Report: LARP Assignment 2

Keval Dodiya (CS19M023) Karan Jivani (CS19M030)

Problem 1

Part 1:

According to Bernstein bound…

$$P(\overline{Xn} \geq \mu + \varepsilon) \leq e^{-\frac{1}{4}*n*\varepsilon^2} \text{ and } P(\overline{Xn} \leq \mu - \varepsilon) \leq e^{-\frac{1}{4}*n*\varepsilon^2},$$

$$P(\overline{Xn} \geq \mu + \varepsilon) \leq e^{-\frac{1}{4}*n*\varepsilon^2} \Rightarrow P(\overline{Xn} < \mu + \varepsilon) \geq 1 - e^{-\frac{1}{4}*n*\varepsilon^2} \quad \text{-----(1).}$$

$$P(\overline{Xn} \leq \mu - \varepsilon) \leq e^{-\frac{1}{4}*n*\varepsilon^2} \Rightarrow P(\overline{Xn} > \mu - \varepsilon) \geq 1 - e^{-\frac{1}{4}*n*\varepsilon^2} \quad \text{-----(2).}$$

Let (1) is the probability of event A(P(A)) and for (2) is P(B).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Here $P(A \cup B) = 1.$

$$1 \geq 1 - e^{-\frac{1}{4}*n*\varepsilon^2} + 1 - e^{-\frac{1}{4}*n*\varepsilon^2} - P(A \cap B).$$

$$\Rightarrow P(A \cap B) \geq 1 - 2\, e^{-\frac{1}{4}*n*\varepsilon^2}.$$

So, $P(\mu \in [\overline{X_n} - \varepsilon, \overline{X_n} + \varepsilon]) \geq 1 - \delta$, where $\delta = 2\, e^{-\frac{1}{4}*n*\varepsilon^2}$.

Hence, we get $\varepsilon = \sqrt{\frac{4}{n} * log(2/\delta).}$

Part 2:

Given $P(\overline{X_n} \geq \mu + \varepsilon) \leq e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2}$ and $P(\overline{X_n} \leq \mu - \varepsilon) \leq e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2}$,

$$P(\overline{X_n} \geq \mu + \varepsilon) \leq e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2} \Rightarrow P(\overline{X_n} < \mu + \varepsilon) \geq 1 - e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2} \quad \text{(3).}$$

$$P(\overline{X_n} \leq \mu - \varepsilon) \leq e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2} \Rightarrow P(\overline{X_n} > \mu - \varepsilon) \geq 1 - e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2} \quad (4).$$

Let (1) is the probability of event A(P(A)) and for (2) is P(B).

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Here, $P(A \cup B) = 1.$

$$1 \geq 1 - e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2} + 1 - e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2} - P(A \cap B).$$

$$\Rightarrow P(A \cap B) \geq 1 - 2e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2}.$$

So, $P(\mu \in [\overline{X_n} - \varepsilon, \overline{X_n} + \varepsilon]) \geq 1 - \delta$, where $\delta = 2e^{-(\frac{2}{(b-a)^2})*n*\varepsilon^2}$.

Therefore, we get $\varepsilon = \sqrt{(\frac{(b-a)^2}{2n}) * log(2/\delta).}$ (5).

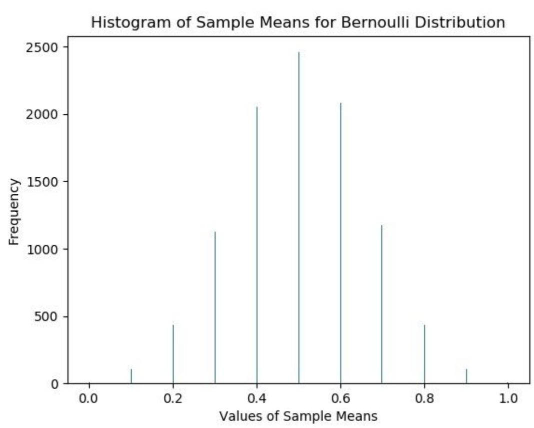
Problem 2

(a,b,c)

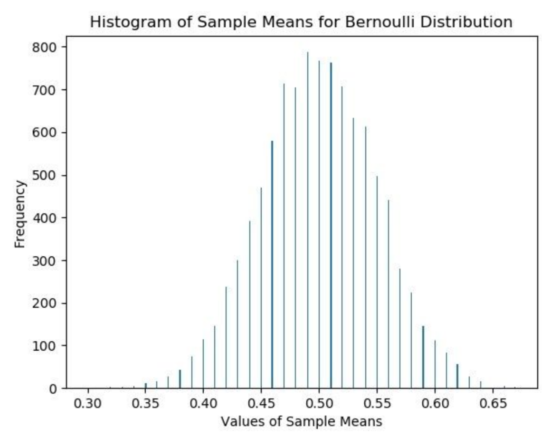
Bernoulli Distribution:

Parameters: $p = 0.5$

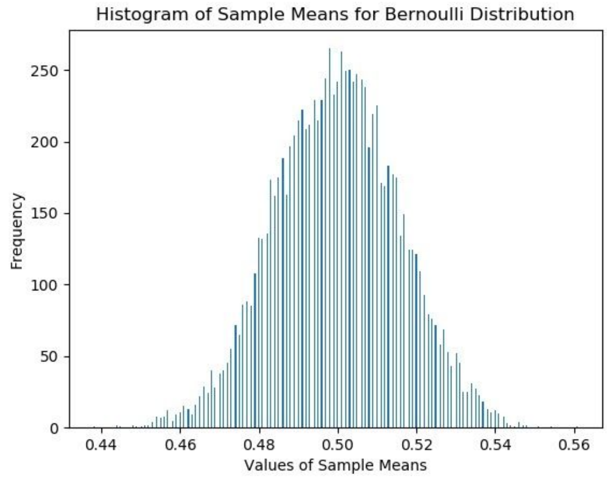
N	Number of Times value of sample mean was between [μ - 0.01, μ + 0.01]	Number of Times value of sample mean was between [μ - 0.1, μ + 0.1]	Number of Times value of true mean outside of Confidence Interval	Confidence Interval (95%)
10	2459	6600	210	[0.5 - 0.3000, 0.5 + 0.3000]
100	2317	9625	363	[0.5 - 0.1000, 0.5 + 0.1000]
1000	4887	10000	474	[0.5 - 0.0310, 0.5 + 0.0310]



(a) n = 10



(b) n = 100



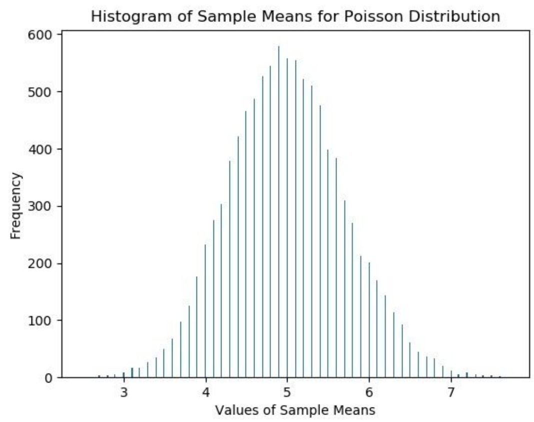
(c) n = 1000

Figure 2.1: Histogram for Sample Means of Bernoulli Distribution for n = 10, n = 100, and n = 1000

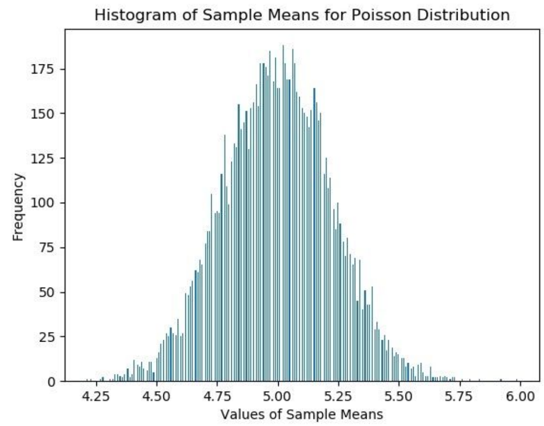
Poisson Distribution:

Parameters: $\lambda = 5$

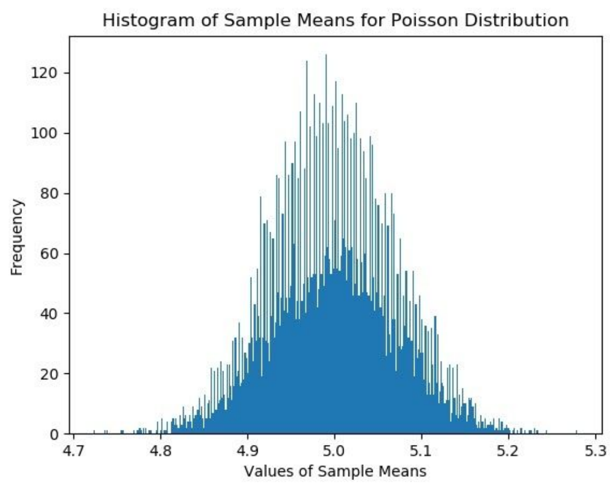
N	Number of Times value of sample mean was between [μ - 0.01, μ + 0.01]	Number of Times value of sample mean was between [μ - 0.1, μ + 0.1]	Number of Times value of true mean outside of Confidence Interval	Confidence Interval (95%)
10	558	1691	389	[5 - 1.40099, 5 + 1.40099]
100	509	3583	479	[5 - 0.4400, 5 + 0.4400]
1000	1195	8411	493	[5 - 0.1380, 5 + 0.1380]



(a) n = 10



(b) n = 100



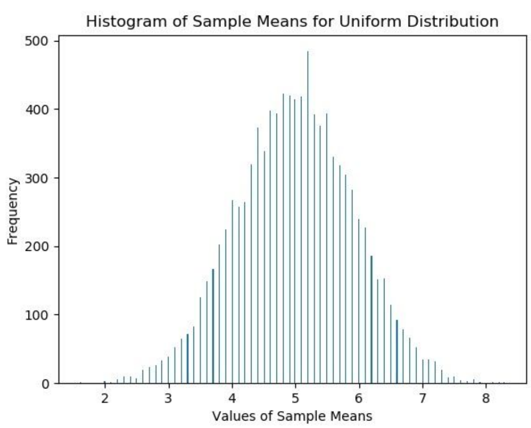
(c) n = 1000

Figure 2.2: Histogram for Sample Means of Poisson Distribution for n = 10, n = 100, and n = 1000

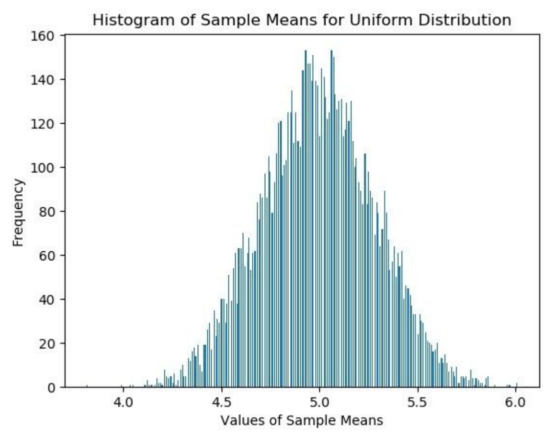
Uniform Distribution:

Parameters: $a = 0, \ b = 10$

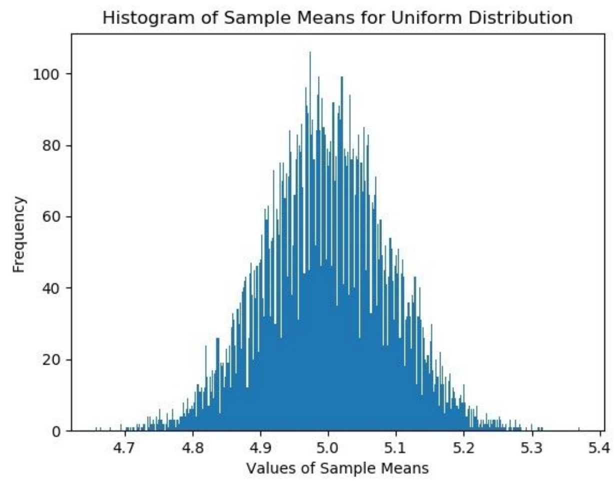
N	Number of Times value of sample mean was between [μ - 0.01, μ + 0.01]	Number of Times value of sample mean was between [μ - 0.1, μ + 0.1]	Number of Times value of true mean outside of Confidence Interval	Confidence Interval (95%)
10	414	1252	470	[5 - 1.80099, 5 + 1.80099]
100	396	2881	479	[5 - 0.5700, 5 + 0.5700]
1000	875	7221	500	[5 - 0.1790, 5 + 0.1790]



(a) n = 10



(b) n = 100



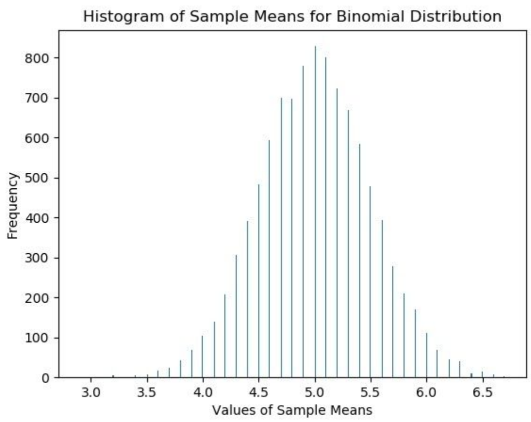
(c) n = 1000

Figure 2.3: Histogram for Sample Means of Uniform Distribution for n = 10, n = 100, and n = 1000

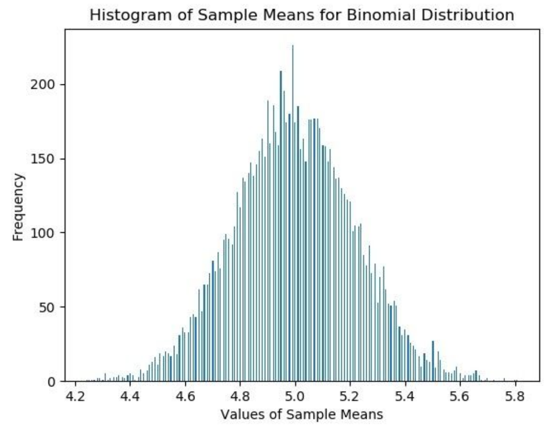
Binomial Distribution: Approximating Poisson Distribution with Binomial Distribution,

Parameters: $n =$ n, $b = \frac{\lambda}{n}$

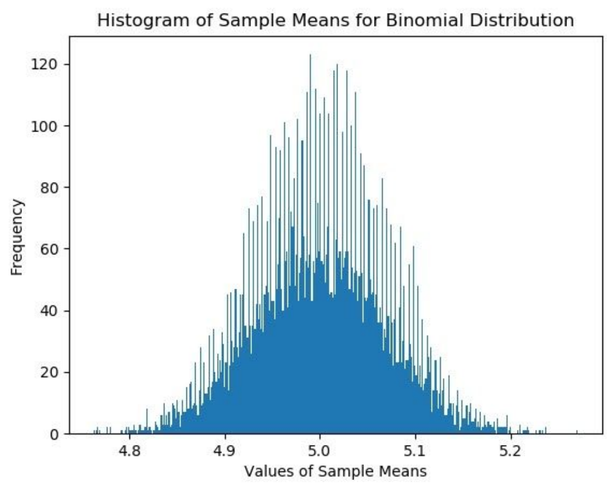
N	Number of Times value of sample mean was between [μ - 0.01, μ + 0.01]	Number of Times value of sample mean was between [μ - 0.1, μ + 0.1]	Number of Times value of true mean outside of Confidence Interval	Confidence Interval (95%)
10	828	2408	381	[5 - 1.0000, 5 + 1.0000]
100	585	3707	481	[5 - 0.4200, 5 + 0.4200]
1000	1179	8505	499	[5 - 0.1360, 5 + 0.1360]



(a) n = 10



(b) n = 100



(c) n = 1000

Figure 2.4: Histogram for Sample Means of Binomial Distribution for n = 10, n = 100, and n = 1000

(a)

- Yes, sample mean is very close to the true mean, because by weak law of large numbers (WLLN) $P(|\overline{X_n} - \mu| \geq \varepsilon)$ becomes zero when n approaches infinity. In other words sample mean converges to its true mean with growth of n.

	N=10 ($\overline{X}_N$)	N=100 ($\overline{X}_N$)	N=1000 ($\overline{X}_N$)
Bernoulli dist.	0.4987	0.5002	0.4999
Poisson dist.	5.007	5.001	4.999
Uniform dist.	5.025	5.002	5.000

Here we can observe that sample means are very close to true means.

(c) Procedure for confidence interval:

- We took true mean and fixed step size as 0.001 (ε). Then increase step size by 0.001 for every iteration and check confidence with interval [μ - step size, μ + step size] each time. And do this procedure until we got 95% confidence.

(D)

- Theorem (1) doesn't work for Poisson Distribution, as acc. to conditions in Theorem (1), $-\infty < a \leq b < \infty$, but in the case of Poisson Distribution, b is not bounded.
- When we approximate poisson r.v. as binomial r.v. with parameter N, $\frac{\lambda}{N}$
  $P(X_N = k) \rightarrow \frac{e^{-\lambda}\lambda^k}{k!}$ as N approaches infinity.
- And for bernoulli trials a=0,b=1 and 95% confidence is given hence $\delta = 0.05$. we put these values into (5) as follows
  $$\varepsilon = \sqrt{(\frac{(b-a)^2}{2n}) * log(2/\delta)} = \sqrt{\frac{1}{2n} * log(40).}$$

	N=10	N=100	N=1000
ε (theoretical)	0.4294	0.1358	0.0429
ε (numerical of bernoulli $\mu = 0.5$)	0.3000	0.1000	0.0310
ε (numerical of poisson $\lambda = 5$)	1.4009	0.4400	0.1380

So, here we can conclude that we are getting closer bound than Hoeffding's bound(bernoulli). And in Poisson dist. For large value of n this Hoeffding bound will become closer to our theoretical epsilon otherwise Hoeffding's inequality does not hold.

(E)

- By equation (5) we calculated N with $\varepsilon = 0.1, 0.01$.

  Bernoulli Distribution:
  Accuracy = 0.1, n = 184, n = 94(empirical).
  Accuracy = 0.01, n = 18444, n = 9498(empirical).
  Poisson Distribution:
  Accuracy = 0.1, n = 184, n = 1933(empirical).
  Accuracy = 0.01, n = 18444, n = 193330(empirical).
  Uniform Distribution:
  Accuracy = 0.1, n = 18444, n = 3664(empirical).
  Accuracy = 0.01, n = 1844430, n = 364333(empirical).

- As it can be seen from the above data, that, for Bernoulli and Uniform Distribution, practical value of n is upper bounded by theoretical values of n, but for Poisson Distribution practical values of n are not upper bounded by theoretical values of n, since the Theorem (1) doesn't hold for Poisson Distribution as stated before.

- So we can observe that if accuracy increase by decimal places k = k * 10^{-1} then N increases by a factor of 100.

Problem 3

Part 1: Value of A to make the $f(k) = \frac{A}{k^2}$ a valid probability mass function is $\frac{3}{\pi^2}$.

Proof: To make $f(k) = \frac{A}{k^2}$ a valid probability mass function,

$$\sum_k f(k) = \sum_k \frac{A}{k^2} = 1 \quad \Rightarrow \quad A \sum_k \frac{1}{k^2} = 1 \quad \Rightarrow \quad A = \frac{1}{\sum_k \frac{1}{k^2}},$$

where $k \in \{\pm 1, \pm 2, \pm 3, \ldots\}$

Now, we need to find the value of $\sum_k \frac{1}{k^2}$.

We know that Taylor series expansion of $\sin(x)$ is given as,

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

Therefore,

$$\frac{\sin(x)}{x} = 1 - \frac{x^2}{3!} + \frac{x^4}{5!} - \frac{x^6}{7!} + \ldots \qquad \text{--- (3.1)}$$

Note that $\frac{\sin(x)}{x}$ is 0 at $x = \pm n\pi$, for n = 1, 2, 3, … and is 1 at $x = 0$. Thus, it can be also written as,

$$\frac{\sin(x)}{x} = (1 - \frac{x}{\pi})(1 + \frac{x}{\pi})(1 - \frac{x}{2\pi})(1 + \frac{x}{2\pi})\ldots$$

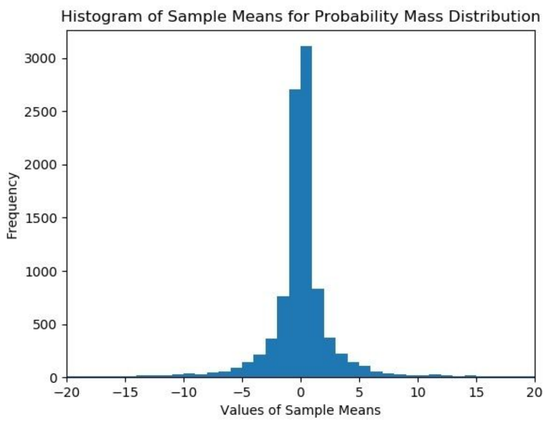
$$\frac{\sin(x)}{x} = (1 - \frac{x^2}{\pi^2})(1 - \frac{x^2}{4\pi^2})(1 - \frac{x^2}{9\pi^2})\ldots \qquad \text{--- (3.2)}$$

Comparing the coefficients of x^2 in Eq. (3.1) and Eq. (3.2), we get,

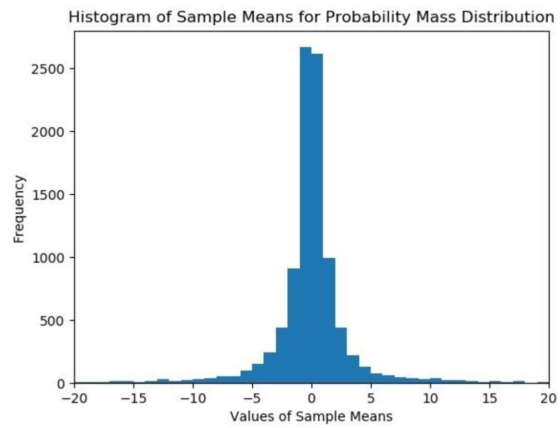
$$-\left(\frac{1}{\pi^2} + \frac{1}{4\pi^2} + \frac{1}{9\pi^2} + \ldots\right) = -\frac{1}{3!} \quad \Rightarrow \quad \sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{6}$$

Therefore, $\sum_{k \neq 0} \frac{1}{k^2} = 2\sum_{k=1}^{\infty} \frac{1}{k^2} = \frac{\pi^2}{3}$ and $A = \frac{3}{\pi^2}$.

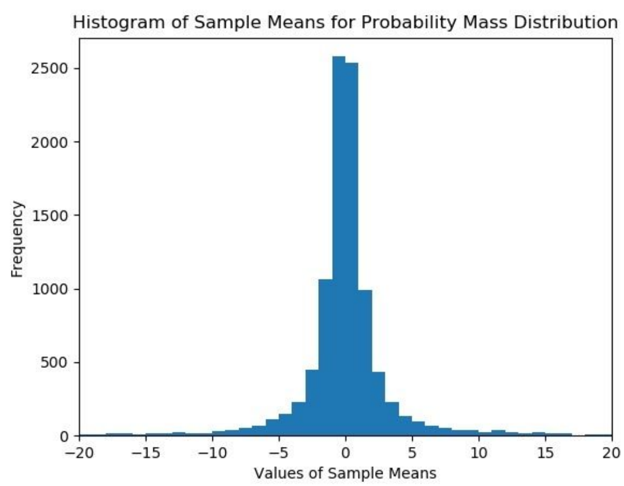
Part 2:



(b) n = 10



(b) n = 100



(c) n = 1000

Figure 3.1: Histogram for Sample Means of Probability Mass Distribution for n = 10, n = 100, and n = 1000

- Sample Mean Concentrates around 0, and this behaviour is expected since Theoretical Expected Value is also 0, which can be calculated as,

$$E[X] = \sum_{k=-\infty}^{\infty} kf(k) = \sum_{k=-\infty}^{-1} k\left(\frac{A}{k^2}\right) + \sum_{k=1}^{\infty} k\left(\frac{A}{k^2}\right)$$

$$E[X] = \left[-\sum_{k=1}^{\infty} k\left(\frac{A}{k^2}\right)\right] + \left[\sum_{k=1}^{\infty} k\left(\frac{A}{k^2}\right)\right] = 0$$

- From the below results, it can be seen that Confidence Interval for n = 10, n = 100 and n = 1000 remains almost the same.
  - 95% Confidence Interval for n = 10, is [0 - 12.5009, 0 + 12.5009]
  - 95% Confidence Interval for n = 100, is [0 - 12.7600, 0 + 12.7600]
  - 95% Confidence Interval for n = 1000, is [0 - 12.3839, 0 + 12.3839]