2N\epsilon^{'2}}{(b-a)^2} &\leq \ln\left(\frac{\delta}{2}\right)\\
\frac{2N\epsilon^{'2}}{(b-a)^2} &\leq \ln\left(\frac{2}{\delta}\right)\\
\epsilon^{'2} &= \frac{(b-a)^2}{2N}\ \ \ln\left(\frac{2}{\delta}\right)\\
\epsilon^{'} &= \sqrt{\frac{(b-a)^2}{2N}\ \ \ln\left(\frac{2}{\delta}\right)}
\end{aligned}
$$

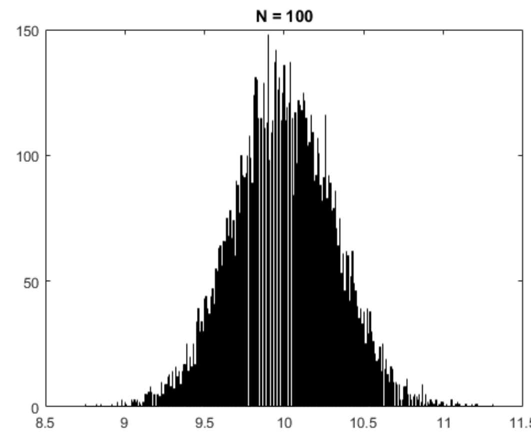
$$
\epsilon^{'} = \sqrt{\frac{(b-a)^2}{2N}\ \ \ln\left(\frac{2}{\delta}\right)} \tag{7}
$$

$$
\delta = 2\ exp\left(-\frac{2N\epsilon^{'2}}{(b-a)^2}\right) \tag{8}
$$

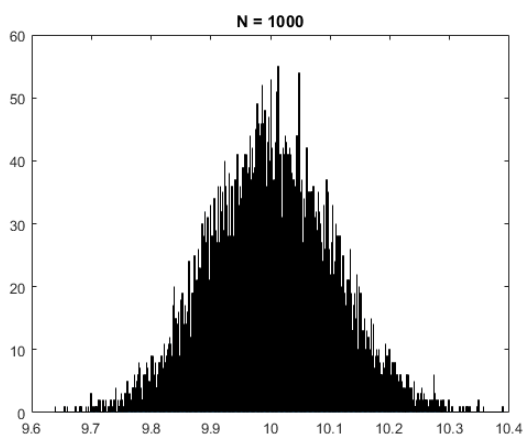
3 Question-3



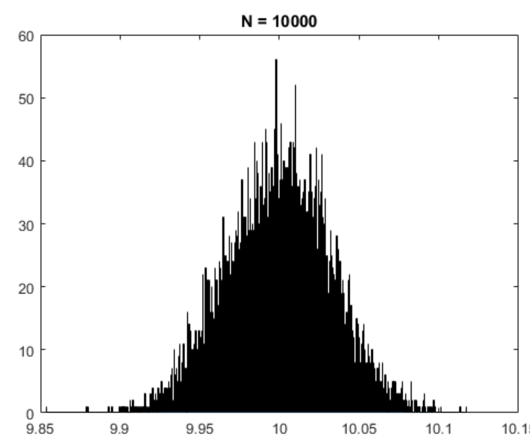
(a) Histogram plot for N=10



(b) Histogram plot for N=100



(c) Histogram plot for N=1000



(d) Histogram plot for N=10000

Figure 1: Histogram plots of sample mean with 1000 bars Mean Value on X-axis
Number of times mean Appeared on Y-axis

Based on Numerical results

a)

Yes,sample mean is close to true mean whenever sample size is large enough. If we draw a sample of individuals from a poission distributed population, the sample will follow a poission distribution.Sample will display the same characterstics.**The Central Limit theorem** says that if we draw a **large** enough sample, the way the sample mean varies around the population mean can be described by a **normal distribution**, irrespective of the population distrubution looks like.**Sample Mean** came from **Normal Distrubution**.And Standard deviation is estimated by

$$\frac{\sigma}{n}$$

.

properties of Normal Distrubution

a)68% of the time, the sample mean and population mean will be within

1^{st} standard deviation of each other.

b)95% of the time, the sample mean and population mean will be within

2^{nd} standard deviation of each other.

c)99% of the time, the sample mean and population mean will be within

3^{rd} standard deviation of each other, and so on.

So If n is large, standard deviation of sample is less, so there is a more chance sample mean and population mean will be closer to each other.

b)

Table 1: Sample Mean Interval

N					
		10	100	1000	10000
Interval	9.90-10.10	1169	2490	6834	9985
	9.99-10.01	400	367	803	2552

c)

Table 2: Confidence Interval

N				
# of times true mean falls	10	100	1000	10000
outside Confidence Interval	811	544	540	496

d) Theorem 1 is applicable for random variables which lies in the finite closed interval,but poission random variables can take any value between $-\infty$ to $+\infty$
e)

$$\textbf{Standard Mean Error} \;=\frac{\sigma}{\sqrt{N}}$$

$$Error_1 = 0.1, Varianceis(\lambda) = 10$$
$$\sqrt{N_1} = \frac{\sqrt{10}}{0.1}$$
$$N_1 = \frac{10}{0.01}$$
$$N_1 = 1000$$
$$Error_2 = 0.01, Varianceis(\lambda) = 10$$
$$\sqrt{N_2} = \frac{\sqrt{10}}{0.01}$$
$$N_2 = \frac{10}{0.0001}$$

$$N_2 = 100000$$

$$N_2 - N_1 = 100000 - 1000 = 99000$$

By 99000 the sample size will increase.
If error is differ by one decimal place
Assume Error is E_1, both errors are differing by 1 decimal place so

$$E_2 = \frac{E_1}{10}$$

$$N_1 = \frac{\sigma^2}{E_1^2}$$

$$N_2 = \frac{10^2\sigma^2}{E_1^2}$$

$$\frac{N_2}{N_1} = 100$$

Jump is 100

$$N_2 = 100N_1$$

4 Question-4

a)

$$f(k) = \frac{A}{k^2} for k = \pm 1, \pm 2, \pm 3.....$$

$$\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

so

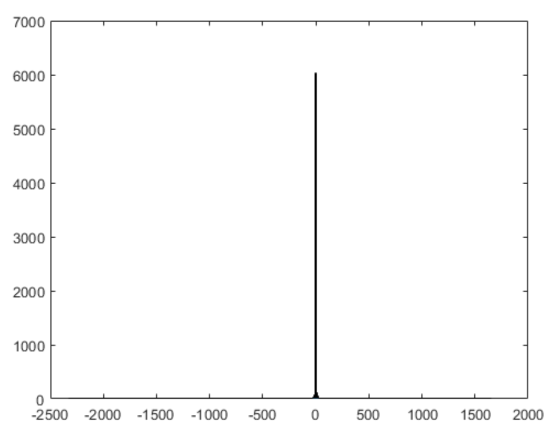
$$\sum_{k \neq 0} f(k) = \frac{2\pi^2}{6}$$

$$\sum_{k \neq 0} Af(k) = A\frac{2\pi^2}{6}$$

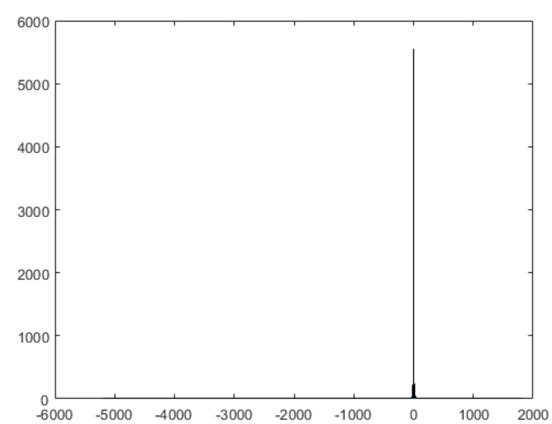
$$A\frac{2\pi^2}{6} = 1$$

$$A = \frac{3}{\pi^2}$$

b)



(a) Histogram plot for N=1000



(b) Histogram plot for N=10000

Figure 2: Histogram plots of sample mean with 1000 bars Mean Value on X-axis Number of times mean Appeared on Y-axis

Table 3: Confidence Interval Limits for Different number of samples

N=1000		N=10000	
Lower Limit	Upper Limit	Lower Limit	Upper Limit
-0.0037	0.0010	-0.0078	0.0224
-0.0006	0.0018	-0.0041	0.0023
-0.0004	0.0006	-0.0003	0.0009
-0.0012	0.0030	-0.0011	0.0066
-0.0008	0.0012	-0.0026	0.0008
-0.0017	0.0009	-0.0006	0.0010
-0.0008	0.0010	-0.0004	0.0032
-0.0007	0.0006	-0.0009	0.0026