CS-6015
Linear Algebra and Random Processes
Programming Assignment 2

Group No : 12
Pritha Ganguly (CS17S013)
Sweta Kumari (BT17D019)
Indian Institute of Technology Madras

November 6, 2017

Course Instructor : Dr. Prashanth L.A.

Problem-1

Let $X_1, \ldots, X_N$ denote the sequence of i.i.d random variables, each with mean μ and variance σ^2. Let $\bar{X_N}$ denote the sample mean, i.e

$$\bar{X_N} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

Therefore,

$$E(\bar{X_N}) = E(\frac{1}{N}\sum_{i=1}^{N} X_i)$$
$$E(\bar{X_N}) = \frac{1}{N}E(X_1 + X_2 \ldots + X_N)$$

Using linearity of expectation,

$$\begin{aligned} E(\bar{X_N}) &= \frac{1}{N}(E(X_1) + E(X_2) \ldots + E(X_N)) \\ &= \frac{1}{N}(N\mu) \\ &= \mu \end{aligned}$$

Similarly,

$$Var(\bar{X_N}) = Var(\frac{1}{N}\sum_{i=1}^{N} X_i)$$
$$Var(\bar{X_N}) = (\frac{1}{N})^2 Var(X_1 + X_2 \ldots + X_N)$$

As all X_i's are i.i.d, variance of their sum is the sum of their variances,

$$Var(\bar{X_N}) = (\frac{1}{N})^2(Var(X_1) + Var(X_2) \ldots + Var(X_N))$$
$$Var(\bar{X_N}) = (\frac{1}{N})^2(N\sigma^2)$$
$$= \frac{\sigma^2}{N}$$

Problem-2

Using the theorem of Hoeffding inequality for bounded random variables, we have the following,

$$P(\bar{X_N} - \mu \geq \epsilon) \leq exp(-\frac{2N\epsilon^2}{(b-a)^2}) \tag{1}$$

$$P(\bar{X_N} - \mu \leq -\epsilon) \leq exp(-\frac{2N\epsilon^2}{(b-a)^2}) \tag{2}$$

We have to to show that, for $\lambda \in (0,1), \epsilon' > 0$

$$P(\mu \in [\bar{X_N} - \epsilon', \bar{X_N} + \epsilon']) \geq 1 - \lambda$$

Equivalently, the complementary event will be:

$$P(\mu \notin [\bar{X_N} - \epsilon', \bar{X_N} + \epsilon']) \leq \lambda$$

which means,

$$P(\mu \leq \bar{X_N} - \epsilon'), P(\mu \geq \bar{X_N} + \epsilon']) \leq \lambda$$
$$P(\epsilon' \leq \bar{X_N} - \mu), P(-\epsilon' \geq \bar{X_N} - \mu) \leq \lambda$$

From equation (1) and (2), we can write,

$$2exp(-\frac{2N\epsilon^2}{(b-a)^2}) \leq \lambda$$

Taking log on both sides,

$$log(2) - \frac{2N\epsilon^2}{(b-a)^2} \leq log(\lambda)$$
$$(b-a)^2 log(\frac{2}{\lambda}) \leq 2N\epsilon^2$$
$$\sqrt{\frac{(b-a)^2 log(\frac{2}{\lambda})}{2N}} \leq \epsilon$$

Therefore,

$$\epsilon \geq \sqrt{\frac{(b-a)^2 log(\frac{2}{\lambda})}{2N}}$$

Problem-3

a) Yes, the sample mean is close to the true mean as we proved in problem-1. Let $X_1, \ldots, X_N$ denote the sequence of i.i.d random variables, each with mean μ and variance σ^2. Let $\bar{X_N}$ denote the sample mean, i.e

$$\bar{X_N} = \frac{1}{N} \sum_{i=1}^{N} X_i$$

$$E(\bar{X_N}) = \frac{1}{N}(E(X_1) + E(X_2) \ldots + E(X_N))$$
$$= \frac{1}{N}(N\mu)$$
$$= \mu$$

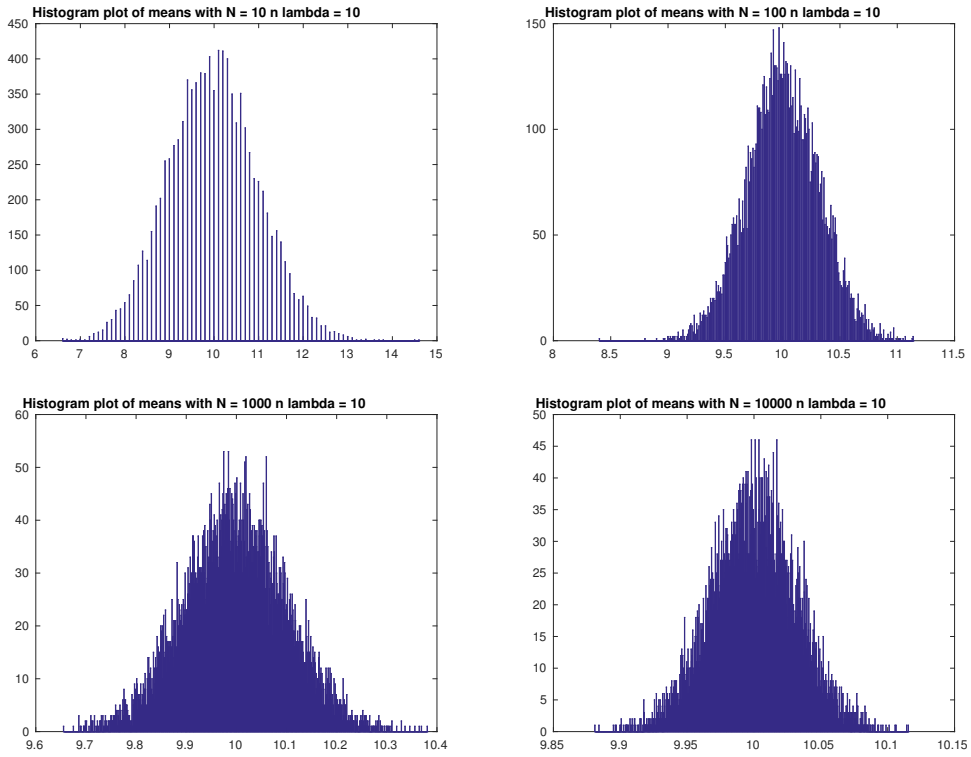

Figure 1: Sample mean histograms for N = 10,100,1000,10000.

The histogram of sample means for different values on N is given in Figure 1.

b) The following table is showing the times was the sample mean in the interval [9:99; 10:01] and [9:9; 10:1] for N=10, 100, 1000, 10000.

N	10	100	1000	10000
interval of 9.99 and 10.01	355	391	822	2456
interval of 9.9 and 10.1	1170	2655	6818	9989

c) The following table is showing the times did the true mean fall outside the confidence interval for N=10, 100, 1000, 10000.

N	10	100	1000	10000
Falls outside the interval	513	508	529	457

d) Theorem 1 is only applicable for bounded random variables but Poisson distribution is not bounded hence we cannot apply the Hoeffding bound on it. We can apply the Hoeffding bound by approximating Poisson distribution with parameter λ by a Binomial distribution with parameters n and $\frac{\lambda}{n}$. A binomial distribution is the sum of independent and identically distributed (i.i.d) Bernoulli random variables. Let $X_1, \ldots, X_n$ denote the sequence of i.i.d Bernoulli($\frac{\lambda}{n}$) random variables with each $\bar{X_i} \in [0, 1]$ for all i = 1,..,n. Therefore, Hoeffding inequality says that,

$$P(|\bar{X_N} - \lambda| \geq \epsilon) \leq 2exp(-2n\epsilon^2)$$

or,

$$P(|\bar{X_N} - \lambda| < \epsilon) \geq 1 - 2exp(-2n\epsilon^2)$$

For 95% confidence interval, we must have,

$$1 - 2exp(-2n\epsilon^2) = 1 - 0.05 = 0.95$$
$$\epsilon = \sqrt{\frac{log(40)}{2n}}$$

Therefore, 95% confidence interval for p = λ can be written as,

$$P(\bar{X_N} - \sqrt{\frac{log(40)}{2n}} < \lambda < \bar{X_N} + \sqrt{\frac{log(40)}{2n}}) \geq 0.95$$

Comparing the above with experimental confidence interval, we get the following table :

N	10	100	1000	10000
Sample mean	9.8	10.36	9.8899	10.03
CI(expt)	[7.40,12.19]	[8.02,12.51]	[8.03,11.95]	[8.08,11.55]
CI(Hoeffding)	[9.37,10.23]	[10.22,10.49]	[9.83,9.92]	[10.017,10.043]

e) The following table is showing the value of N for accuracy 0.1 and 0.01.

Accuracy	0.1	0.01
N	62	611

When the accuracy is 0.001, the sample size is approximately 6100. So, we can generalize and say that on decreasing the accuracy value by one decimal place, the sample size increases 10 times.

Problem-4

a)Given a random variable X that takes values ±1, ±2,... with p.m.f f defined as:

$$\bar{f_k} = \frac{A}{k^2}$$

for k= ±1, ±2,... As we know that sum of the total probability is

$$\sum_k f_k = 1$$

Therefore,

$$\sum_k \frac{A}{k^2} = 1$$

Now we can write it as:

$$A[\frac{1}{(-1)^2} + \frac{1}{(+1)^2} + \frac{1}{(-2)^2} + \frac{1}{(+2)^2} + ...] = 1$$

$$A[\frac{2}{(1)^2} + \frac{2}{(2)^2} + ...] = 1$$

$$2A[\frac{1}{(1)^2} + \frac{1}{(2)^2} + ...] = 1$$

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

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

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

Therefore for $A = \frac{3}{\pi^2}$, f would be a valid p.m.f.

b) The histogram of sample means for different values on N is given in Figure 2. The sample mean does not stay close to a particular value, it varies. Given

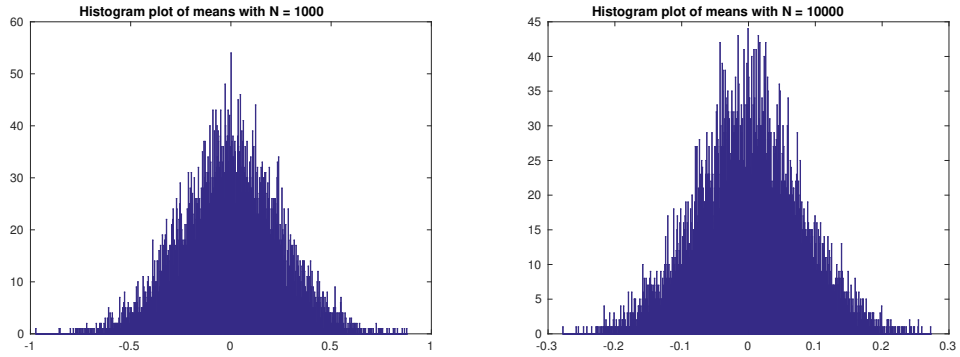

Figure 2: Sample mean histograms for N = 1000,10000.

the PMF, the expectation of X,

$$\sum k.f_k = A\sum_{k!=0}\frac{1}{k}$$

does not converge absolutely. Therefore, the sample mean does not converge to a particular value hence confidence intervals cannot be defined.