CS6015: Linear Algebra and Random Processes
Rajan Kumar Soni (CS18S038) and Kitty Varghese (CS18D300)

Programming Assignment 2

1 Question No 1

1.1 Give an explicit expression for ϵ' as a function of N and δ

$$P\left(\bar{X}_N \geq \mu + \epsilon'\right) \leq e^{\frac{-\epsilon'^2 N}{4}} \dots \quad (1) \qquad P\left(\bar{X}_N \leq \mu - \epsilon'\right) \leq e^{\frac{-\epsilon'^2 N}{4}} \dots \quad (2)$$

these two can be written as

$$P\left(\bar{X}_N - \epsilon' \geq \mu\right) \leq e^{\frac{-\epsilon'^2 N}{4}} \dots \quad (3) \qquad P\left(\bar{X_N} + \epsilon' \leq \mu\right) \leq e^{\frac{-\epsilon'^2 N}{4}} \dots \quad (4)$$

this shows that probability that μ is outside the confidence interval is bounded by $\leq e^{\frac{-\epsilon'^2 n}{4}}$ Let probability that

$$p\left(\mu \in [p - \epsilon', p + \epsilon']\right) = x \tag{5}$$

we know that 3 and 4 are disjoint so

$$P\left(\bar{X}_N - \epsilon' \geq \mu\right) + P\left(\bar{X}_N + \epsilon' \leq \mu\right) + x = 1 \tag{6}$$

since

$$P\left(\bar{X}_N - \epsilon' \geq \mu\right) + P\left(\bar{X}_N + \epsilon' \leq \mu\right) \leq 2 * e^{\frac{-\epsilon'^2 N}{4}} \tag{7}$$

so,

$$1 - x \leq 2 * e^{\frac{-\epsilon'^2 N}{4}} \tag{8}$$

$$x \geq 1 - 2 * e^{\frac{-\epsilon'^2 N}{4}} \implies p\left(\mu \in [p - \epsilon', p + \epsilon']\right) \geq 1 - 2 * e^{\frac{-\epsilon'^2 N}{4}} \tag{9}$$

we have given

$$p\left(\mu \in [p - \epsilon', p + \epsilon']\right) \geq 1 - \delta \tag{10}$$

comparing 9 and 10 we get

$$\delta = 2 * e^{\frac{-\epsilon'^2 N}{4}} \tag{11}$$

expression for

$$\boldsymbol{\epsilon' = \sqrt{-4N \log\left(\frac{\delta}{2}\right)}}$$

1.2 Give an explicit expression for ϵ' as a function of N and δ for bounded random variables

$$P\left(\bar{X}_N - \mu \geq \epsilon'\right) \leq e^{\frac{-\epsilon'^2 2N}{(b-a)}} \ldots \tag{12}$$

$$P\left(\bar{X}_N - \mu \leq -\epsilon'\right) \leq e^{\frac{-\epsilon'^2 2N}{(b-a)}} \ldots \tag{13}$$

these two can be written as

$$P\left(\bar{X}_N - \epsilon' \geq \mu\right) \leq e^{\frac{-\epsilon'^2 2N}{(b-a)}} \ldots \tag{14}$$

$$P\left(\bar{X_N} + \epsilon' \leq \mu\right) \leq e^{\frac{-\epsilon'^2 2N}{(b-a)}} \ldots \tag{15}$$

this shows that probability that μ is outside the confidence interval is bounded by $\leq e^{\frac{-\epsilon'^2 2N}{(b-a)}}$ Let probability that

$$p\left(\mu \in [p-\epsilon', p+\epsilon']\right) = x \tag{16}$$

we know that 3 and 4 are disjoint so

$$P\left(\bar{X}_N - \epsilon' \geq \mu\right) + P\left(\bar{X}_N + \epsilon' \leq \mu\right) + x = 1 \tag{17}$$

since

$$P\left(\bar{X}_N - \epsilon' \geq \mu\right) + P\left(\bar{X}_N + \epsilon' \leq \mu\right) \leq 2 * e^{\frac{-\epsilon'^2 2N}{(b-a)}} \tag{18}$$

so,

$$1 - x \leq 2 * e^{\frac{-\epsilon'^2 2N}{(b-a)}} \tag{19}$$

$$x \geq 1 - 2 * e^{\frac{-\epsilon'^2 2N}{(b-a)}} \implies p\left(\mu \in [p-\epsilon', p+\epsilon']\right) \geq 1 - 2 * e^{\frac{-\epsilon'^2 2N}{(b-a)}} \tag{20}$$

we have given

$$p\left(\mu \in [p-\epsilon', p+\epsilon']\right) \geq 1 - \delta \tag{21}$$

comparing 20 and 21 we get

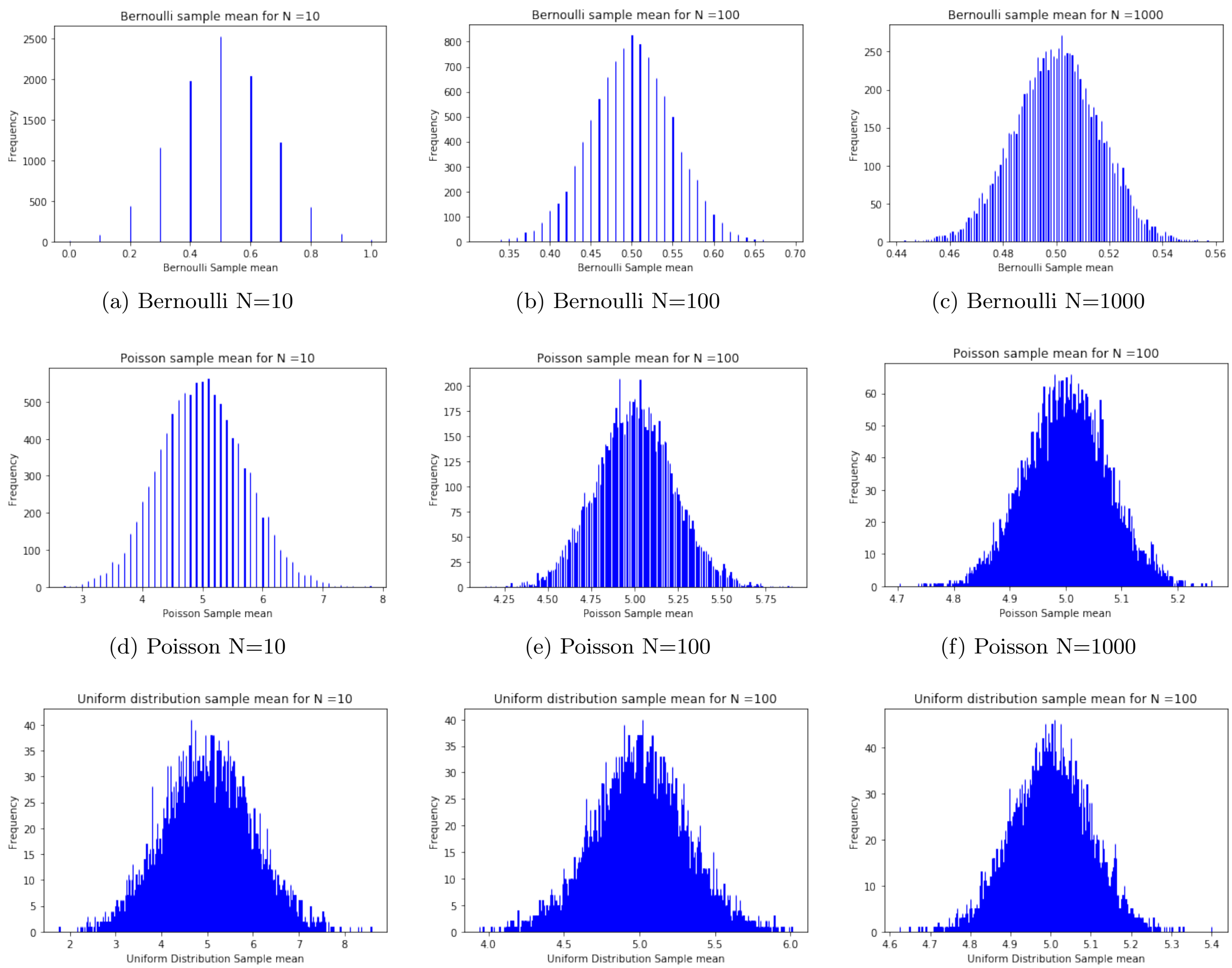
$$\delta = 2 * e^{\frac{-\epsilon'^2 2N}{(b-a)}} \tag{22}$$

expression for

$$\boldsymbol{\epsilon' = \sqrt{\log\left(\frac{\delta}{2}\right)\frac{(a-b)}{2N}}}$$

2 Question No 2

Plot histogram of the sample mean with 1000 bars Observations



(a) Bernoulli N=10 (b) Bernoulli N=100 (c) Bernoulli N=1000

(d) Poisson N=10 (e) Poisson N=100 (f) Poisson N=1000

(g) Uniform N=10 (h) Uniform N=100 (i) Uniform N=1000

Figure 1: Histogram of sample mean with 1000 bars when the experiment is performed for $10,000$ times

- It is observed that as the number of samples increases the curve takes a shape of normal distribution.
- It is observed that the peak of the histogram is obtained near its true mean. This infer that the when the number of sample size increases it reaches it true mean (Law of Large Numbers).
- From fig.1, we notice that as the number of samples increases the curve becomes narrower.

2.1 Is sample mean close to true mean

	Bernoulli	**Poisson**	**Uniform**
True Mean	***0.5***	***5***	***5***
N=10	0.5018	4.993	4.996
N=100	0.499	5.001	5.001
N=1000	0.500	5.000	4.999

Table 1: Sample mean for various number of samples

Observation

- From fig.1, it is observed that as the number of samples increases it becomes nearer to its true mean.
- The deviation of sample mean from true mean is caused due biased sampling methods or randomness inherent in drawing a sample from a population.
- This randomness is decreased when we increase the number of sample and the sample mean becomes equal to true mean. (Law of Large Numbers)

2.2 Sample mean in the interval $[\mu + 0.01, \mu - 0.01]$ and $[\mu + 0.1, \mu - 0.1]$

	Bernoulli		**Poisson**		**Uniform**	
	mean+/- 0.01	***mean+/- 0.1***	***mean+/- 0.01***	***mean+/- 0.1***	***mean+/- 0.01***	***mean+/- 0.1***
N=10	2533	6544	554	1669	80	872
N=100	2390	9634	550	3638	318	2766
N=1000	4924	10000	1186	8404	930	7278

Table 2: Number of times the sample mean was in interval $[\mu - 0.01, \mu + 0.01]$ and $[\mu - 0.1, \mu + 0.1]$ for various N and distributions

Observation

- it is observed that number of sample in $[\mu - 0.1, \mu + 0.1]$ is more than in $[\mu - 0.01, \mu + 0.01]$ as the area in the former interval is more than in latter.
- From table.2 we observe that as the number of samples increases more sample mean fall in the above intervals.

2.3 95% Confidence Interval and Procedure

	Bernoulli			**Poisson**			**Uniform**		
	Class Interval	ϵ	***true mean outside CI***	***Class Interval***	ϵ	***true mean outside CI***	***Class Interval***	ϵ	***true mean outside CI***
N=10	0.2-0.8	0.3	237	3.7-6.4	1.35	490	3.2-6.78	1.77	499
N=100	0.4-0.6	0.09	384	4.56-5.44	0.440	479	4.43-5.56	0.568	500
N=1000	0.46-0.53	0.031	458	4.86-5.14	0.138	491	4.81-5.10	0.180	498

Table 3: Class interval for various value of N for different distribution

Procedure

- To find the confidence interval we replace the unknown parameter with statistics parameter as we don't know the true parameters of the distribution given.
- The following method was used to calculate the confidence interval:
  - Since we observed that the histogram takes a shape of normal distribution which is symmetric along it's mean. Therefore to calculate the 95% confidence we know out of $10,000$ sample means 500 sample means should fall outside. This is divided into 250 to right and left of the distribution respectively.
  - Then we sort the sample mean and take the 249th and 9749th mean which becomes the upper and lower limit between which 95% values fall.
  - To find the ϵ, the difference between the upper limit and lower limit is calculated and then divided by 2 since its symmetric.
- We cross-verified empirical result with the below formulas and it was observed to be similar to our results.

$$CIforBernoulli : \bar{X}_n \pm z\sqrt{\frac{\bar{X}_n\left(1-\bar{X}_n\right)}{N}} \tag{23}$$

$$CIforPoisson : \bar{X}_n \pm z\sqrt{\frac{\bar{X}_n}{N}}$$

$$CIforUniform : \bar{X}_n \pm z\sqrt{\frac{(b-a)^2}{12*N}} \tag{24}$$

Observations

- It is observed that the length of class-interval decreases as we increase the number of sample size.
- This shows that wider the interval, then greater is the uncertainty to find the sample mean.

- It is also observed that the number of sample mean increases as the N increases even though the class interval length decreases.
- This proves that when the sample size increases, we are more certain that the sample mean reflects true mean and hence narrower is the confidence interval.

2.4 Poisson and Binomial distribution

Theorem 1 is not applicable for poisson rvs because it is not bounded as hoeffding inequality is for bounded interval.

2.5 Number of samples versus accuracy

Bernoulli		**Poisson**		**Uniform**	
N	ϵ	**N**	ϵ	**N**	ϵ
100	0.1	1947	0.10	3000	0.10
500	0.043	4000	0.06	10000	0.05
700	0.037	10000	0.04	40000	0.028
3000	0.01	54000	0.01	90000	0.01

Table 4: Empirical results for N for accuracy of 0.1 [green] and 0.01 [yellow]

Observations

- It is observed that as the N increases, then the absolute difference between the sample mean and true mean decreases.

		Bernoulli	**Poisson**	**Uniform**
Accuracy =0.1	**Empirical**	100	1947	3000
	Theoretical	1475	10000	14979
Accuracy =0.01	**Empirical**	3000	54000	90000
	Theoretical	147555	1000000	1497966

Table 5: Comparison of value of N obtained by empirical and theoretical results

Observations

- The empirical value of N is lesser than its theoretical value as the various inequalities used to determine N gives the upper bound value.
- From table. 4 we observe that the as N increases the confidence interval decreases in all the various distribution.

- We used Bernstein inequalities to calculate value of N in Bernoulli distribution, Chebyshev's inequality is used to calculate N in Poisson Distribution while Hoeffding's Inequality is used for Uniform Distribution.

- Bernoulli
$$N = \frac{4ln\left(\frac{2}{\delta}\right)}{\epsilon^2} \tag{25}$$
where 1-δ is confidence and ϵ confidence interval.

- Poisson
$$N = \frac{\lambda}{\delta\epsilon^2} \tag{26}$$

- Uniform
$$N = \frac{(b-a)^2 ln\frac{2}{\delta}}{2\epsilon^2} \tag{27}$$

- From the above formulas we understand that
$$N \propto \frac{1}{\epsilon^2} \tag{28}$$

- **Hence when we change the decimal place 10^{-n} then N will be $x \times 10^{2n}$**

3 Question No 3

3.1 Finding value of A

As we know that $\sum 1/n^2$ where $n = 1$ to $\infty\pi^2/6$
$\sum 1/n^2 = \frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16}.... = \pi^2/6$ Given function $f(k) = \frac{A}{k^2}$ for$k = \pm1, \pm2, \pm3, \pm...$
we want $f(k)_{k\neq0} = \frac{A}{k^2} = 1$ since the function is symmetric about the origin, so we can write $2 * \frac{A}{k^2} = 1 \implies A2 * \pi^2/6 = 1 \implies A = 3/\pi^2$

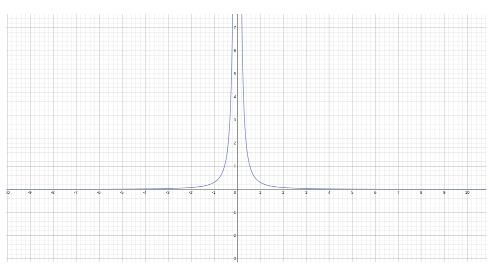


Figure 2: plot of the function $f(k) = \frac{A}{k^2}$

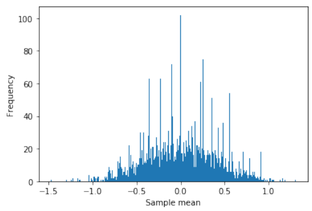


Figure 3: for N = 10

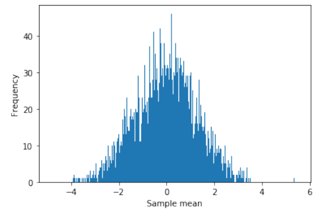


Figure 4: for N = 100

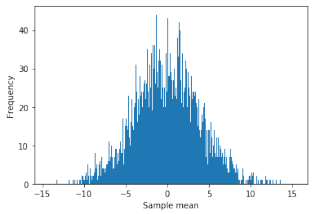


Figure 5: for N = 1000

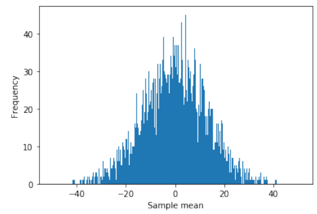


Figure 6: for N = 10000

3.2

$$E[X] = \sum_{i \neq 0}^{\infty} x f(x) \tag{29}$$

$$E[X] = \sum_{i \neq 0}^{\infty} x \frac{3}{\pi^2 x^2}$$

$$E[X] = \sum_{i \neq 0}^{\infty} \frac{3}{\pi^2 x}$$

$$E[X] = \frac{3}{\pi^2} \sum_{i \neq 0}^{\infty} \frac{1}{x}$$

$$E[X] = ... + \frac{1}{-3} + \frac{1}{-2} + \frac{1}{-1} + \frac{1}{1} + \frac{1}{2} + ... = 0$$

$$E[X^2] = \sum_{i \neq 0}^{\infty} x^2 \frac{3}{\pi^2 x^2} \tag{30}$$

$$E[X^2] = \sum_{i \neq 0}^{\infty} \frac{3}{\pi^2} = \infty$$

Means is close to zero as the as the actual mean is zero which is less then infinity. As expected mean is close to zero. No, this behaviour is not accepted, we can see as the sample size is increasing confidence interval is also increasing,but According to central limit theorem confidence interval should decrease, this is happening because one assumption of central limit theorem is not satisfied that is variance should be $var < \infty$ but for the given density function var is ∞

Confidence interval N	means	confidence interval	epsilon
10	-0.004	(-0.78, 0.77)	.775
100	-0.016	(-2.52, 2.44)	2.48
1000	-0.026	(-7.9, 7.8)	7.85
10000	0.0370	(-25.2, 25.4)	25.3