CSE 544.01 Probability and Statistics for Data Scientists

Assignment - 3 Team Members

Akhil Arradi - 114353508

Sai Vikas Balabadhrapatruni - 114777850

Varshith Adavala - 114778433

Akhila Juturu - 114777498

① Show that

$$MSE = bias^2(\hat{\theta}) + var(\hat{\theta})$$

$\hat{\theta}$ is estimator of θ ← true parameter of unknown distribution X

W.K.T

MSE = Mean Squared Error = $MSE(\hat{\theta}) = E((\theta-\hat{\theta})^2)$

$$MSE(\hat{\theta}) = E[(\theta-\hat{\theta})^2]$$

~~LHS~~ Let $E(\hat{\theta})$ be a expectation of Random Variable

$\because E[\hat{\theta}]$ is a constant

$$E[\hat{\theta} - E(\hat{\theta})) \overset{LOE}{=} E(\hat{\theta}) - E(\hat{\theta}) = 0$$

Consider LHS

$$E((\theta-\hat{\theta})^2) = E((\theta - E(\hat{\theta}) + E(\hat{\theta}) - \hat{\theta})^2] \quad \begin{bmatrix}\text{Adding \&} \\ \text{subtracting} \\ E(\hat{\theta})\end{bmatrix}$$

$$= E((\theta - E(\hat{\theta}))^2 + (E(\hat{\theta}) - \hat{\theta})^2 + 2((\theta - E(\hat{\theta}))(E(\hat{\theta}) - \hat{\theta}))$$

$$\underline{\underline{LOE}}\ E((\theta - E(\hat{\theta}))^2] + E[(E[\hat{\theta}] - \hat{\theta})^2]$$

$$+ 2(\theta - E[\hat{\theta}]) E(E(\hat{\theta}) - \hat{\theta}) \quad -①$$

[$\because \theta$ is constant, $\theta - E[\hat{\theta})$ is constant, θ is true parametric constanst.

W.K.T

$$bias(\hat{\theta}) = E[\hat{\theta}] - \theta \quad \& \quad Var(\hat{\theta}) = E\left((\hat{\theta} - E[\hat{\theta}])^2\right]$$

From ①

$$E\left((\hat{\theta}-E[\hat{\theta}])^2\right) + 2(\theta - E[\hat{\theta}])(0)$$

$$= \left(E[\hat{\theta}] - \theta\right]^2 + E\left((\hat{\theta} - E(\hat{\theta}])^2\right)$$

$$= bias^2(\hat{\theta}) + Var(\hat{\theta})$$

$$= R.H.S$$

$$\therefore \; MSE = bias^2(\hat{\theta}) + Var(\hat{\theta})$$

Hence proved.

Another way of solving

Consider L.H.S

$$MSE = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 + \theta^2 - 2\hat{\theta}\theta]$$

$$= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2 \quad \text{— ①}$$

R.H.S

$$bias^2(\hat{\theta}) + Var(\hat{\theta}) = (E(\hat{\theta}) - \theta)^2 + E[(\hat{\theta} - E[\hat{\theta}])^2]$$

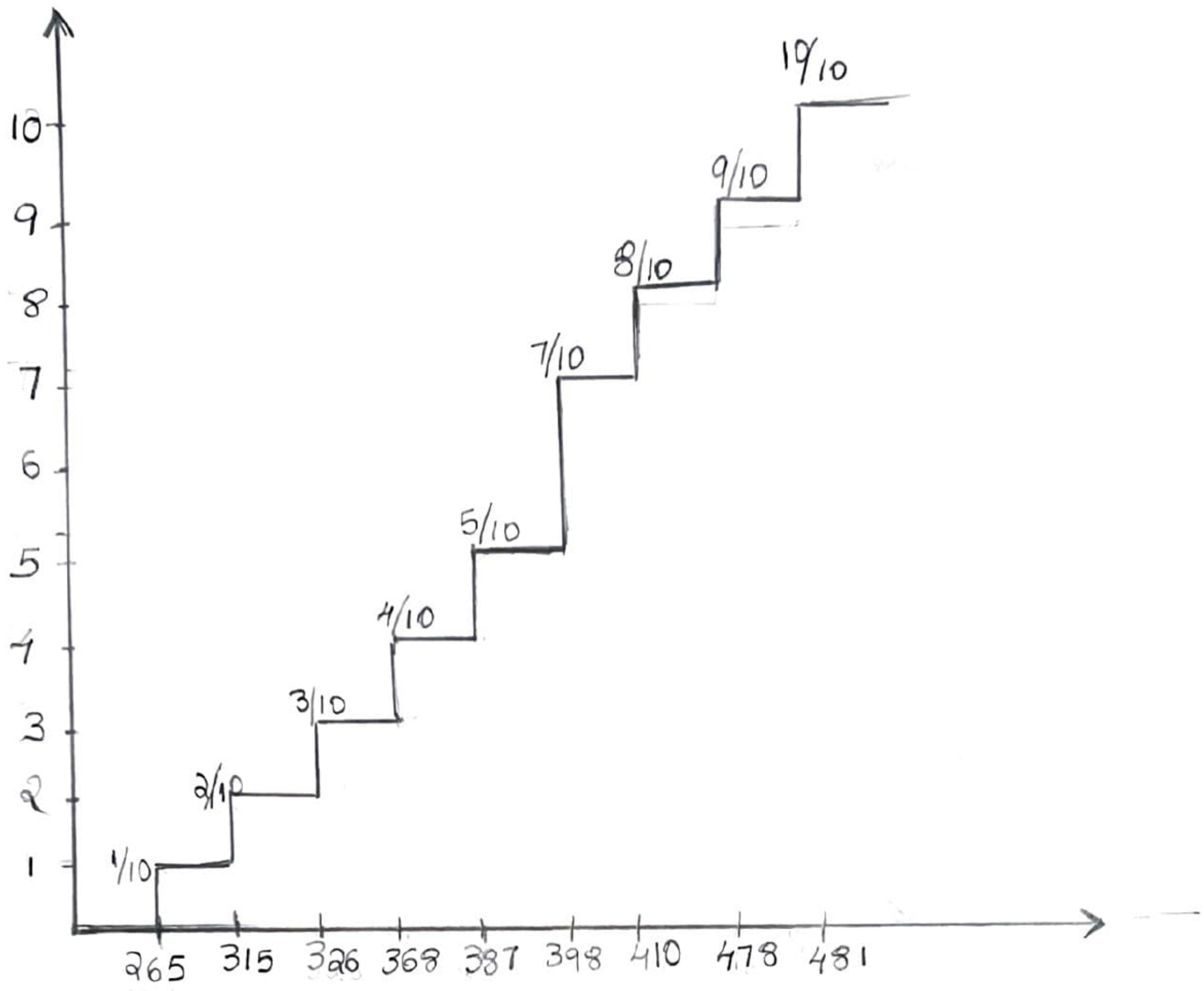
$$= \cancel{E[\hat{\theta}]^2} + \theta^2 - 2E[\hat{\theta}]\theta + E[\hat{\theta}^2] - \cancel{(E[\hat{\theta}])^2}$$

$$E[\hat{\theta}^2] - 2E[\hat{\theta}]\theta + \theta^2 \quad \text{— ②}$$

From ① & ② L.H.S = R.H.S

$$\therefore \; MSE = bias^2(\hat{\theta}) + Var(\hat{\theta})$$

2)



D = {265, 315, 326, 368, 387, 398, 398, 410, 478, 481}

Q) 3 a) Given $D = \{x_1, x_2, \dots x_n\}$ be a set of i.i.d.

$\hat{F}$ is the eCDF

$$\hat{F}(\alpha) = \sum_{i=1}^{n} \frac{I(x_i \le \alpha)}{n}$$

eCDF

$$E(\hat{F}(\alpha)) = E\left(\sum_{i=1}^{n} \frac{I(x_i \le \alpha)}{n}\right)$$

$$\stackrel{LOE}{=} \frac{1}{n} \sum_{i=1}^{n} E(I(x_i \le \alpha)) \stackrel{iid}{=} \frac{1}{\cancel{n}} (\cancel{n}) E(I(x_j \le \alpha))$$

$$= E(I(x \le \alpha))$$

We know $E(I(x)) = P_r(x)$

$$= P_r(x \le \alpha)$$

$$= F(\alpha)$$

$\therefore E[\hat{F}(\alpha)] = F(\alpha)$

3 (b) $\text{Bias}(\hat{F}) = 0$

$$\left[\ \text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta\ \right]$$

From (a) we know that $E[\hat{F}(\alpha)] = F$

$$\therefore \text{bias}(\hat{F}) = \hat{F} - F = 0$$

$$\left[\text{bias}(\hat{F}) = 0\right]$$

(c) $se(\hat{F}) = \sqrt{Var(\hat{F})}$

$$Var(\hat{F}(\alpha)) = Var\left(\sum_{i=1}^{n} \frac{I(x_i \le \alpha)}{n}\right)$$

$$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} I(x_i \le \alpha)\right) \underset{\text{i.i.d}}{\overset{\text{Lov}}{=}} \frac{1}{n^2} \sum_{i=1}^{n} Var(I(x_i \le \alpha))$$

$$\underset{\text{iid}}{=} \frac{1}{n^2} \cdot n\, Var(I(\cancel{x} \le \alpha))$$

$$= \frac{1}{n} Var(I(X \le \alpha))$$

The distribution is in the form of ~~binomial~~ bernoulli distribution.

[W.K.T
Variance of ~~binomial distribution~~ indicator RV is $P(1-P)$]

$$= \frac{1}{n} Var(I(X \le \alpha))$$

$$= \frac{1}{n} P_r(X \le \alpha)(1 - P_r(X \le \alpha))$$

$$= \frac{1}{n} F(\alpha)(1 - F(\alpha))$$

$$Var(\hat{F}(\alpha)) = \frac{1}{n} F(\alpha)(1 - F(\alpha))$$

$$\boxed{\Rightarrow se(\hat{F}) = \sqrt{\frac{F(1-F)}{n}}}$$

(d) If $bias(\hat{\theta}) \xrightarrow{n \to \infty} 0$ and $sec(\hat{\theta}) \xrightarrow{n \to \infty} 0$ then $\hat{\theta}$ is a ~~Constant~~ Consistent estimator of θ

From (b) $bias(\hat{F}) = 0$ (irrespective of n)

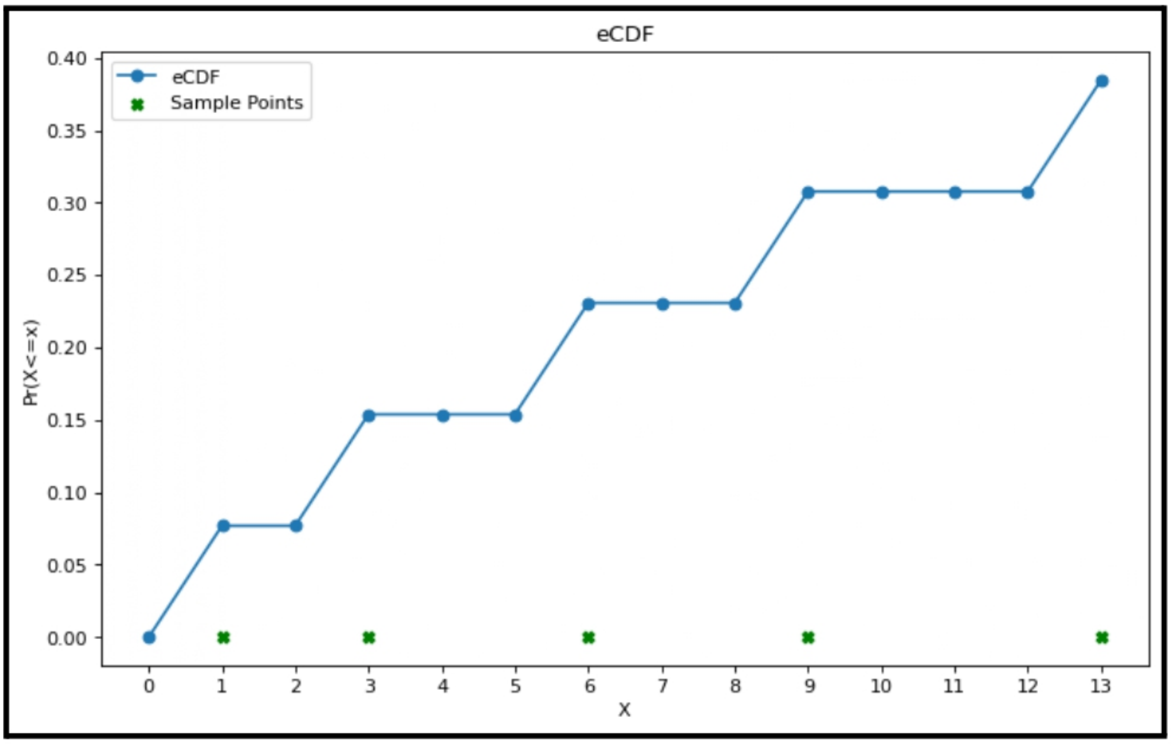
From (c) $se(\hat{F}) = \sqrt{\frac{1}{n} F(1-F)}$ as $n \to \infty$ $\frac{1}{n} = 0$

$$\therefore se(\hat{F}(\alpha)) \to 0 \quad \text{as } n \to \infty$$

$\therefore \hat{F}$ is a consistent estimator of F.

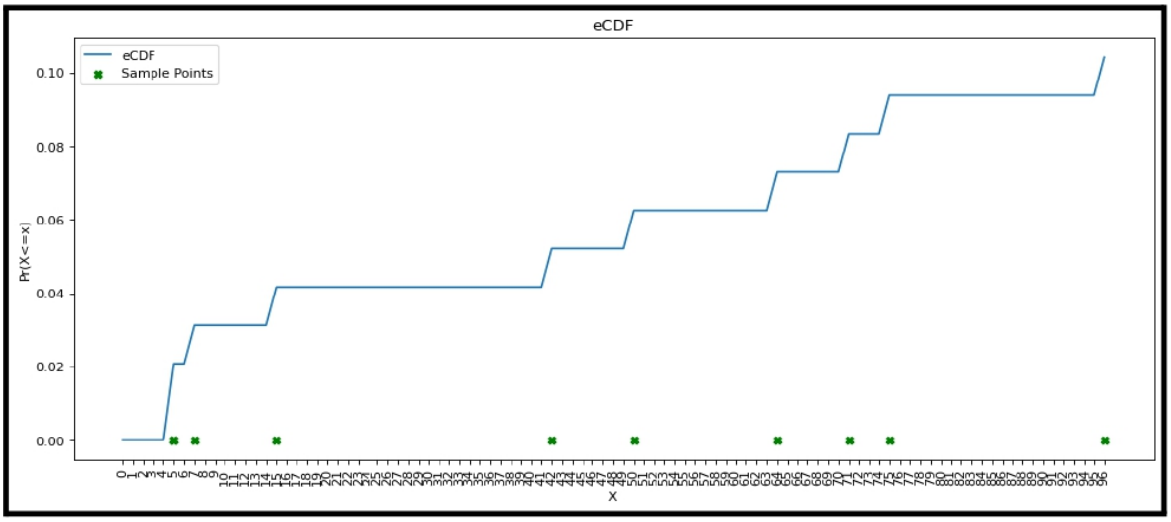
4

A)

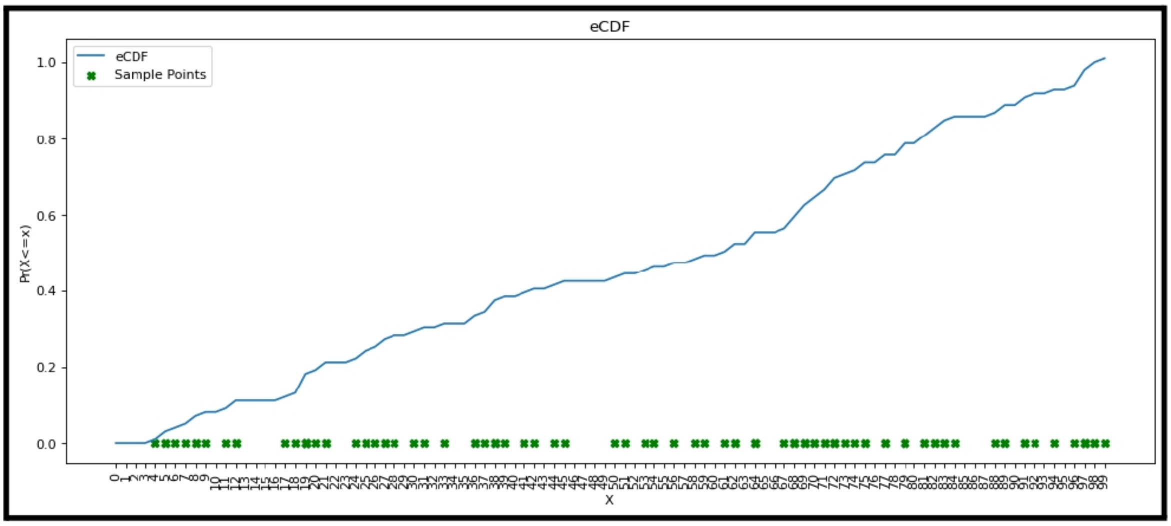

eCDF
eCDF
Sample Points
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Pr(X<=x)
0
1
2
3
4
5
6
7
8
9
10
11
12
13
X

B)

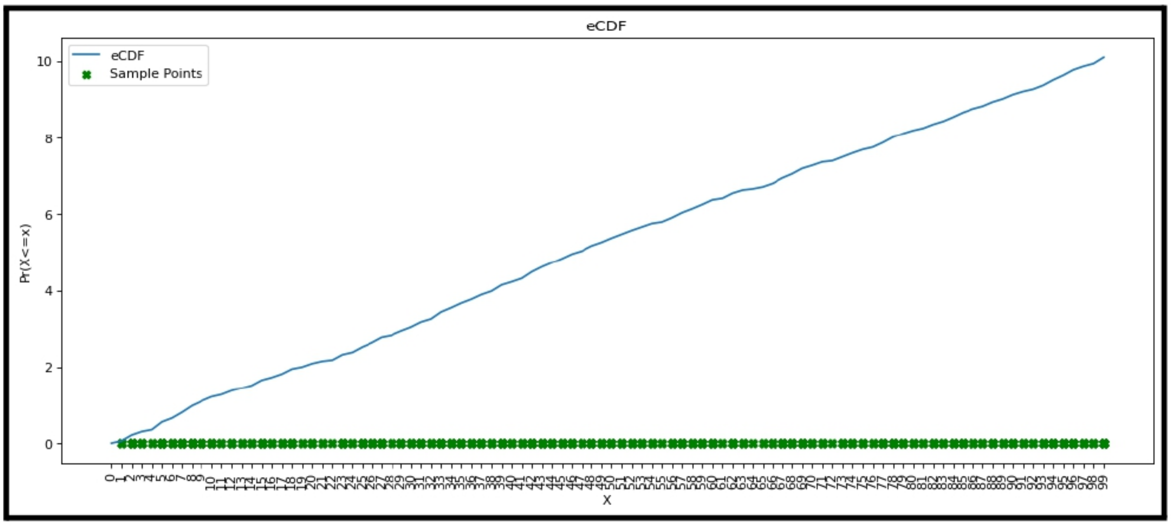
N=10


eCDF
eCDF
Sample Points
0.10
0.08
0.06
0.04
0.02
0.00
Pr(X<=x)
X

N=100

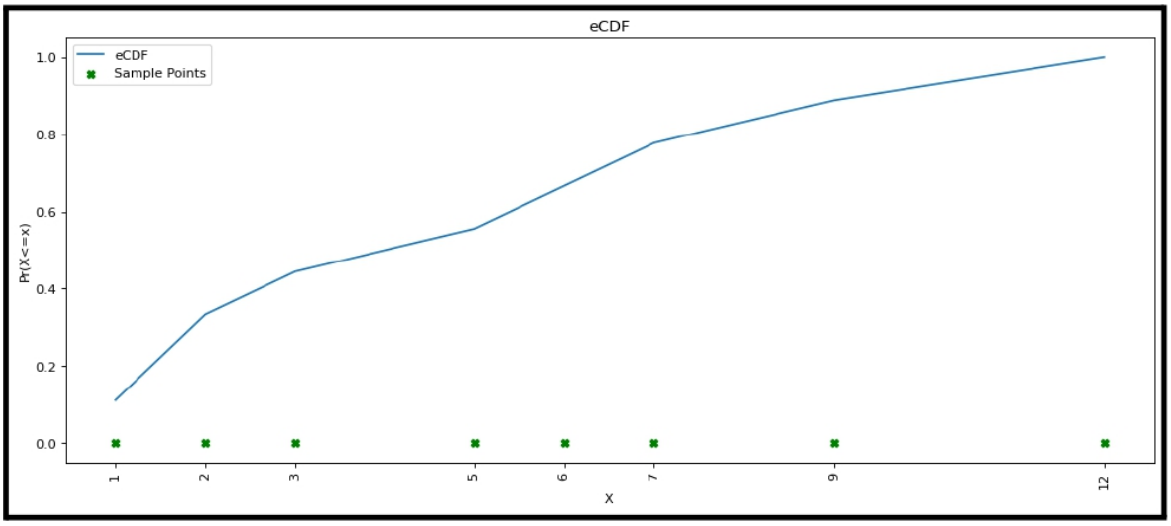


N=1000



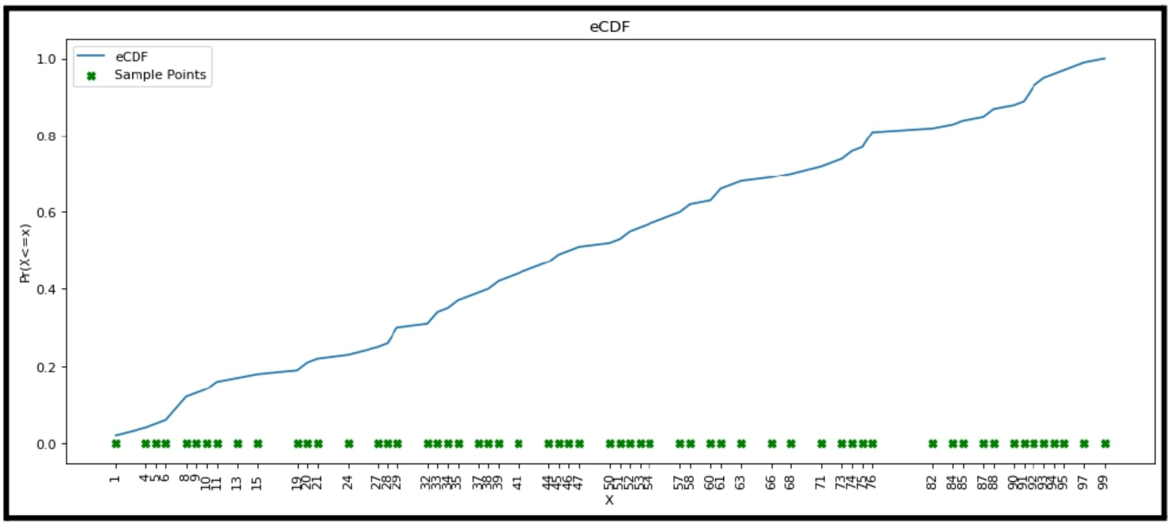
As the N value increases, the CDF becomes a straight line (The steps decrease in the CDF).

C)

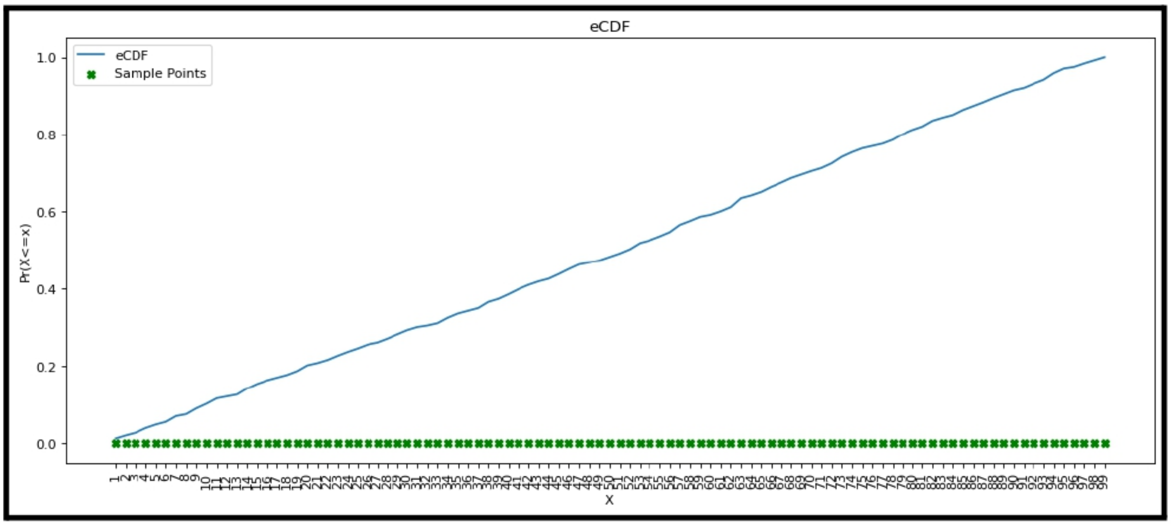

eCDF
eCDF
Sample Points
Pr(X<=x)
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
5
6
7
9
12
X

D)

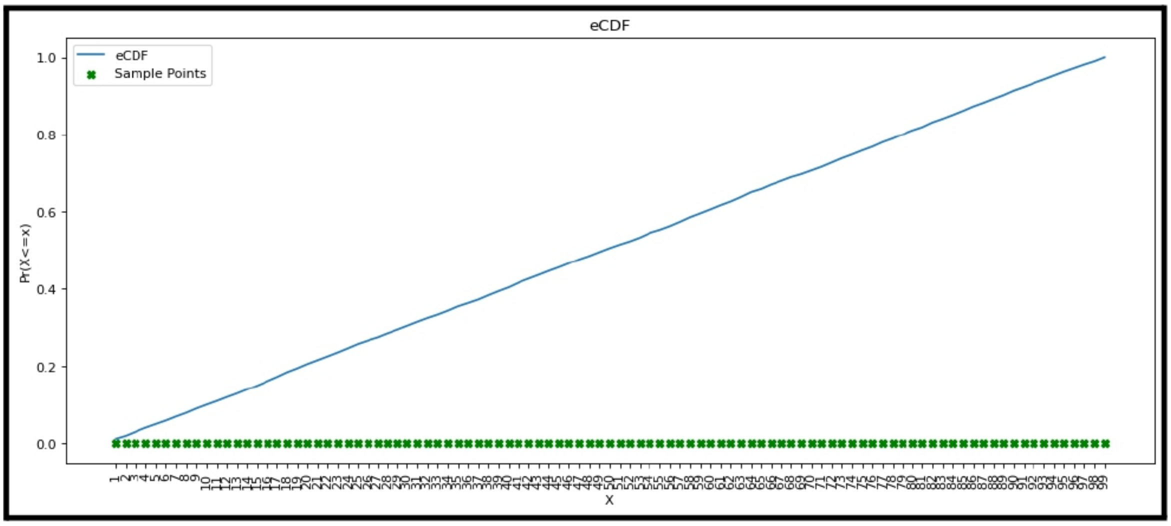
N=10, M=10


eCDF
eCDF
Sample Points
Pr(X<=x)
1.0
0.8
0.6
0.4
0.2
0.0
X

N=10, M=100

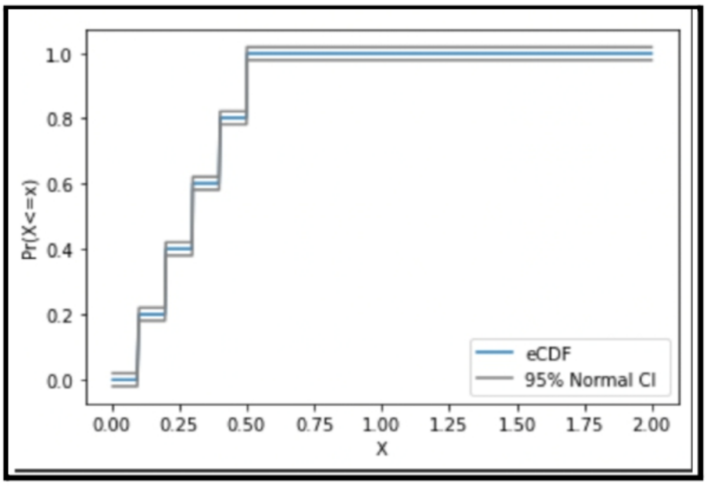


N=10, M=1000

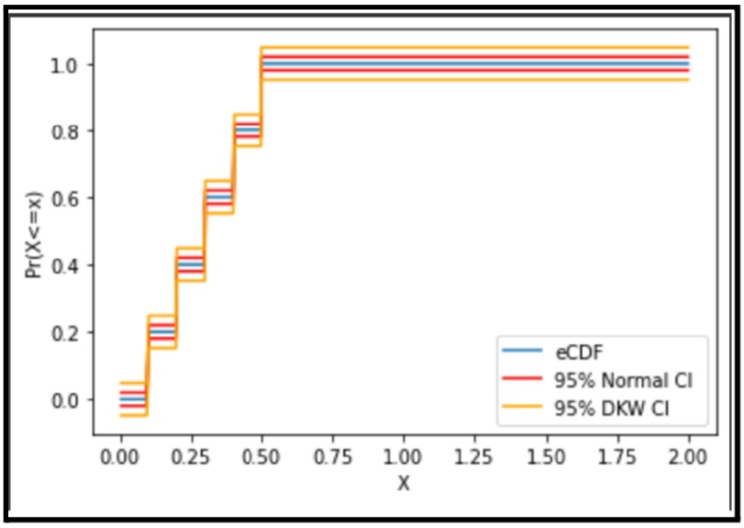


As the M value increases, the CDF becomes a straight line, the steps disappear as all the values are picked at least once in the M values.

E)



F)



After observing both the Intervals, We can conclude that the Normal-Based Confidence Intervals are tighter than compared to the DKW Confidence Intervals

5 (a) Show plug-in estimator of the variance of x is $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$, where $\overline{X}_n$ is the sample mean $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$

$$\sigma^2 = E[X^2] - (E[X])^2$$

$$= \sum_i X_i^2 P_r(X_i) - \left(\sum_i X_i P_r(X_i)\right)^2$$

$D \stackrel{i.i.d}{=} \{X_1, X_2 \dots X_n\}$ $\quad \hat{P}_x(\alpha) = \begin{cases} \frac{1}{n} & \text{if } x \in D \\ 0 & \text{otherwise} \end{cases}$

Plugin estimate

$$\hat{\sigma}^2 = \sum_{x \in D} x^2 \hat{P}(x) - \left(\sum_{x \in D} x_i \hat{P}(x)\right)^2$$

$$\hat{\sigma}^2 = \sum_{i=1}^{n} (x_i)^2 \left(\frac{1}{n}\right) - \left(\sum_{i=1}^{n} x_i \left(\frac{1}{n}\right)\right)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\sum_{i=1}^{n} x_i$$

[Given $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$)

$$= \frac{1}{n}\sum_{i=1}^{n} (X_i)^2 - \overline{X}_n^2 \quad \to (1)$$

R.H.S

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}\left(x_i^2 + (\bar{x}_n)^2 - 2x_i\bar{x}_n\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{n}\sum_{i=1}^{n}\bar{x}_n^2 - \frac{1}{n}(2)\sum_{i=1}^{n}x_i\bar{x}_n$$

$\bar{x}_n$ is a constant

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \frac{1}{\cancel{n}}\bar{x}_n^2(\cancel{n}) - \frac{2(\bar{x}_n)}{n}\sum_{i=1}^{n}x_i$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}_n^2 - 2\bar{x}_n\sum_{i=1}^{n}\frac{x_i}{n}$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}x_i^2 + \bar{x}_n^2 - 2\bar{x}_n(\bar{x}_n) \quad \left[\text{Given } \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n}x_i\right]$$

$$= \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}_n^2 \rightarrow 2$$

$1 = 2$

$$\therefore \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2$$

5 (b) $\text{bias}(\hat{\sigma}^2) = E[\hat{\sigma}^2] - \sigma^2$

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2\right]$$

$$= E\left[\frac{1}{n}\left(\sum_{i=1}^{n}(X_i^2 + \bar{X}_n^2 - 2X_i\bar{X}_n)\right)\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n}X_i^2 + \sum_{i=1}^{n}\bar{X}_n^2 - 2X_n\sum_{i=1}^{n}X_i\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n}X_i^2 + \bar{X}_n^2(n) - 2\bar{X}_n(n)(\bar{X}_n)\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n}X_i^2 - n\bar{X}_n^2\right]$$

$$= \frac{1}{n} E\left[\sum_{i=1}^{n}X_i^2 - n\bar{X}_n^2\right] \quad — (1)$$

$$\overset{LOE}{=} \frac{1}{n}\left(E\left[\sum_{i=1}^{n}X_i^2\right] - E\left[n\bar{X}_n^2\right]\right)$$

$\left[\cancel{E[X_n^2] = Var[X_n] + (E[X_n])^2}\right.$

$\left[E[X^2] = Var[X] + (E[X])^2\right]$

Consider $E[\bar{X}_n^2] = Var[\bar{X}_n] + (E[\bar{X}_n])^2$

$$= Var\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right] + \left(E\left[\frac{1}{n}\sum_{i=1}^{n}X_i\right]\right)^2$$

$$\overset{LOV}{\underset{LOE}{=}} \frac{1}{n^2}\sum_{i=1}^{n}Var(X_i) + \left(\frac{1}{n}\sum_{i=1}^{n}E[X_i]\right)^2$$

$$\underset{=}{\text{i.i.d}} \quad \frac{1}{n^2} \cancel{n} \, Var[x] + \left(\frac{1}{\cancel{n}} (\cancel{n}) E[x] \right)^2$$

$$E[\bar{X}_n^2] = \frac{1}{n} Var[x] + (E[x])^2 \qquad - ②$$

Substitute ② in ①

$$= \frac{1}{n} \left(\sum_{i=1}^{n} E[X_i^2] - n \, E[\bar{X}_n^2] \right)$$

$$= \frac{1}{n} \left(\sum_{i=1}^{n} E[X_i^2] - n \left(\frac{1}{n} Var[x] + (E[x])^2 \right) \right)$$

$$\underset{=}{\text{i.i.d}} \quad \frac{1}{n} (\cancel{n}) E[X^2] - \frac{Var[x]}{n} + (E[x])^2$$

$$= E[X^2] - (E[X])^2 - \frac{Var[X]}{n}$$

$$\therefore E[\hat{\sigma}^2] = \sigma^2 - \frac{\sigma^2}{n} \qquad \rightarrow ③$$

$$bias(\hat{\sigma}^2) = E[\hat{\sigma}^2] - \sigma^2$$

$$= \sigma^2 - \frac{\sigma^2}{n} - \sigma^2$$

$$= -\frac{\sigma^2}{n}$$

$$\therefore \text{bias}(\hat{\sigma}^2) = -\frac{\sigma^2}{n}$$

Hence proved.

5(c) $\text{Kurt}[X] = \frac{E[(X-\mu)^4]}{\sigma^4}$

Let $\hat{\sigma}^2, \hat{\mu}$ be the plugin estimator for σ^2 & μ

From 5(a) we know that

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^2 \quad — (1)$$

$$\& \ \hat{\mu} = \bar{X}_n$$

$\left[E[X^i] = \sum_{i=0}^{n} x^i P(x)\right]$

Consider $E[(X-\mu)^4]$

$$= \sum_{i=1}^{n} (x_i - \mu)^4 . P(X_i)$$

plugin estimator $= \sum_{i=1}^{n} (X_i - \hat{\mu})^4 \hat{P}(X_i)$ $\left[\text{W.K.T } \hat{P}(X_i) = 1/n\right]$

$$= \sum_{i=1}^{n} (X_i - \hat{\mu})^4 \left(\frac{1}{n}\right) \rightarrow (2)$$

$$= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X}_n)^4$$

From eq (1)

$$\hat{\sigma}^4 = \frac{1}{n^2} \left(\sum_{i} (X_i - \bar{X}_n)^2\right)^2 \quad — (3)$$

From eq ② & ③

$$\text{Kurt}[\hat{X}] = \frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X}_n)^4}{\frac{1}{n^2}\left(\sum_{i=1}^{n}[X_i - \overline{X}_n]^2\right)^2}$$

$$= \frac{n\sum_{i}(X_i - \overline{X}_n)^4}{\left(\sum_{i}(X_i - \overline{X}_n)^2\right)^2}$$

5(d) $$\rho = \frac{E[XY] - E[X]E[Y]}{\sigma_X \sigma_Y}$$

$$\sigma_X = E[X^2] - (E[X])^2$$

$$\sigma_Y = E[Y^2] - (E[Y])^2$$

n iid

$D: \{ (\cancel{x}, Y_1), (X_2, Y_2), (X_3, Y_3) \; .... \; (X_n, Y_n) \}$

$$\left[\text{ePMF}_{ij} : \begin{cases} 1/n & \text{if } i=j \rightarrow \hat{P}(X = x_i \cap Y = y_j) \\ 0 & \text{otherwise} \end{cases} \right]$$

$$\rho = \frac{\sum_{x \in \Omega} \sum_{y \in \Omega} P(\cancel{x}\, x \cap Y = y) - \left(\sum_{x \in \Omega} P(X = x)\right)\left(\sum_{y \in \Omega} y\, P(Y = y)\right)}{\sqrt{\left(\sum_{x \in \Omega} x^2 P_x(x) - \left(\sum_{x \in \Omega} x P_x(x)\right)^2\right)\left(\sum_{y \in \Omega} y^2 P_y(y) - \left(\sum_{y \in \Omega} y P_y(y)\right)^2\right)}}$$

$$\hat{\rho} = \frac{\sum_{x \in D} \sum_{y \in D} xy\, \hat{P}(\cancel{X} = x \cap Y = y) - \left(\sum_{x \in D} x\, \hat{P}_X(x)\right)\left(\sum_{y \in D} y\, \hat{P}_Y(y)\right)}{\sqrt{\left(\sum_{x \in \Omega} x^2 \hat{P}_x(x) - \left(\sum_{x \in \Omega} x P_x(x)\right)^2\right)\left(\sum_{y \in \Omega} y^2 P_y(y) - \left(\sum_{y \in \Omega} y P_y(y)\right)^2\right)}}$$

$$= \frac{\frac{1}{n} \sum_{i=1}^{n} X_i Y_i - \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} Y_i\right)}{\sqrt{\left(\frac{1}{n} \sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum X_i\right)^2\right)\left(\frac{1}{n} \sum_{i=1}^{n} Y_i^2 - \frac{1}{n^2}\left(\sum Y_i\right)^2\right)}}$$

$$\hat{R} = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{\sqrt{\left(n\sum_{i=1}^{n} x_i^2 - (\sum x_i)^2\right)\left(n\sum y_i^2 - (\sum y_i)^2\right)}}$$

(6) Properties of estimators

~~(a)~~ Given $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i$ where X_i are i.i.d $\sim$ Poisson(θ)

i) bias $= E(\hat{\theta}) - \theta$

$$E(\hat{\theta}) = E\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n} E\left(\sum_{i=1}^{n} X_i\right) \underset{LOE}{=} \frac{1}{n}\sum_{i=1}^{n} E[X_i]$$

$$\underset{\text{i.i.d}}{=} \frac{\cancel{n}}{\cancel{n}} E[X] = E[X]$$

$$E[\hat{\theta}] = E[X]$$

$$\text{bias} = E[\hat{\theta}] - \theta = E[X] - \theta$$

$$X \sim \text{Poisson}(\theta) = \frac{\theta^{x} e^{-\theta}}{x!}$$

W.K.T $\boxed{E[X] = \theta}$

$$\text{bias} = E[X] - \theta = \theta - \theta$$
$$= 0$$

ii) $se = \sqrt{Var(\hat{\theta})}$

$$Var(\hat{\theta}) = E[\hat{\theta}^2] - (E[\hat{\theta}])^2$$

$$Var(\hat{\theta}) = Var\left(\frac{1}{n}\sum_i X_i\right)$$

$$= \frac{1}{n^2} Var\left(\sum X_i\right) \overset{LOV}{=} \frac{1}{n^2} \sum Var(X_i)$$

($\because X_i$'s are iid)

$$= \frac{1}{n^{\cancel{2}}} \cdot \cancel{n} \, Var(X)$$

$$= \frac{1}{n} Var(X)$$

$$X \sim Poisson(\theta) = \frac{\theta^x e^{-\theta}}{x!}$$

[For Poisson distribution
Variance $= \theta$]

$$\Rightarrow \frac{1}{n} Var(X) = \frac{\theta}{n}$$

$$se = \sqrt{Var(\hat{\theta})} = \sqrt{\frac{\theta}{n}}$$

iii) $MSE = bias^2(\hat{\theta}) + Var(\hat{\theta})$

From i) & ii)

$$= 0^2 + \frac{\theta}{n}$$

$$\boxed{MSE = \frac{\theta}{n}}$$

7) Given $\hat{\theta} = \hat{F}_n(b) - \hat{F}_n(a)$

$\theta = F(b) - F(a)$

(a) $\hat{se}(\hat{\theta})$

$se\ \hat{\theta} = \hat{F}_n(b) - \hat{F}_n(a) \Rightarrow \frac{1}{n}\sum_{i=1}^{n} I(X_i \le b) - \frac{1}{n}\sum_{i=1}^{n} I(X_i \le a)$

$se(\hat{\theta}) = \sqrt{Var(\hat{\theta})}$

$Var(\hat{\theta}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} I(X_i \le b) - \frac{1}{n}\sum_{i=1}^{n} I(X_i \le a)\right)$

$= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} I(X_i \le b) - \sum_{i=1}^{n} I(X_i \le a)\right)$

$= \frac{n}{n^2} Var(I(X \le b) - I(X \le a))$

The distribution is in the form of binomial distribution

[W.K.T Variance of indicator RV ~~binomial~~ ~~distribution~~ is $P(1-P)$]

$= \frac{1}{n} Var(I(X \le b) - I(X \le a))$

$= \frac{1}{n}\left(Pr(X \le b) - Pr(X \le a)\right)\left(1 - (Pr(X \le b) - Pr(X \le a))\right)$

$= \frac{1}{n}(F(b) - F(a))\left(1 - (F(b) - F(a))\right)$

$= \frac{1}{n}$

$= \frac{\theta(1-\theta)}{n}$

$$SE[\hat{\theta}] = \sqrt{Var[\hat{\theta}]}$$

$$= \sqrt{\frac{\theta(1-\theta)}{n}}$$

(b) Find an approximate $(1-\alpha)$ CI for θ

$\theta = F(b) - F(a)$ true value

$\hat{\theta} = F(\hat{b}) - F(\hat{a}) \rightarrow$ eCDF

W.K.T If $\hat{\theta} \simeq \text{Normal}(\theta, \hat{se}^2)$ then a $(1-\alpha)$ CI for θ is

$$\left(\hat{\theta} - Z_{\alpha/2}\, se,\ \hat{\theta} + Z_{\alpha/2}\, se\right)$$

We need to show that $\hat{\theta}$ is a normal distribution

$$\hat{\theta} = \hat{F}_n(b) - \hat{F}_n(a)$$

$$= \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq b) - \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq a) \rightarrow ①$$

[$I[X \leq b]$ is the indicator of event and a Bernouli random Variable with mean P & Variance as $P(1-P)$]

From ①

$$\hat{\theta} = \frac{\cancel{n}}{\cancel{n}}{}^{1} I(X \leq b) - \frac{\cancel{n}}{\cancel{n}}{}^{1} I(X \leq a)$$

$$\hat{\theta} = I(X \leq b) - I(X \leq a)$$

7)b

Using CLT, it can be approximated to a normal distribution with

$$\mu = E[I(X \le b) - I(X \le a)]$$
$$= E[I(X \le b)] - E[I(X \le a)]$$
$$= Pr(X \le b) - Pr(X \le a)$$
$$= F(b) - F(a) = \theta$$

se

From (a) Variance $= \dfrac{(F(b)-F(a))(1-(F(b)-F(a)))}{n}$

$$= \frac{\theta(1-\theta)}{n}$$

$$se = \sqrt{\frac{\theta(1-\theta)}{n}}$$

$\therefore$ An approximate $1-\alpha$ CI for θ is $\hat{\theta} \pm Z_{\alpha/2}\, se$

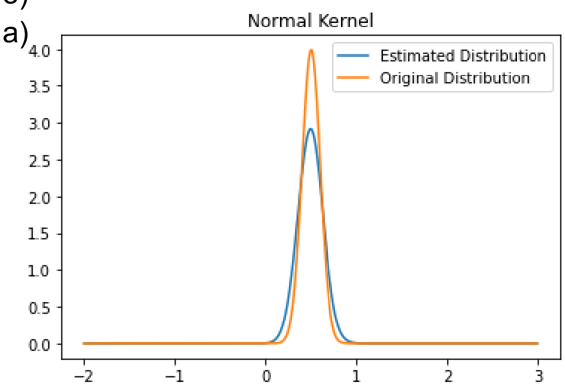
$$= \hat{\theta} \pm Z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}}$$

Since θ is unknown in the CI, we will get estimated CI by replacing θ with $\hat{\theta}$

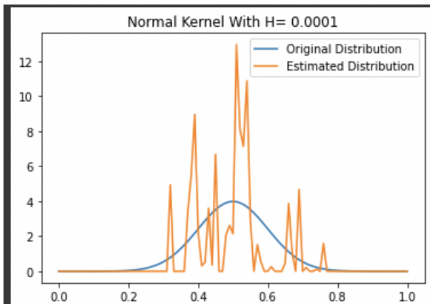
est $\hat{C.I} = \hat{\theta} \pm Z_{\alpha/2}\sqrt{\dfrac{\hat{\theta}(1-\hat{\theta})}{n}}$

8)

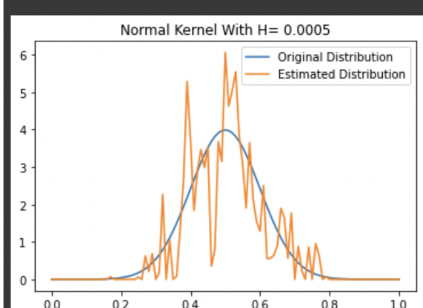
a)



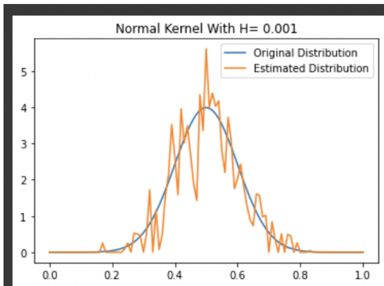
b)



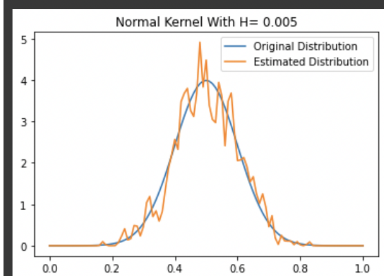

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.49389189546542334 Standard Deviation = 0.1021160365478767
The mean of original distribution is 1.22 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 2.12 percent lesser than the standard deviation of estimated distribution
```



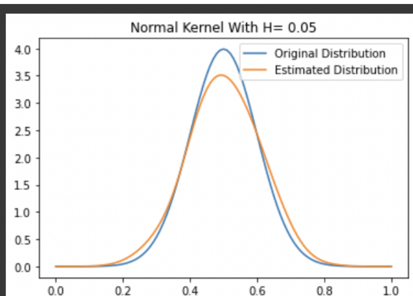

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4757198474598633 Standard Deviation = 0.15308392795549175
The mean of original distribution is 4.86 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 53.08 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4783513188411217 Standard Deviation = 0.15075778934495448
The mean of original distribution is 4.33 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 50.76 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.5019868670626897 Standard Deviation = 0.10185701069419788
The mean of original distribution is 0.4 percent lesser than the mean of estimated distribution
The standard deviation of original distribution is 1.86 percent lesser than the standard deviation of estimated distribution
```



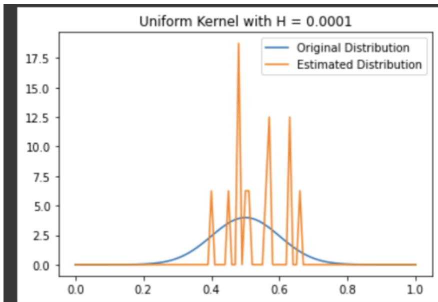

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.5021041376030938 Standard Deviation = 0.11297870948099324
The mean of original distribution is 0.42 percent lesser than the mean of estimated distribution
The standard deviation of original distribution is 12.98 percent lesser than the standard deviation of estimated distribution
```

H	Percentage Deviation in Mean	Percentage Deviation in Standard Deviation
0.0001	27.26	156.32
0.0005	6.03	57.93
0.001	3.31	28.12
0.005	0.73	3.66
0.05	0.69	8.78

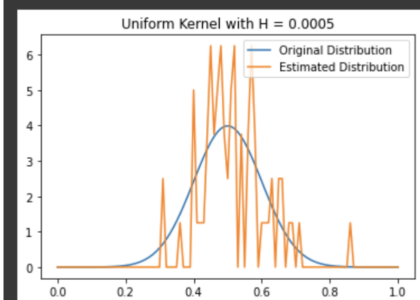
Based on the observation of above plots the H value of 0.05 looks like performing best.

And the deviations from original mean and variance are least for the H value of 0.005

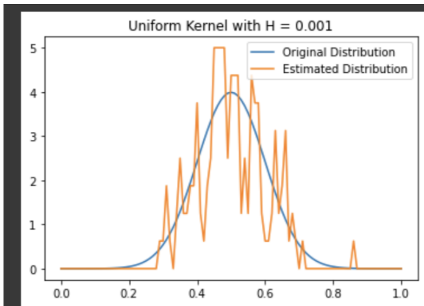
c)



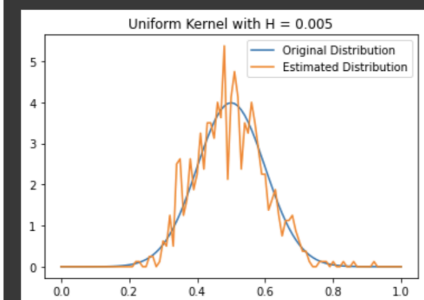

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4325 Standard Deviation = 0.21847482692520895
The mean of original distribution is 13.5 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 118.47 percent lesser than the standard deviation of estimated distribution
```



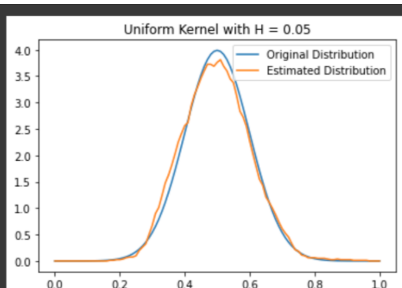

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4590000000000001 Standard Deviation = 0.1861155555024884
The mean of original distribution is 8.2 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 86.12 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.46668750000000003 Standard Deviation = 0.15901541542803327
The mean of original distribution is 6.66 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 59.02 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.49718749999999984 Standard Deviation = 0.10051474938410829
The mean of original distribution is 0.56 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 0.51 percent lesser than the standard deviation of estimated distribution
```

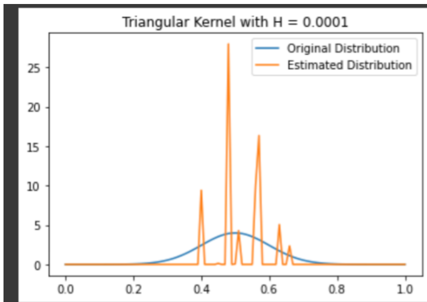



```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.49711250000000035 Standard Deviation = 0.10441581462474743
The mean of original distribution is 0.58 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 4.42 percent lesser than the standard deviation of estimated distribution
```

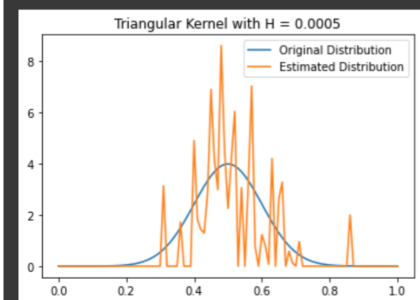
H	Percentage Deviation in Mean	Percentage Deviation in Standard Deviation
0.0001	13.5	118.47
0.0005	8.2	86.12
0.001	6.66	59.02
0.005	0.56	0.51
0.05	0.58	4.42

- Based on the observation of above plots the H value of 0.05 looks like performing the best.

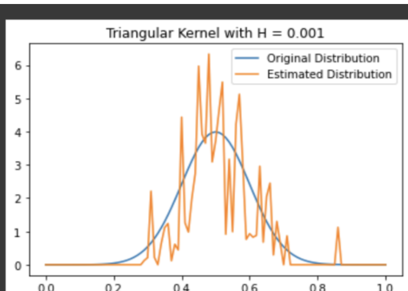
  And the deviations from original mean and variance are least for the H value of 0.005



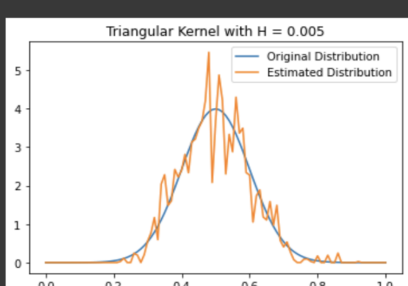

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.39118009509918783 Standard Deviation = 0.22983014624892953
The mean of original distribution is 21.76 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 129.83 percent lesser than the standard deviation of estimated distribution
```



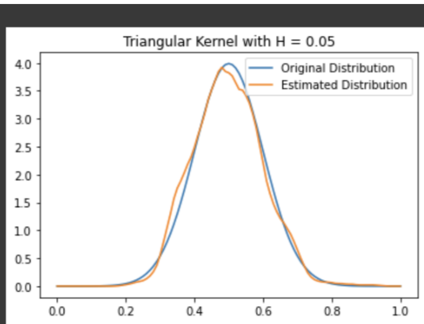

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4451453654582157 Standard Deviation = 0.19679724560191428
The mean of original distribution is 10.97 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 96.8 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.45133959579516647 Standard Deviation = 0.18747773437990636
The mean of original distribution is 9.73 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 87.48 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.49585523888407823 Standard Deviation = 0.10675493091137168
The mean of original distribution is 0.83 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 6.75 percent lesser than the standard deviation of estimated distribution
```



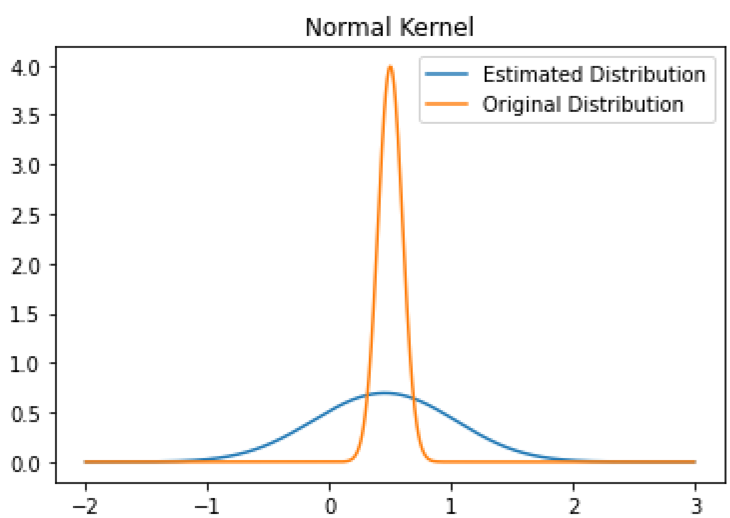

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4972073793000693 Standard Deviation = 0.10246496679673502
The mean of original distribution is 0.56 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 2.47 percent lesser than the standard deviation of estimated distribution
```

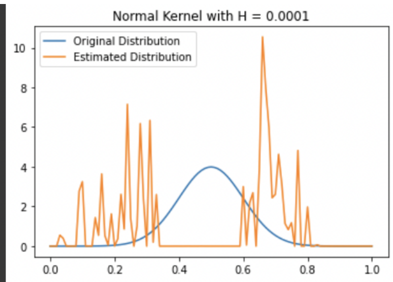
H	Percentage Deviation in Mean	Percentage Deviation in Standard Deviation
0.0001	21.76	129.83
0.0005	10.97	96.8
0.001	9.73	87.48
0.005	0.83	6.75
0.05	0.56	2.47

Based on the observation of above plots the H value of 0.05 looks like performing the best.

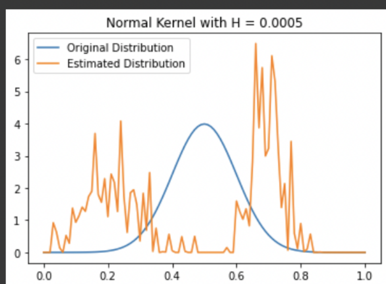
And the deviations from original mean and variance are least for the H value of 0.05

d)

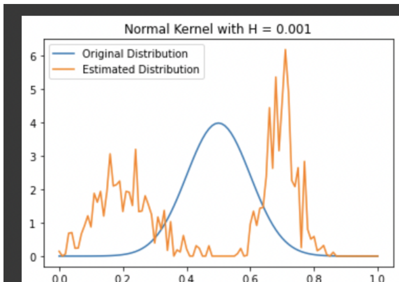




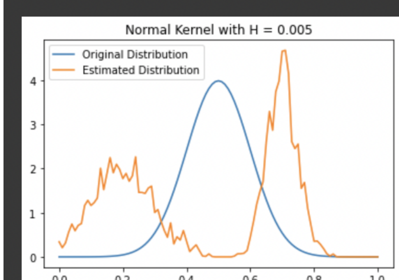

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.5080970912058647 Standard Deviation = 0.21536076144782934
The mean of original distribution is 1.62 percent lesser than the mean of estimated distribution
The standard deviation of original distribution is 115.36 percent lesser than the standard deviation of estimated distribution
```



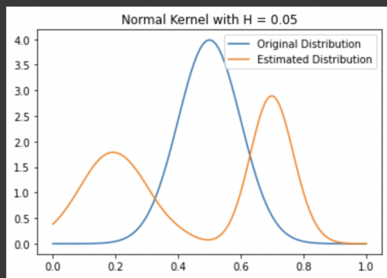

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.48697731318403975 Standard Deviation = 0.23839712135235744
The mean of original distribution is 2.6 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 138.4 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4679898661169175 Standard Deviation = 0.2535151888562762
The mean of original distribution is 6.4 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 153.52 percent lesser than the standard deviation of estimated distribution
```



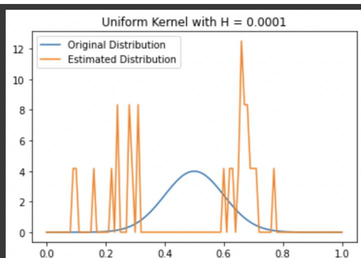

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4482347692271753 Standard Deviation = 0.2638636347882409
The mean of original distribution is 10.35 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 163.86 percent lesser than the standard deviation of estimated distribution
```



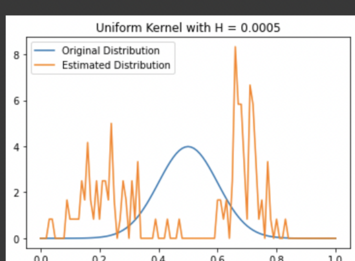

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4485429342716778 Standard Deviation = 0.26778918892667736
The mean of original distribution is 10.29 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 167.79 percent lesser than the standard deviation of estimated distribution
```

H	Percentage Deviation in Mean	Percentage Deviation in Standard Deviation
0.0001	1.62	115.36
0.0005	2.6	138.4
0.001	6.4	153.52
0.005	10.35	163.86
0.05	10.29	167.79

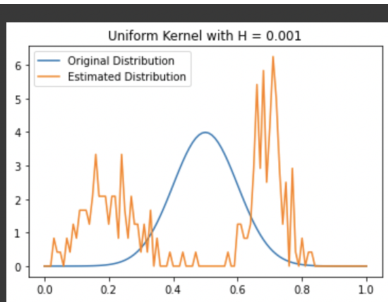
Based on the observation of above plots and the values in the above table the H value of 0.0001 looks like performing the best.



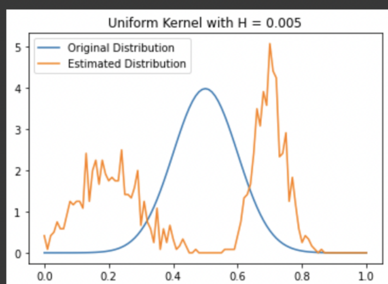

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.52375 Standard Deviation = 0.1840700161170562
The mean of original distribution is 4.75 percent lesser than the mean of estimated distribution
The standard deviation of original distribution is 84.07 percent lesser than the standard deviation of estimated distribution
```



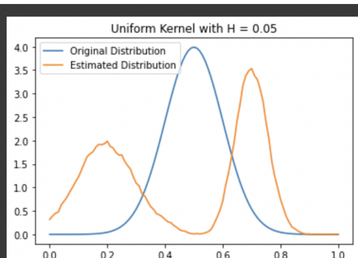

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.49908333333333355 Standard Deviation = 0.22695078700507318
The mean of original distribution is 0.18 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 126.95 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.45683333333333334 Standard Deviation = 0.26106507277347957
The mean of original distribution is 8.63 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 161.07 percent lesser than the standard deviation of estimated distribution
```



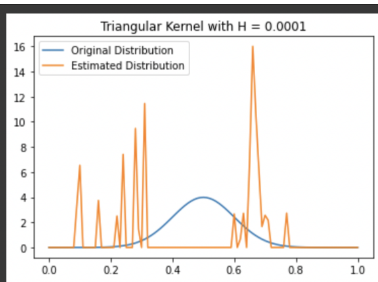

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.44812500000000005 Standard Deviation = 0.2638090869833713
The mean of original distribution is 10.37 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 163.81 percent lesser than the standard deviation of estimated distribution
```



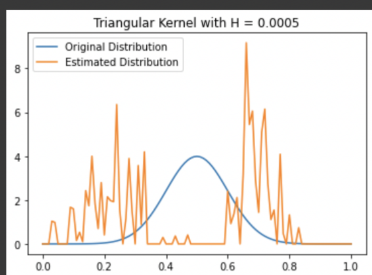

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.44824500000000006 Standard Deviation = 0.26526691584452555
The mean of original distribution is 10.35 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 165.27 percent lesser than the standard deviation of estimated distribution
```

H	Percentage Deviation in Mean	Percentage Deviation in Standard Deviation
0.0001	4.75	84.07
0.0005	0.18	126.95
0.001	8.63	161.07
0.005	10.37	163.81
0.05	10.35	165.27

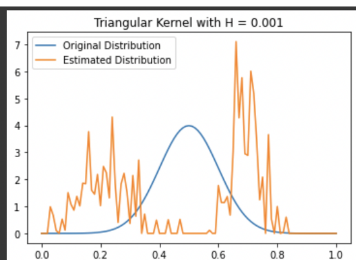
Based on the observation of above plots and the values in the above table the H value of 0.001 looks like performing the best.



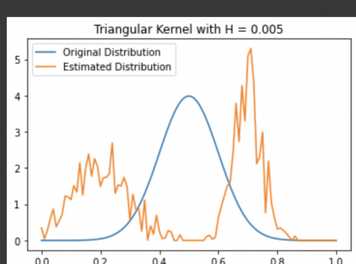

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4723761597035165 Standard Deviation = 0.22337543709654156
The mean of original distribution is 5.52 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 123.38 percent lesser than the standard deviation of estimated distribution
```



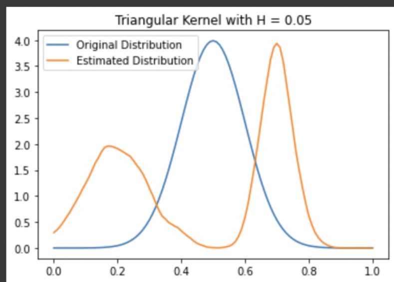

```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.5136074904983706 Standard Deviation = 0.20989575977926755
The mean of original distribution is 2.72 percent lesser than the mean of estimated distribution
The standard deviation of original distribution is 109.9 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4912300043334525 Standard Deviation = 0.23440230441775936
The mean of original distribution is 1.75 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 134.4 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4540199733545321 Standard Deviation = 0.2611668736973374
The mean of original distribution is 9.2 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 161.17 percent lesser than the standard deviation of estimated distribution
```




```
For Original Distribution Mean = 0.4999997815123576 Standard Deviation = 0.09999994582855165
For Estimated Distribution Mean = 0.4481913939074269 Standard Deviation = 0.26454555398348106
The mean of original distribution is 10.36 percent greater than the mean of estimated distribution
The standard deviation of original distribution is 164.55 percent lesser than the standard deviation of estimated distribution
```

H	Percentage Deviation in Mean	Percentage Deviation in Standard Deviation
0.0001	5.52	123.38
0.0005	2.72	109.9
0.001	1.75	134.4
0.005	9.2	161.17
0.05	10.36	164.55

Based on the observation of above plots and the values in the above table the H value of 0.001 looks like performing the best.

[ ] ↳ *1 cell hidden*