ASSIGNMENT

Piyush Soni [SBU ID # 113273375]

Abhishek Bhattad [SBUID# 113277131]

Aditya Nandan Bhide [SBU ID# 113216565]

Niket Shah [SBU ID# 113274949]

ASSIGNMENT 3

1. To prove $MSE = bias^2(\hat{\theta}) + Var(\hat{\theta})$.

A. We know that:

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

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta$$

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

LHS:

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

$$= E[\theta^2 + \hat{\theta}^2 - 2\theta.\hat{\theta}]$$

$$\overset{L.O.E}{=} E[\theta^2] + E[\hat{\theta}^2] - E[2\theta.\hat{\theta}]$$

$$= \theta^2 + E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] \longrightarrow (1).$$

$\because$ Since θ = const. is true value $E[\theta^2] = \theta^2$

RHS:

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

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

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

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

Since (1) = (2) i.e L.H.S = RHS

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

Hence Proved

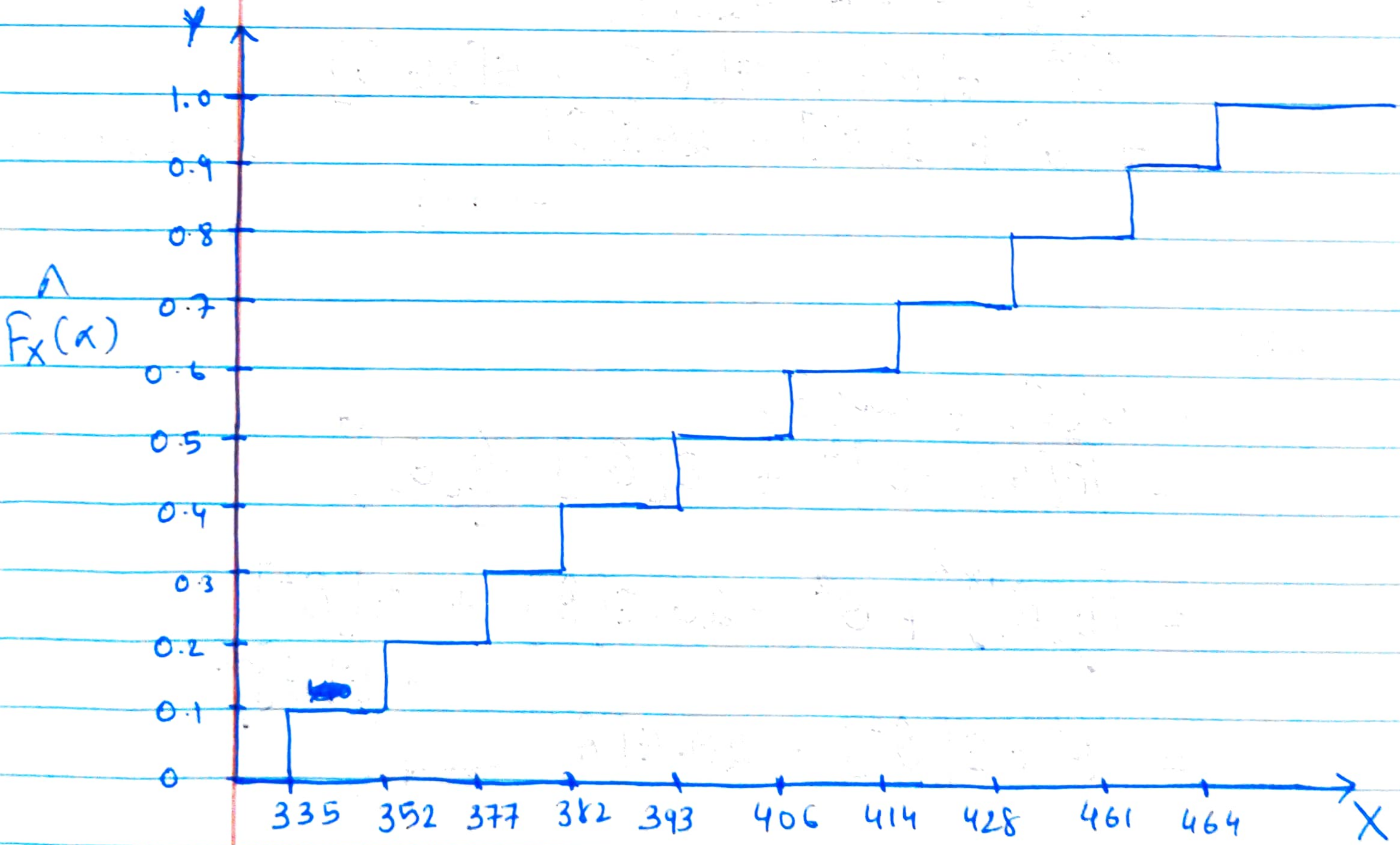
(2) 10 samples from collision.csv.

$D = \{393, 377, 414, 382, 335, 461, 428, 406, 464, 352\}$

$D_{sort} = \{335, 352, 377, 382, 393, 406, 414, 428, 461, 464\}$

We know that $\hat{F}_X(\alpha) = \sum_{i=1}^{n} \frac{I(X \leq \alpha)}{n}$

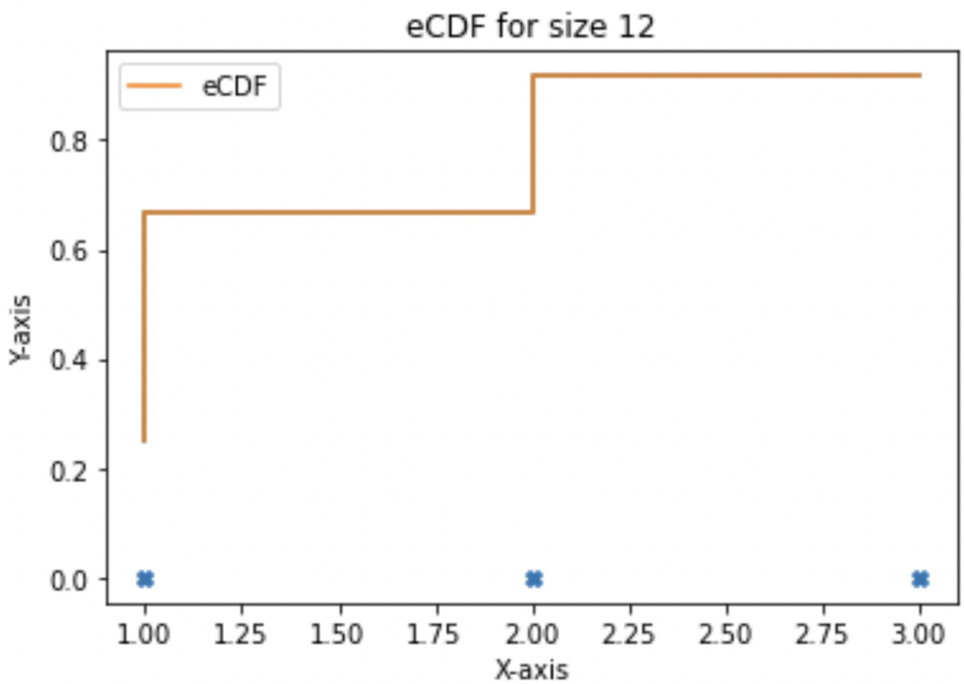
Since each element is unique, $p_X(\alpha)$ for $\alpha \in D$ is $\frac{1}{10}$



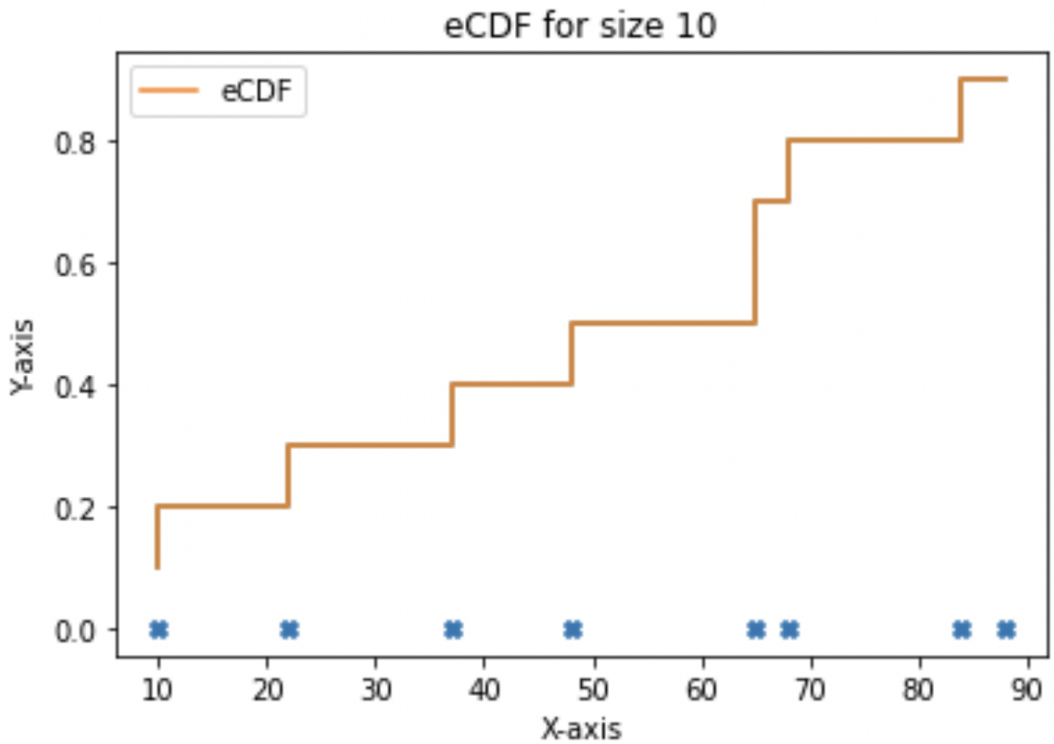
Assignment 3

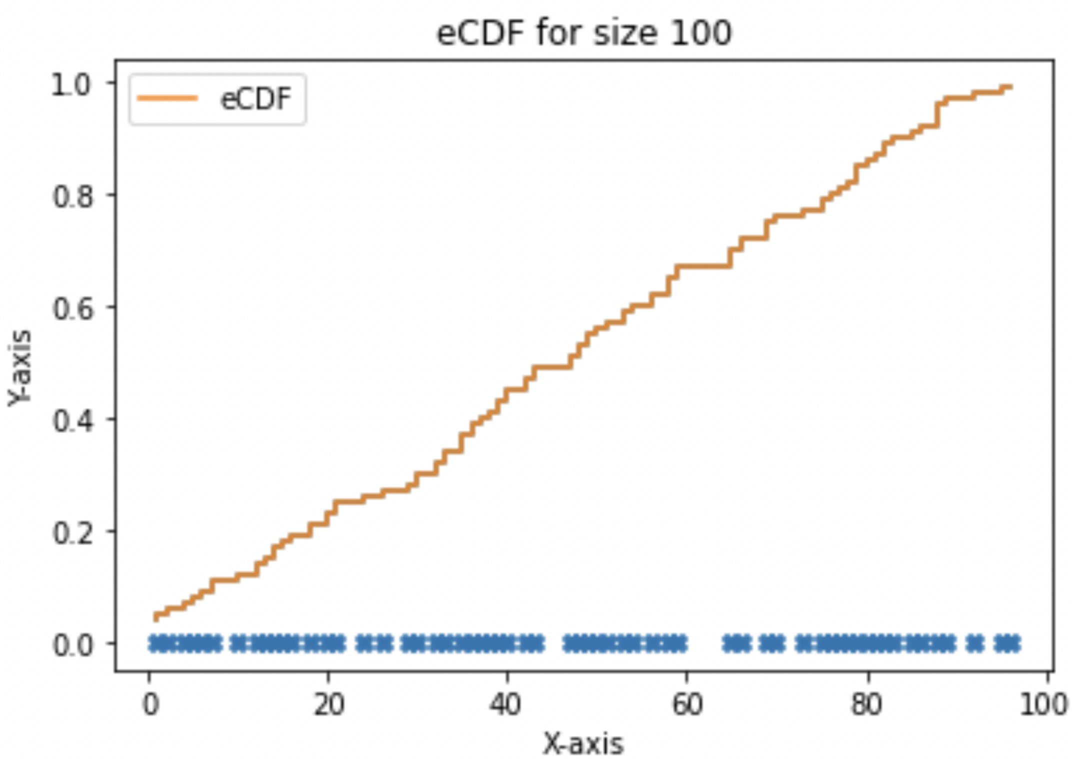
Q3 a]

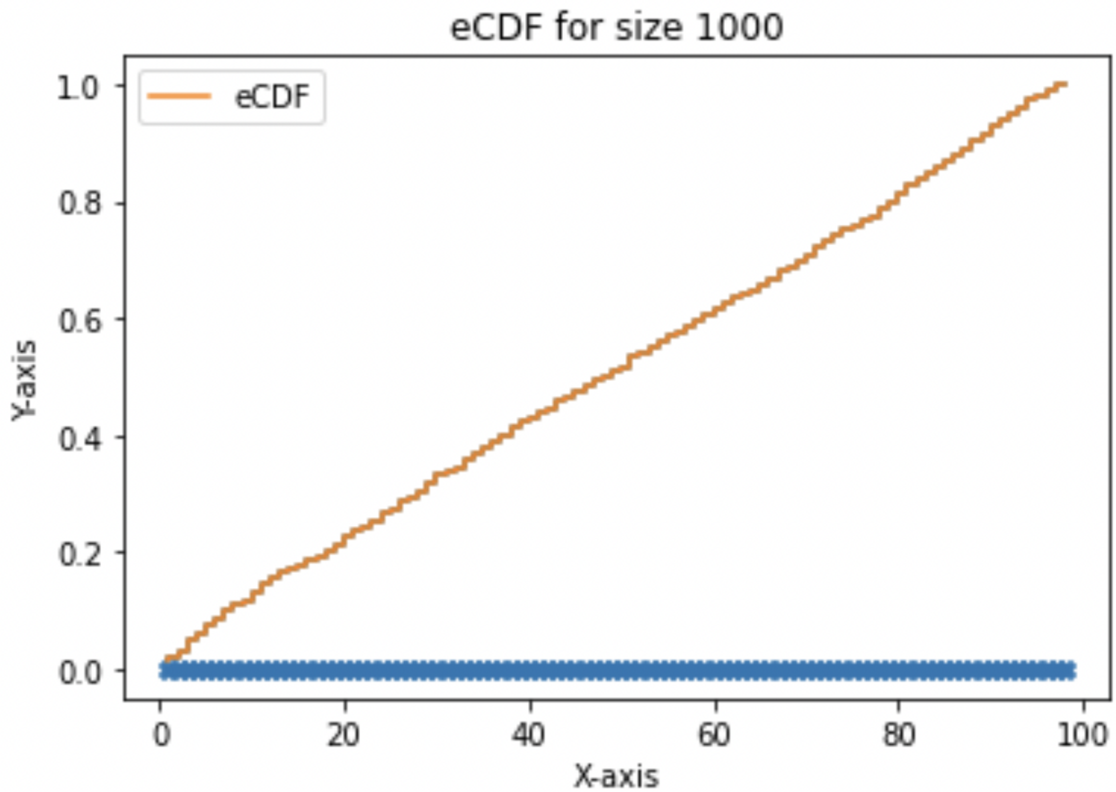
```
Give input numbers. (Input q to stop)
 2
 1
 2
 3
 2
 1
 2
 3
 2
 1
 2
 3
 q
Answer for question 3 a:
```



Q3 b] As value of n increases estimated CDF becomes smoother and estimated CDF approaches the true CDF.

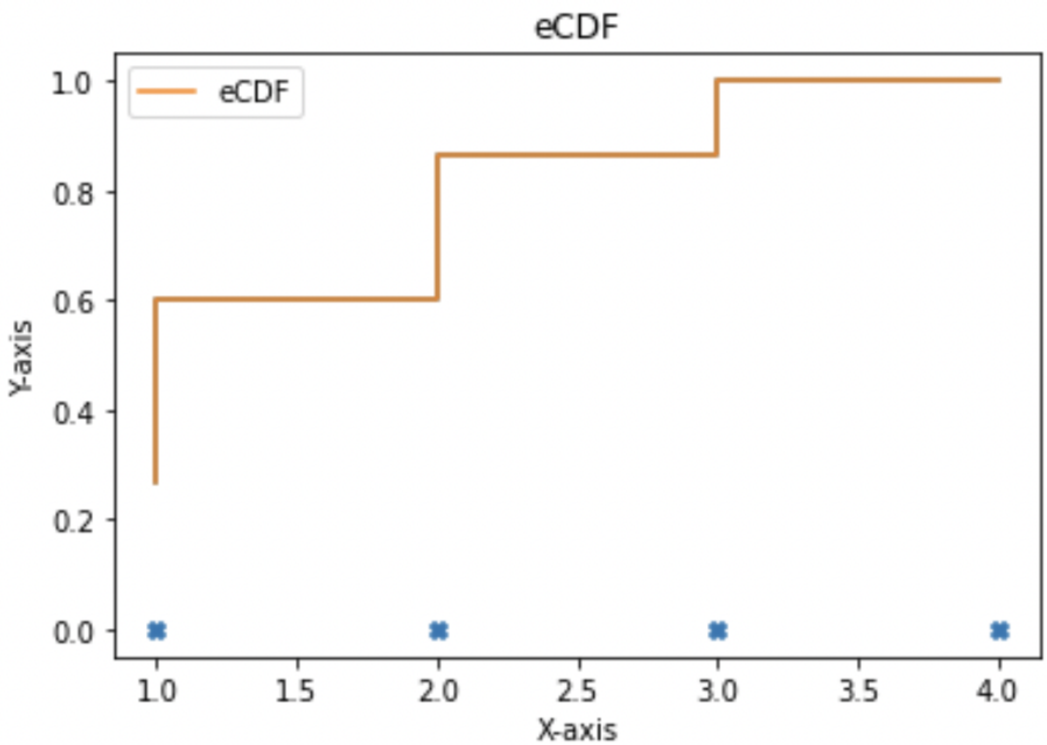




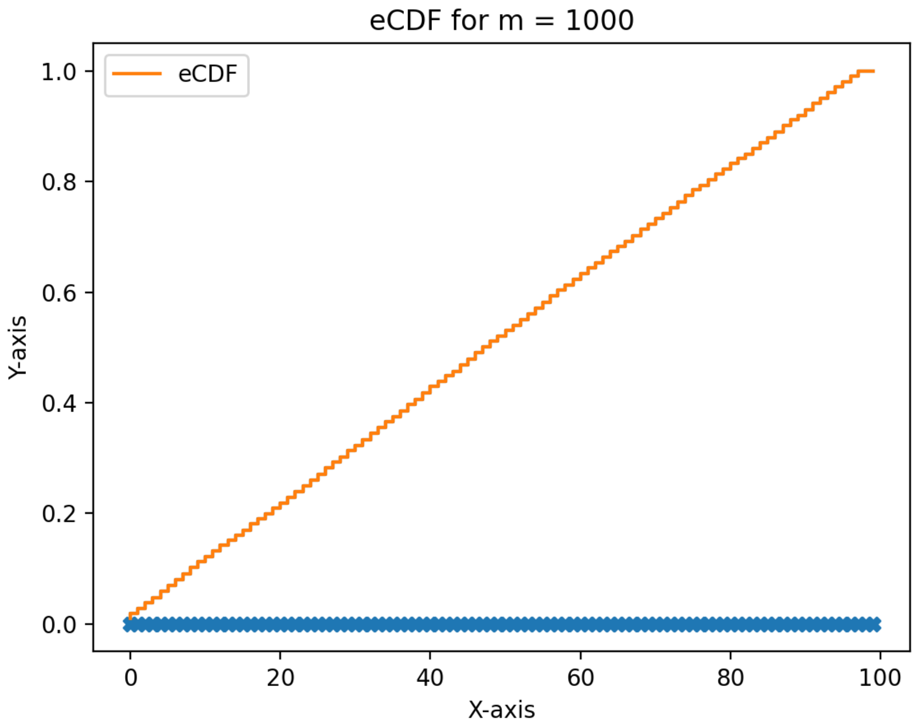

eCDF for size 1000
eCDF
1.0
0.8
0.6
0.4
0.2
0.0
Y-axis
0
20
40
60
80
100
X-axis

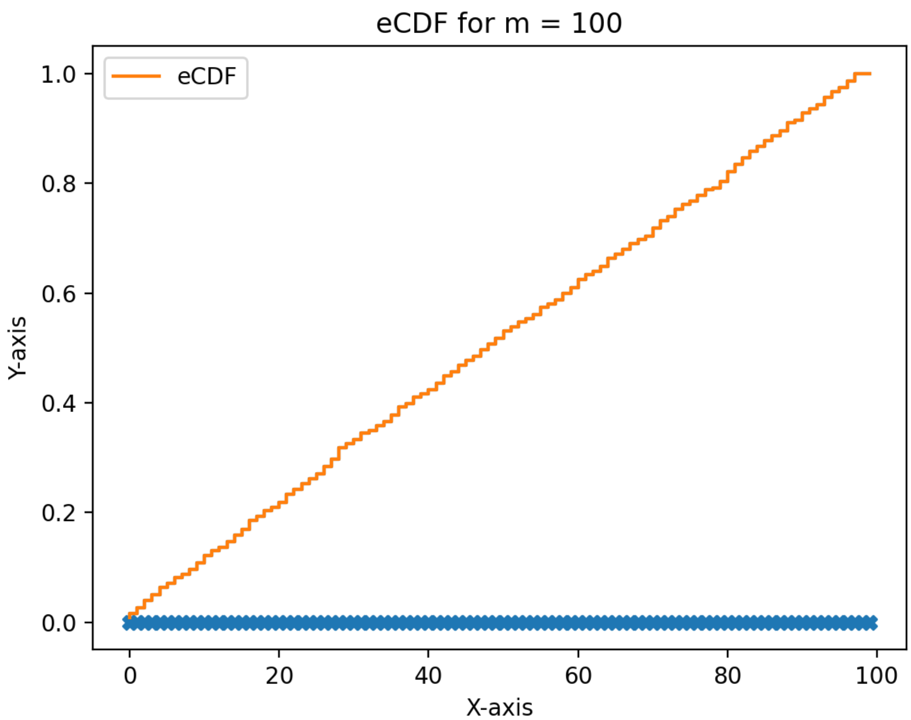
Q3 c]

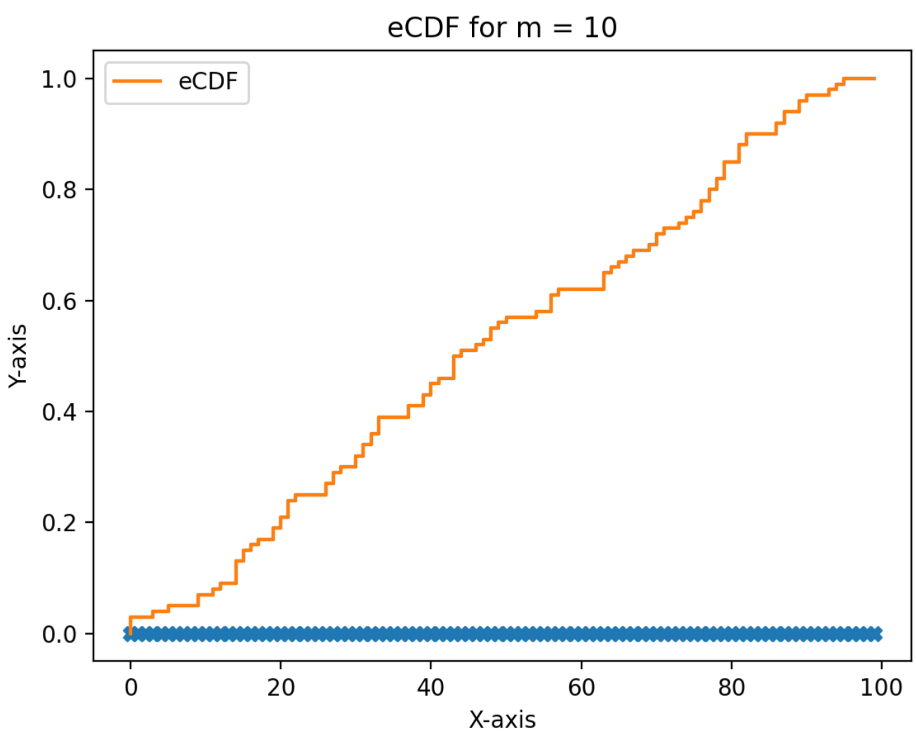
```
Answer for question 3 c:
Size of sample:  5
Total number of samples:  3
 4
 3
 2
 1
 3
 2
 1
 2
 1
 2
 3
 2
 1
 4
 3
[[4. 3. 2. 1. 3.]
 [2. 1. 2. 1. 2.]
 [3. 2. 1. 4. 3.]]
```



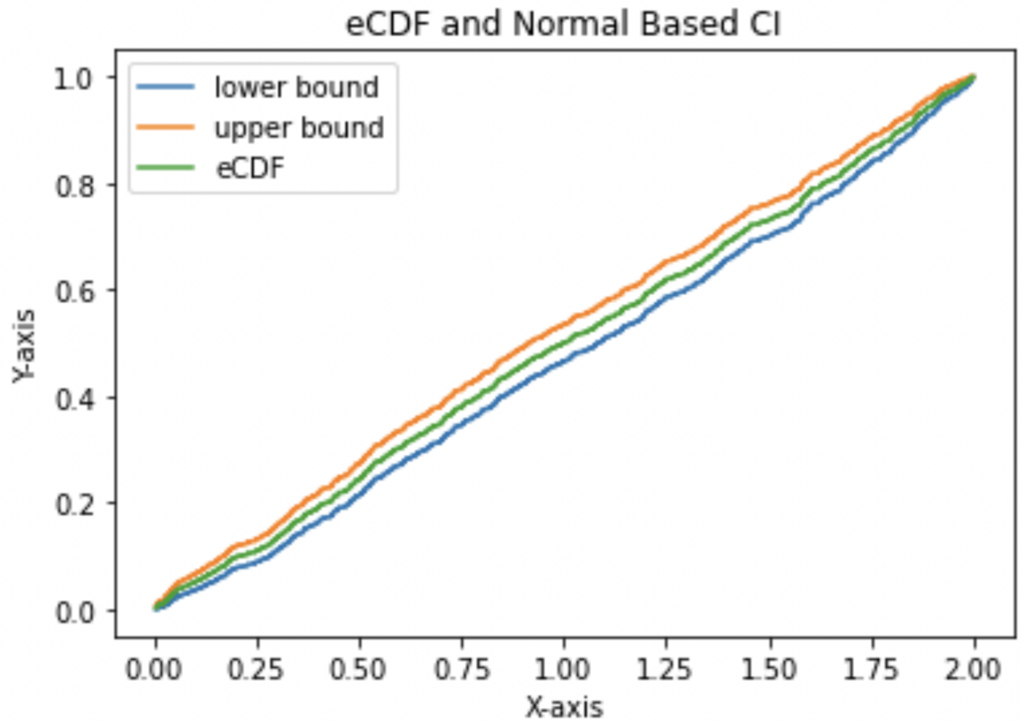
Q3 d] As value m increases estimated CDF becomes smoother and estimated CDF approaches the true CDF.




eCDF for m = 100
eCDF
1.0
0.8
0.6
0.4
0.2
0.0
Y-axis
0
20
40
60
80
100
X-axis

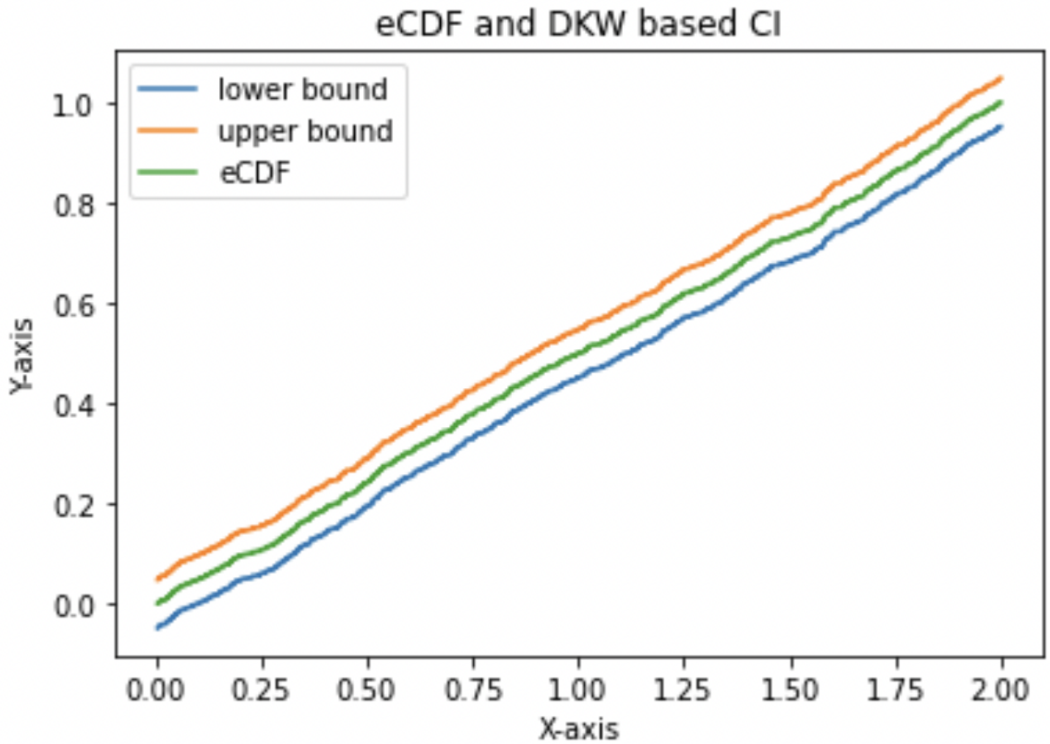

eCDF for m = 10
eCDF
1.0
0.8
0.6
0.4
0.2
0.0
Y-axis
0
20
40
60
80
100
X-axis

Q3 e]



eCDF and Normal Based CI
lower bound
upper bound
eCDF
Y-axis
X-axis
0.0
0.2
0.4
0.6
0.8
1.0
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

Q3 f]


eCDF and DKW based CI
lower bound
upper bound
eCDF
1.0
0.8
0.6
0.4
0.2
0.0
Y-axis
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
X-axis

4. (a) Given, $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ (Sample mean)

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

$\Rightarrow$ L.H.S. :

$$\hat{\sigma}^2 = \hat{E}[X^2] - (\hat{E}[X])^2 \quad \text{— (1)}$$ {by definition of variance}

$$\Rightarrow E[X] = \sum_{x \in \Omega} x\, p(x)$$

$$= \sum_{i=1}^{n} x_i \hat{P}_X(x_i)$$ {Plug in estimator}

$$= \sum_{i=1}^{n} x_i (1/n)$$ {Best guess for $Pr(X = x_i) = 1/n$ $\because$ x_i are iid}

$$= \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$$ {Given} — (2)

$$\Rightarrow E[X^2] = \sum_{x \in \Omega} x^2 p(x)$$

$$= \sum_{i=1}^{n} x_i^2 \hat{P}_X(x_i)$$

$$= \sum_{i=1}^{n} x_i^2 (1/n)$$

$$= \frac{1}{n}\sum_{i=1}^{n} x_i^2 \quad \text{— (3)}$$

Substituting ② and ③ in ① :-

LHS

$\Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum x_i^2 - \frac{1}{n}(\bar{x}_n)^2$ — ④

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

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

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

{ $\because \bar{x}_n$ is const for $\sum_i$ }

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

$\bar{x}_n$ (Given)

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

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

$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}x_i^2 - \bar{x}_n^2$ = L.H.S.

Hence proved

4. (b) from part (a) — , let μ, σ^2 are true mean & variance

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

from def^n —

$$\text{bias}(\hat{\sigma}^2) = E(\hat{\sigma}^2) - \sigma^2 \quad — (1)$$

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

Adding & subtracting true mean (μ) —

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

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

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

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

Since, $\bar{X}_n - \mu$ would be const for $\sum_i$ over i,

Thus,

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

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

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

By definition of variance,

$$\sigma^2 = \frac{1}{n}\sum(X_i-\mu)^2 = E[(X-\mu)^2] \Rightarrow \text{true variance which is a const}$$

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

~~$\because$ $(\bar{X}_n-\mu)$ is const~~

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

$$\because \quad \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$= \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu + \mu)$$

$$= \frac{1}{n}\Big(\sum_{i=1}^{n} \underset{\text{const}}{\mu}\Big) + \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)$$

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

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu) = \bar{X}_n - \mu \qquad \text{— (3)}$$

Substituting eqⁿ (3) in eqⁿ (2) —

$$\Rightarrow E[\hat{\sigma}^2] = E[\underset{\text{const}}{\sigma^2}] + E\big[(\bar{X}_n - \mu)^2 - 2(\bar{X}_n - \mu)(\bar{X}_n - \mu)\big]$$

$$= \sigma^2 + E\big[(\bar{X}_n - \mu)^2 - 2(\bar{X}_n - \mu)^2\big]$$

$$\Rightarrow \boxed{E[\hat{\sigma}^2] = \sigma^2 - E[(\bar{X}_n - \mu)^2]} \qquad \text{— (4)}$$

Now, $$E[(\bar{X}_n - \mu)^2] = E[\bar{X}_n^2 + \mu^2 - 2\mu\bar{X}_n]$$

$$\overset{\text{LOE}}{=} E[\bar{X}_n^2] + E[\underset{\text{const}}{\mu^2}] - E[\underset{\text{const}}{2\mu}\bar{X}_n]$$

$$= E[\bar{X}_n^2] + \mu^2 - 2\mu E[\bar{X}_n] \qquad \text{— (5)}$$

from def^n of variance -

$$var(\bar{X}_n) = E[\bar{X}_n^2] - (E[\bar{X}_n])^2$$

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

Substituting $E[\bar{X}_n^2]$ in eq^n 5 -

$$\Rightarrow E[(\bar{X}_n - \mu)^2] = var(\bar{X}_n) + \underbrace{(E[\bar{X}_n])^2 + \mu^2 - 2\mu E[\bar{X}_n]}_{(a+b)^2}$$

$$= var(\bar{X}_n) + (E(\bar{X}_n) - \mu)^2$$

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

$$\overset{LOE}{=} var(\bar{X}_n) + \left[\frac{1}{n}\sum_{i=1}^{n} \underset{\substack{\text{true mean} \\ (\mu)}}{E[X_i]} - \mu\right]^2$$

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

$$= var(\bar{X}_n) + \left[\frac{\cancel{n}\mu}{\cancel{n}} - \mu\right]^2$$

$$E[(\bar{X}_n - \mu)^2] = var(\bar{X}_n) \quad -6$$

Substituting 6 to 4 -

$$E[\hat{\sigma}^2] = \sigma^2 - var(\bar{X}_n)$$

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

$$\overset{cov}{\underset{iid}{=}} \sigma^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} var(X_i)\right)$$

$$\overset{iid}{=} \sigma^2 - \frac{1}{n^2} n \times var(X_1)$$ {iid var will be same}

$$= \sigma^2 - \frac{\sigma^2}{n} \quad \text{— (7)}$$

from eq^n 1 -

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

$$= \cancel{\sigma^2} - \sigma^2/n - \cancel{\sigma^2}$$

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

4. (c) RV, $X \sim \mu, \sigma^2$

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

$$\Rightarrow E[(X-\mu)^4] = \sum_{i=1}^{n} (x_i-\mu)^4 \, p(x)$$

for plug-in estimator - $\hat{p}(x) = \frac{1}{n}$.

$$\Rightarrow \hat{E}[(x-\mu)^4] = \sum_{i=1}^{n} (x_i-\mu)^4 \, \hat{p}(x)$$

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

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

$$\Rightarrow \hat{\sigma}^4 = \left[\frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2\right]^2$$

Plug-in estimator of Kurt[X] -

$$\widehat{\text{Kurt}}[X] = \frac{\hat{E}[(x-\mu)^4]}{\hat{\sigma}^4} = \frac{1/n \left(\sum_{i=1}^{n} (x_i-\mu)^4\right)}{\left[1/n \sum_{i=1}^{n} (x_i-\bar{x}_n)^2\right]^2}$$

$$\boxed{\widehat{\text{Kurt}}(X) = \frac{n\left(\sum_{i=1}^{n} (x_i-\mu)^4\right)}{\left[\sum_{i=1}^{n} (x_i-\bar{x}_n)^2\right]^2}}$$

4. (d) $\rho = \dfrac{E[XY] - E[X]E[Y]}{\sigma_x \sigma_y}$ — (1)

$X \sim iid$
$Y \sim iid$
they apear in pair –
$\{(x_1, y_1), (x_2, y_2) \ldots (x_n, y_n)\}$

$\Rightarrow$ ePMF $(X = x_1, Y = y_1) = \dfrac{\sum I(x = x_1, Y = y_1)}{n}$

$= \dfrac{1}{n}$

{as, (x_1, y_1) occurs only in pair}

$\Rightarrow$ ePMF $(X = x_1, Y = y_2) = 0$

$\because (x_i, y_i)$ only occurs in pair

$\Rightarrow$ ~~ePMF = $\hat{P}_{XY}(x,y) = \sum\sum \frac{1}{n}$~~

$\Rightarrow$ ePMF, $\hat{P}_{XY}(x = x_i, y = y_i) = \dfrac{1}{n}$

$\Rightarrow \hat{E}[XY] = \sum_{j=1}^{n} \sum_{i=1}^{n} x_i y_j \hat{P}_{XY}(x = x_i, y = y_j)$

$= \sum_{i=1}^{n} \cancel{\sum_{i=1}^{n}} x_i y_i \times 1/n$

$\because \hat{P}_{xy}(x = x_i, y = y_j) = 0 \quad \forall \{i \neq j\}$

$= 1/n \quad i = j$

thus, $\hat{E}[xy] = \sum_{i=1}^{n} x_i y_i \times 1/n$

$\hat{E}[xy] = \frac{1}{n} \sum_{i=1}^{n} x_i y_i \quad - (2)$

$\Rightarrow \hat{E}[x] = \sum_{i=1}^{n} x_i \hat{P}_x(x)$

$= \frac{1}{n} \sum_{i=1}^{n} x_i \quad \left\{ \hat{P}(x) = 1/n \text{ for plugin estimator} \right\}$

$\Rightarrow \hat{E}(y) = \sum_{i=1}^{n} y_i \hat{P}_y(y)$

$= \frac{1}{n} \sum_{i=1}^{n} y_i$

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

$\Rightarrow \sigma_x^2 = E[x^2] - (E[x])^2$

$= \sum x^2 p(x) - \left(\sum x\, p(x) \right)^2$

for plugin estimator

$\Rightarrow \hat{\sigma}_x^2 = \sum_{i=1}^{n} x_i^2 \hat{P}_x(x_i) - \left(\sum_{i=1}^{n} x_i P_x(x_i) \right)^2$

$\Rightarrow \hat{P}_x(x_i) = 1/n$

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

$\Rightarrow \sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} x_i\right)^2}$

Similarly,

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

substituting all values in ρ -

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

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

Q.5| Given,

$D = \{X_1, X_2, \dots, X_m\} \rightarrow$ iid $\sim X$.

with true CDF F.

a| $E[\hat{F}] = ?$

By definition and from class, we have,

$$\hat{F}_X(\alpha) = P_r(\hat{X} \leq \alpha) = \frac{\sum_{i=1}^{m} I(x_i \leq \alpha)}{m}$$

$$\therefore E[\hat{F}_\alpha(\alpha)] = E\left[\frac{\sum_{i=1}^{m} I(X_i \leq \alpha)}{m}\right]$$

$$= \frac{1}{m} E\left[\sum_{i=1}^{m} I(X_i \leq \alpha)\right]$$

$$\overset{LOE}{=} \frac{1}{m} \sum_{i=1}^{m} E[I(X_i \leq \alpha)].$$

$\because X_{i_s}$ are iid $E[X_1] = E[X_i]$

$$= \frac{1}{m} \times m \times E[I(X_1 \leq \alpha)]$$

$$= E[I(X_1 \leq \alpha)]$$

$$= P_r(X_1 \leq \alpha) \rightarrow \text{Expectation of Indicator RV.}$$

$$= F_{X_1}(\alpha)$$

$$\boxed{\therefore E[\hat{F}_X(\alpha)] = F_X(\alpha)} \leftarrow \text{true CDF } F - Eq(i)$$

b) $Bias(\hat{F}) = ?$

We know that,

$$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

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

From eq.(i).

$$= F - F$$

$$\therefore Bias(\hat{F}) = 0 \quad \text{— Eq(ii)}.$$

c) $se(\hat{F}) = ?$.

We know that,

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

$$Var(\hat{F}) = Var\left(\frac{\sum_{i=1}^{m} I(X_i \le \alpha)}{m}\right) \text{ — By Definition}$$

$$= \frac{1}{m^2} Var\left(\sum_{i=1}^{m} I(X_i \le \alpha)\right)$$

$$\overset{LOV}{=} \frac{1}{m^2} \sum_{i=1}^{m} Var(I(X_i \le \alpha))$$

$\because$ X_i's are iid, $Var(X_i) = Var(X_i)$.

$$= \frac{1}{m^2} \times m \times Var(I(X_1 \le \alpha)).$$

$$= \frac{1}{n} \times Var(I(X_1 \leq \alpha))$$

By $Var(I) = p(1-p)$

$$= \frac{1}{n} \times P(X_1 \leq \alpha)(1 - P(X_1 \leq \alpha))$$

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

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

$$\therefore \boxed{se(\hat{F}) = \sqrt{\frac{F(1-F)}{n}}} \quad - Eq(iii)$$

d) To show $\hat{F}$ is consisment estimator, we can show;

① $Bias(\hat{F}) = 0$

② $se(\hat{F}) \to 0$ as $n \to \infty$.

From Eq(ii), we have $Bias(\hat{F}) = 0$.
From Eq(iii), as $n \to \infty$, $se(\hat{F}) \to 0$.

Both the properties are satisfied, so, $\hat{F}$ is consistent estimator.

Q. 6) a) Given, $\hat{\theta} = \frac{1}{m} \sum_{i=1}^{m} X_i$, X_is $\overset{iid}{\sim}$ Bernoulli (θ).

To find,

i) Bias $(\hat{\theta}) = ?$

ii) se $(\hat{\theta}) = ?$

iii) MSE $(\hat{\theta}) = ?$

i) Bias $(\hat{\theta}) = E[\hat{\theta}] - \theta$.

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

$$= \frac{1}{m} E\left[\sum_{i=1}^{m} X_i\right]$$

$$\overset{LOE}{=} \frac{1}{m} \times \sum_{i=1}^{m} E[X_i]$$

$$\overset{iid}{=} \frac{1}{m} \times m \times E[X_1]$$

$$= E[X].$$

$\because X \sim$ Bernoulli (θ).

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

$$\therefore \text{Bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

$$= \theta - \theta$$

$$\boxed{\text{Bias}(\hat{\theta}) = 0} \quad \text{— Eq (i).}$$

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

$$Var(\hat{\theta}) = Var\left(\frac{1}{m}\sum_{i=1}^{m} x_i\right)$$

$$= \frac{1}{m^2} Var\left(\sum_{i=1}^{m} X_i\right)$$

$$\overset{LOV}{=} \frac{1}{m^2}\sum_{i=1}^{m} Var(X_i)$$

$$\overset{iid}{=} \frac{1}{m^2} \times m \times Var(X_1)$$

$$= \frac{1}{m} \times Var(X_1).$$

$\because X \sim$ Bernoulli (θ)

$$Var(\hat{\theta}) = \frac{1}{m} \times \theta(1-\theta).$$

$$\therefore \boxed{se(\hat{\theta}) = \sqrt{\frac{\theta(1-\theta)}{m}}} \quad - \; Eq(ii).$$

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

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

$$\overset{LOE}{=} E[\theta^2] - E[2\theta\hat{\theta}] + E[\hat{\theta}^2].$$

$\because \theta$ is true value, it is constant.

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

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

$\therefore E[\hat{\theta}^2] = Var(\hat{\theta}) + (E[\hat{\theta}])^2$ — (iv).

Substituting (iv) in (iii),

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

From Part (i) and (ii),

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

$Var(\hat{\theta}) = \dfrac{\theta(1-\theta)}{n}$.

$$\therefore MSE(\hat{\theta}) = \theta^2 - 2\theta \times \theta + \frac{\theta(1-\theta)}{n} + \theta^2.$$

$$\therefore \boxed{MSE(\hat{\theta}) = \frac{\theta(1-\theta)}{n}}.$$

b) Given $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}_n$

$X_i \overset{iid}{\sim}$ Benoulli (θ).

By CLT,

if $A_1, A_2, \ldots A_n$ are iid RVs with true mean μ and Var σ^2, then

$$\frac{\sum A_i}{n} \overset{n\to\infty}{\sim} Nor\left(\mu, \frac{\sigma^2}{n}\right).$$

$$\therefore \quad \hat{\theta} \overset{n\to\infty}{\sim} \text{Normal}\left(E[X_i], \frac{Var(X_i)}{n} \right)$$

$$\sim \text{Normal}\left(\theta, \frac{\theta(1-\theta)}{n} \right).$$

By Normal based CI, $(1-\alpha)$ CI of $\hat{\theta}$ is

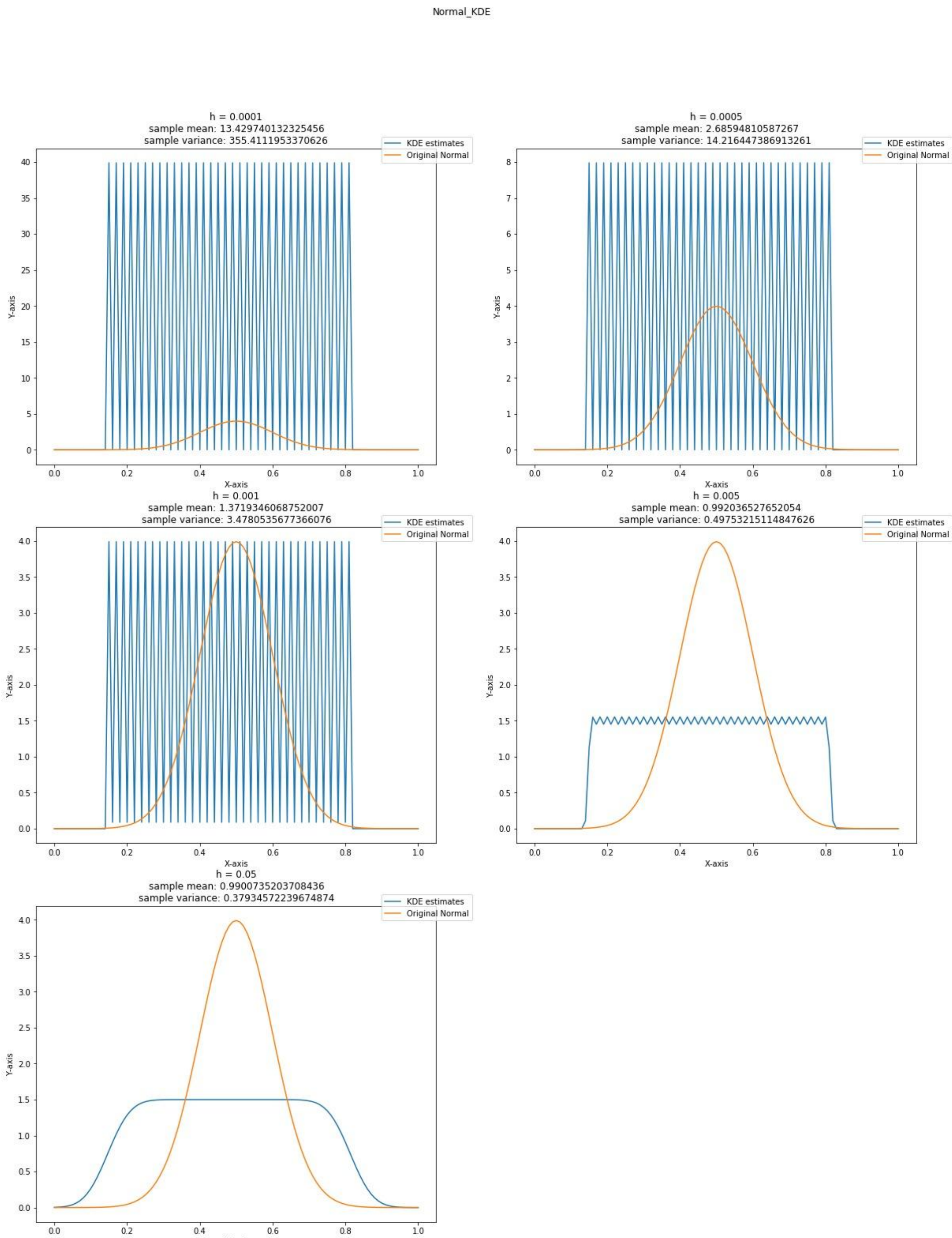
$$(\hat{\theta} - Z_{\alpha/2}\, se \,,\ \hat{\theta} + Z_{\alpha/2}\, se).$$

$\therefore$ $(1-\alpha)$ CI for $\hat{\theta}$ is,

$$\boxed{\left(\hat{\theta} - Z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}} \,,\ \hat{\theta} + Z_{\alpha/2}\sqrt{\frac{\theta(1-\theta)}{n}} \right)}$$

Q7 a] Code attached in the zip file.

Q7 b] H = 0.05 performs best because deviation from true mean and variance is decreasing as h is increasing.



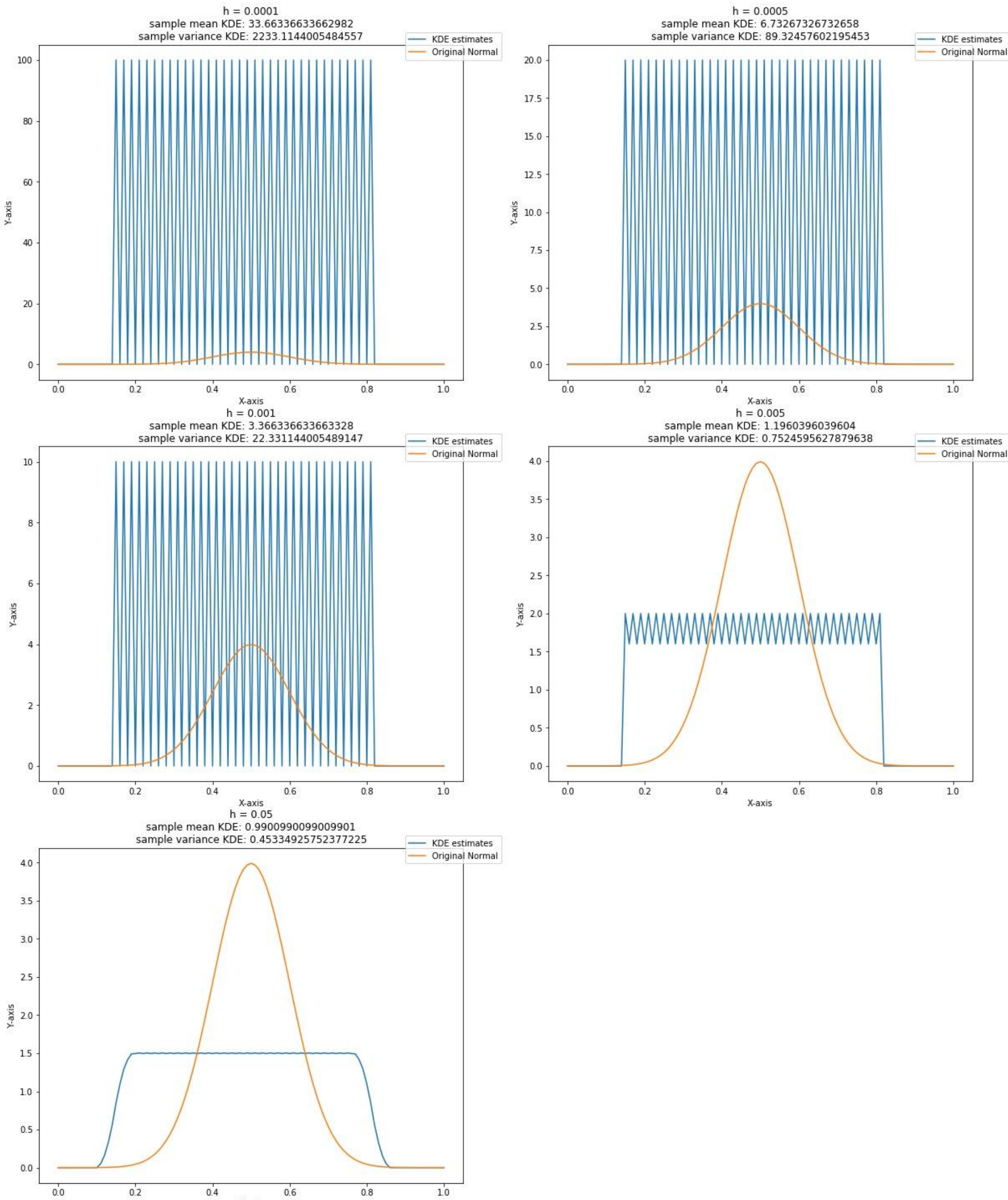
	TRIANGULAR KDE			
H	Sample Mean	Sample Variance	Deviation from Mean (in %)	Deviation from Variance(in %)
h=0.0001	33.66	2233	6632	22331044
h=0.0005	6.73	89.32	1246	893145
h=0.001	3.36	22.33	573	223211
h=0.005	1.19	0.75	139.2	7424
h=0.05	0.99	0.45	98.01	4433

	UNIFORM KDE			
H	Sample Mean	Sample Variance	Deviation from Mean (in %)	Deviation from Variance(in %)
h=0.0001	16.83	558.27	3266	5582686
h=0.0005	3.36	22.33	573.26	223211
h=0.001	1.68	5.58	236.63	55727
h=0.005	0.99	0.66	98.01	6532
h=0.05	0.99	0.43	98.81	4232

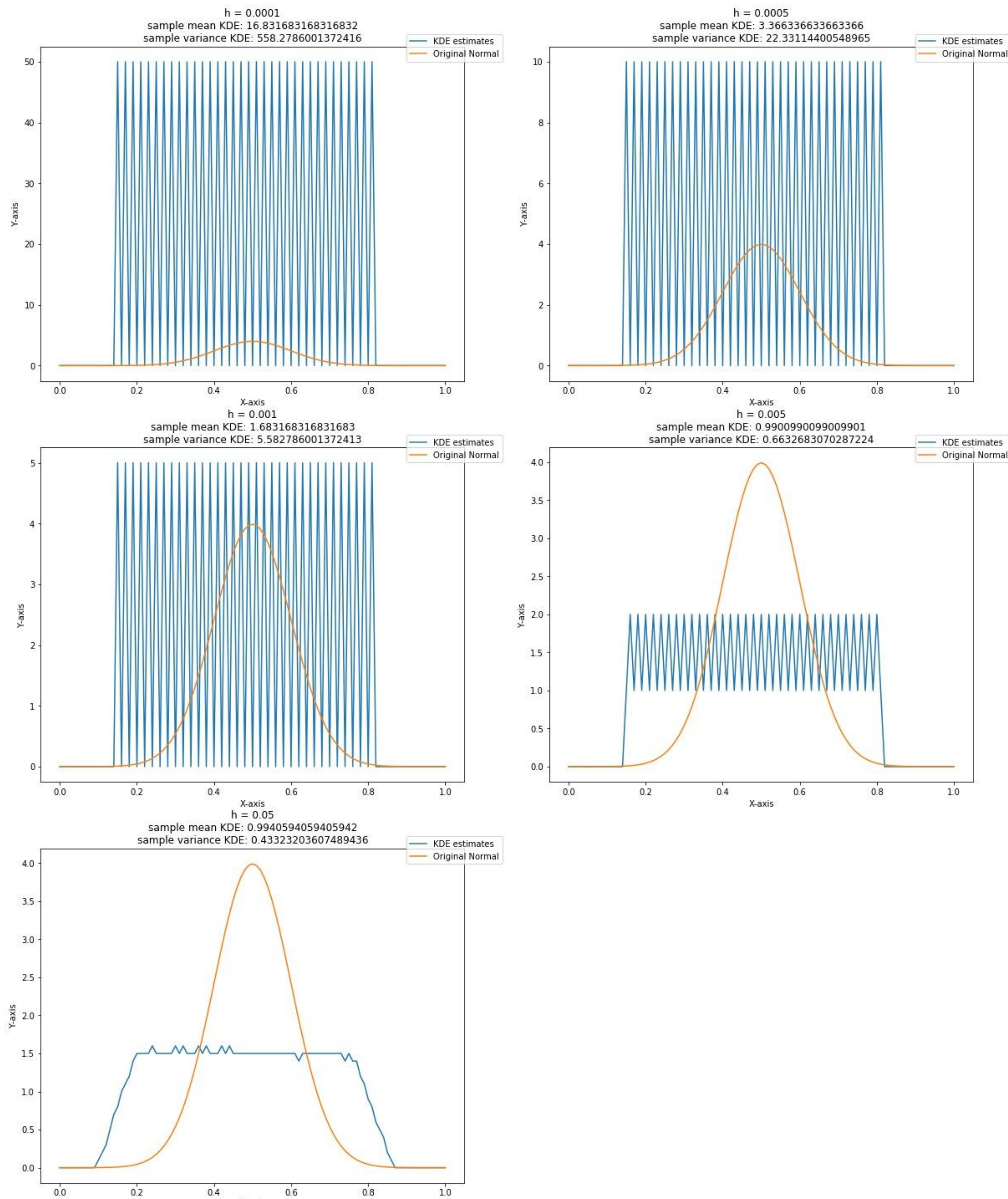
	NORMAL KDE			
H	Sample Mean	Sample Variance	Deviation from Mean (in %)	Deviation from Variance(in %)
h=0.0001	13.42	355.41	2585.948	3554011
h=0.0005	2.68	14.21	437.1896	142064
h=0.001	1.37	3.47	174.3869	34680
h=0.005	0.992	0.49	98.4073	4875
h=0.05	0.99	0.37	98.0147	3693

Q7 c] Triangular h=0.05 performs best.
Uniform h=0.05 performs best.

Triangular_KDE



Uniform_KDE


h = 0.0001
sample mean KDE: 16.831683168316832
sample variance KDE: 558.2786001372416
KDE estimates
Original Normal
Y-axis
X-axis
h = 0.0005
sample mean KDE: 3.366336633663366
sample variance KDE: 22.33114400548965
KDE estimates
Original Normal
Y-axis
X-axis
h = 0.001
sample mean KDE: 1.683168316831683
sample variance KDE: 5.582786001372413
KDE estimates
Original Normal
Y-axis
X-axis
h = 0.005
sample mean KDE: 0.9900990099009901
sample variance KDE: 0.6632683070287224
KDE estimates
Original Normal
Y-axis
X-axis
h = 0.05
sample mean KDE: 0.9940594059405942
sample variance KDE: 0.43323203607489436
KDE estimates
Original Normal
Y-axis
X-axis