Problem 1

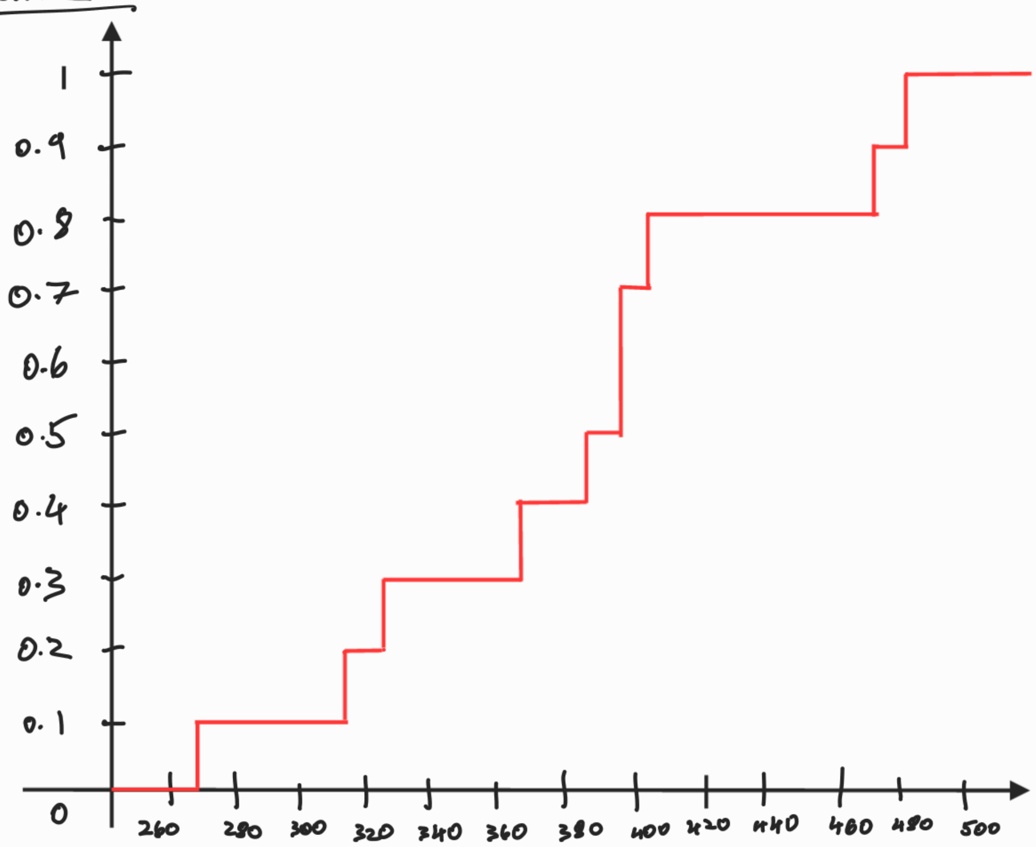
$$MSE = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2]$$
$$MSE = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2$$

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

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

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

Problem 2



Problem 3

(a)

$$\hat{F}(\alpha) = \frac{\sum_{i=1}^{n} I(X_i < \alpha)}{n}$$
$$E[\hat{F}(\alpha)] = E[\frac{\sum_{i=1}^{n} I(X_i < \alpha)}{n}]$$
$$E[\hat{F}(\alpha)] = \frac{\sum_{i=1}^{n} E[I(X_i < \alpha)]}{n} \qquad \text{By L.O.E}$$
$$E[\hat{F}(\alpha)] = \frac{n * E[I(X_i < \alpha)]}{n} \qquad X_i s \text{ are iid}$$
$$E[\hat{F}(\alpha)] = E[I(X_i < \alpha)] = Pr(X_i < \alpha)$$
$$E[\hat{F}(\alpha)] = F(\alpha)$$

(b)

$$Bias(\hat{F}(\alpha)) = E[\hat{F}(\alpha)] - F(\alpha) = F(\alpha) - F(\alpha) = 0$$

(c)

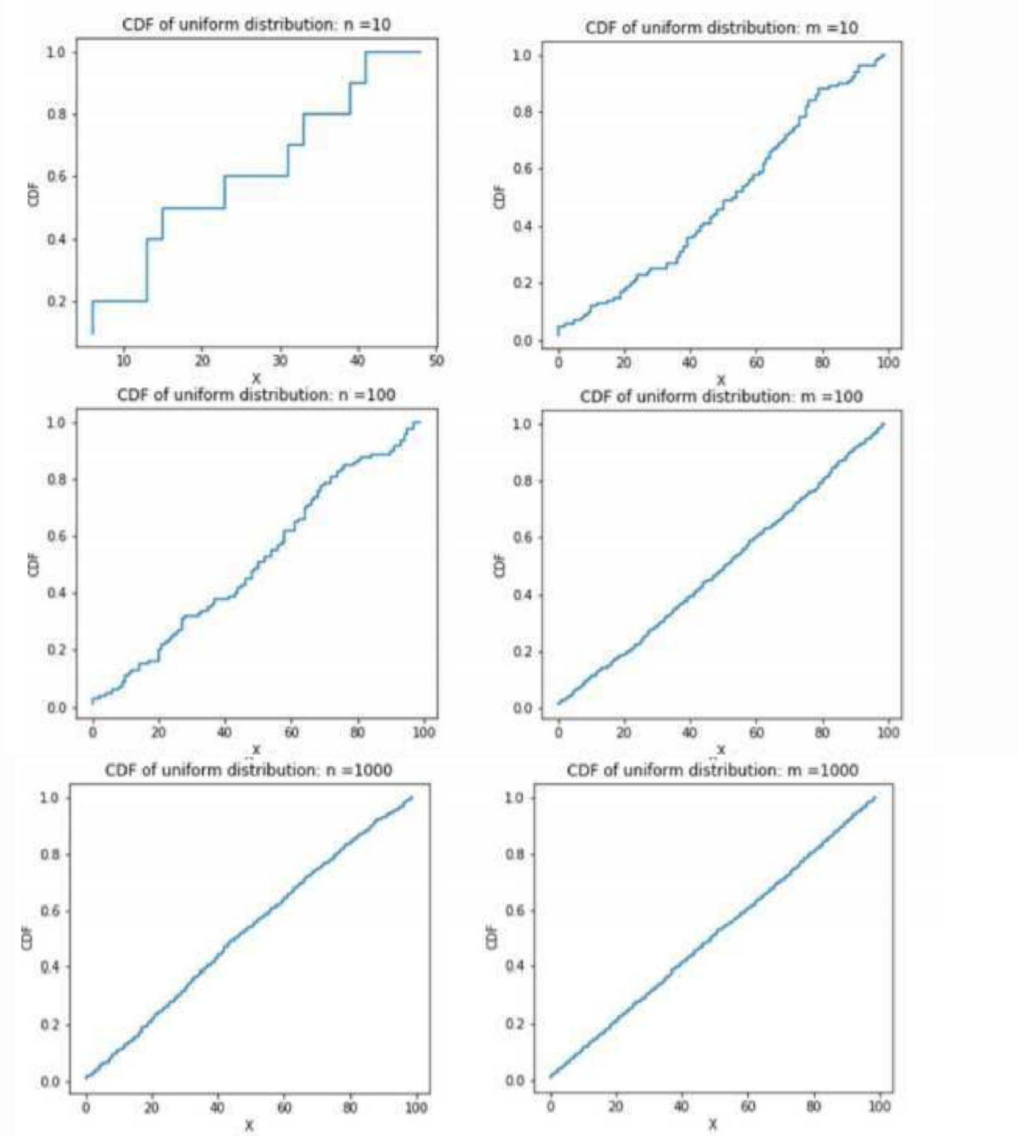
$$SE(\hat{F}(\alpha)) = \sqrt{Var(\hat{F}(\alpha))}$$
$$Var(\hat{F}(\alpha)) = Var(\frac{\sum_{i=1}^{n} I(X_i < \alpha)}{n})$$
$$Var(\hat{F}(\alpha)) = \frac{1}{n^2} Var(\sum_{i=1}^{n} I(X_i < \alpha))$$
$$Var(\hat{F}(\alpha)) = \frac{n}{n^2} Var(I(X_i < \alpha)) \qquad X_i s \text{ are iid}$$
$$Var(\hat{F}(\alpha)) = \frac{1}{n} Var(I(X_i < \alpha))$$

(d) $As\ n \to \infty\ Bias\left(\hat{F}(\alpha)\right) = 0\ and\ Se\left(\hat{F}(\alpha)\right) \to 0, \therefore\ \hat{F}$ is consistent estimator of F.

Problem 4

4b

4d

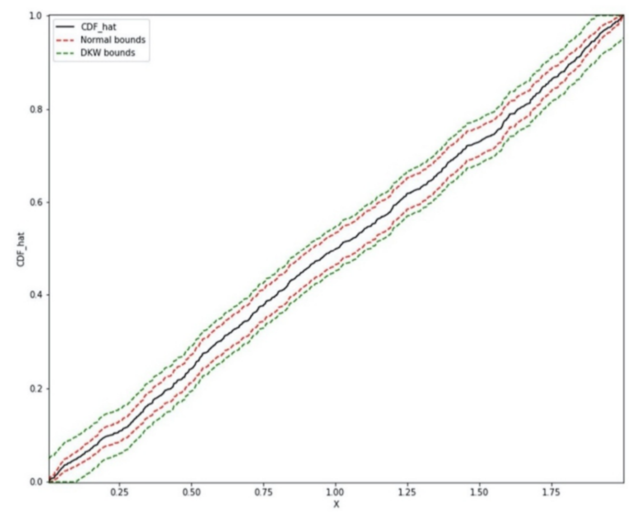


OBSERVATIONS

3b) As value of n (sample size) increases the CDF estimate becomes smoother and the estimated CDF approaches the true CDF.

3d) As value of m (# of rows) increases the CDF estimate becomes smoother and approaches the true CDF even with small sample size because m list of n samples is equivalent to a sample of size n*m.

4e, 4f)



From the figure we can see that Normal bound is tighter than DKW bound.

Problem 5

a) Let $\hat{\sigma}$ be plugin estimator of σ & $\bar{x}_n$ be plugin estimator for mean μ.

We know, $E[X] = \sum_i x_i\, p(x_i)$. Using plugin estimator for $p(x_i)$

we get $p(x_i) = 1/n$ where n = sample size.

$\therefore E[X] = \frac{1}{n}\sum_i x_i = \bar{x}_n$; $E[X^2] = \sum_i x_i^2\, p(x_i) = \sum_i x_i^2\, \hat{p}(x_i)$
$= \frac{1}{n}\sum x_i^2$

$\therefore \hat{\sigma}^2 = E[X^2] - (E[X])^2 = \frac{1}{n}\sum_i x_i^2 - \left(\frac{1}{n}\sum x_i\right)^2 = \frac{1}{n}\sum_i x_i^2 - \bar{x}_n^2$ — ①

$$\text{R.H.S.} = \frac{1}{n}\sum_i (x_i - \bar{x}_n)^2 = \frac{1}{n}\sum_i (x_i^2 - 2x_i\bar{x}_n + \bar{x}_n^2)$$

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

$$= \frac{1}{n}\sum_i x_i^2 - 2\bar{x}_n\cdot\bar{x}_n + \bar{x}_n^2\cdot\cancel{\frac{n}{n}}$$

$$= \frac{1}{n}\sum_i x_i^2 - \bar{x}_n^2 \quad \text{—②}$$

From ① & ②

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

(b)

The bias of estimator $\hat{\sigma}^2$ is

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

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

$$= E\left[\frac{1}{n}\sum_i (x_i-\mu)^2 - \frac{2}{n}(\bar{x}_n-\mu)\sum_i (x_i-\mu) + (\bar{x}_n-\mu)^2\right]$$

$$= E\left[\frac{1}{n}\sum_i (x_i-\mu)^2 - \frac{2}{n}(\bar{x}_n-\mu)\cdot n\cdot(\bar{x}_n-\mu) + (\bar{x}_n-\mu)^2\right]$$

$$= E\left[\frac{1}{n}\sum_i (x_i-\mu)^2 - (\bar{x}_n-\mu)^2\right]$$

$$= E\left[\frac{1}{n}\sum_i (x_i-\mu)^2\right] - E\left[(\bar{x}_n-\mu)^2\right]$$

$$= \sigma^2 - E\left[(\bar{x}_n-\mu)^2\right]$$

$$= \sigma^2 - \mathrm{Var}(\bar{x}_n) = \sigma^2 - \mathrm{Var}\left(\frac{1}{n}\sum_i x_i\right)$$

$$= \sigma^2 - \frac{1}{n^2}\sum_i \mathrm{Var}(x_i) = \sigma^2 - \frac{1}{n}\cdot\sigma^2$$

Thus bias is $E[\hat{\sigma}^2] - \sigma^2 = -\frac{1}{n}\sigma^2$

(c) Let $\hat{\sigma}^2$, $\hat{\mu}$ be plugin estimator for σ^2 & μ.

From part A we know $\hat{\sigma}^2 = \frac{1}{n} \sum_i (x_i - \bar{x}_n)^2$ — ①

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

$$E[(X-\mu)^4] = \sum_i (x_i-\mu)^4 \cdot P(x_i) = \sum_i (x_i-\mu)^4 \cdot \hat{P}(x_i)$$

$$= \frac{1}{n} \sum_i (x_i - \hat{\mu})^4 \quad \{\text{Plugin estimator}\} \quad — ②$$

From ①

$$\hat{\sigma}^4 = \frac{1}{n^2} \left(\sum_i (x_i - \bar{x}_n)^2 \right)^2 \quad — ③$$

From ② & ③ & by definition of Kurt[X], we have:

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

$$= \frac{n \sum_i (x_i - \bar{x}_n)^4}{\left(\sum_i (x_i - \bar{x}_n)^2 \right)^2}$$

(d)

$$\rho = \frac{E[XY] - E[X]E[Y]}{\sigma_x \sigma_y}$$

$$\rho = \frac{E[(X - E[X])(Y - E[Y])]}{\sigma_x \sigma_y}$$

$$\hat{\rho} = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{(\sum_{i=1}^{n} (X_i - \bar{X})^2} \sqrt{(\sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

Problem 6

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

$$\Rightarrow \; Bias(\hat{\theta}) \; = E\left[\frac{1}{n}\sum_i X_i\right] - \theta$$

$$\Rightarrow Bias(\hat{\theta}) \; = \frac{1}{n}\sum_i E[X_i] - \theta \quad [By\, L.O.E]$$

$$\Rightarrow Bias(\hat{\theta}) = \frac{1}{n} n.\theta - \theta$$

$$\Rightarrow Bias(\hat{\theta}) = 0$$

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

$$\Rightarrow Var(\hat{\theta}) = \frac{1}{n^2}\sum_i Var(\,X_i) \quad [\because X_i s\ are\ iid]$$

$$\Rightarrow Var(\hat{\theta}) = \frac{1}{n^2} n\theta = \frac{\theta}{n}$$

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

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

$$MSE(\hat{\theta}) = \frac{\theta}{n}$$

Problem 7

a

First, we know that $\hat{\theta} = \hat{F_n}(b) - \hat{F_n}(a) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \in (a,b])$. This is the sum of n independent Bernoulli random variables with success probability $F(b) - F(a)$.

Thus,

$$
\begin{aligned}
Var(\hat{\theta}) &= Var(\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(X_i \in (a,b])) \\
&= \frac{1}{n^2} Var(\sum_{i=1}^{n} \mathbf{1}(X_i \in (a,b])) \\
&= \frac{1}{n} Var(\mathbf{1}(X_i \in (a,b])) \\
&= \frac{1}{n}(F(b) - F(a))(1 - F(b) + F(a)) \\
&= \frac{\theta(1-\theta)}{n}
\end{aligned}
$$

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

b

Way 1

Now we could get the CI for $\hat{\theta}$ based on the DKW inequality.

$$
\begin{aligned}
L(x) &= max(\hat{\theta} - \epsilon_n, 0) \\
R(x) &= min(\hat{\theta} + \epsilon_n, 1)
\end{aligned}
$$

where $\epsilon_n = \sqrt{\frac{1}{2n}\log(\frac{2}{\alpha})}$

Way 2

Since we already know that $E[\hat{\theta}] = \theta$, the confident interval can be derived by

$$
\begin{aligned}
&P(-z_\alpha \le \frac{\hat{\theta} - E[\hat{\theta}]}{se(\hat{\theta})} \le z_\alpha) \\
=&P(-z_\alpha \le \frac{\hat{\theta} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \le z_\alpha) \\
=&P(-z_\alpha\sqrt{\frac{\theta(1-\theta)}{n}} \le \hat{\theta} - \theta \le z_\alpha\sqrt{\frac{\theta(1-\theta)}{n}}) \\
=&P(\theta - z_\alpha\sqrt{\frac{\theta(1-\theta)}{n}} \le \hat{\theta} \le \theta + z_\alpha\sqrt{\frac{\theta(1-\theta)}{n}})
\end{aligned}
$$

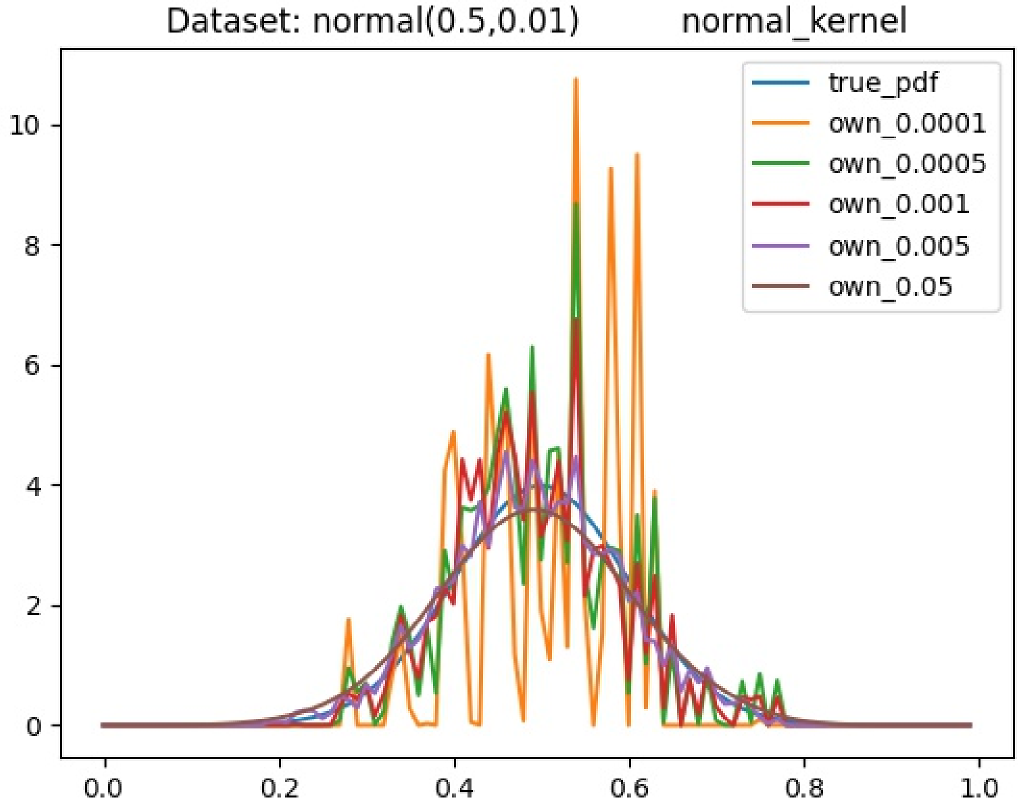
Thus, the CI for $\hat{\theta}$ is

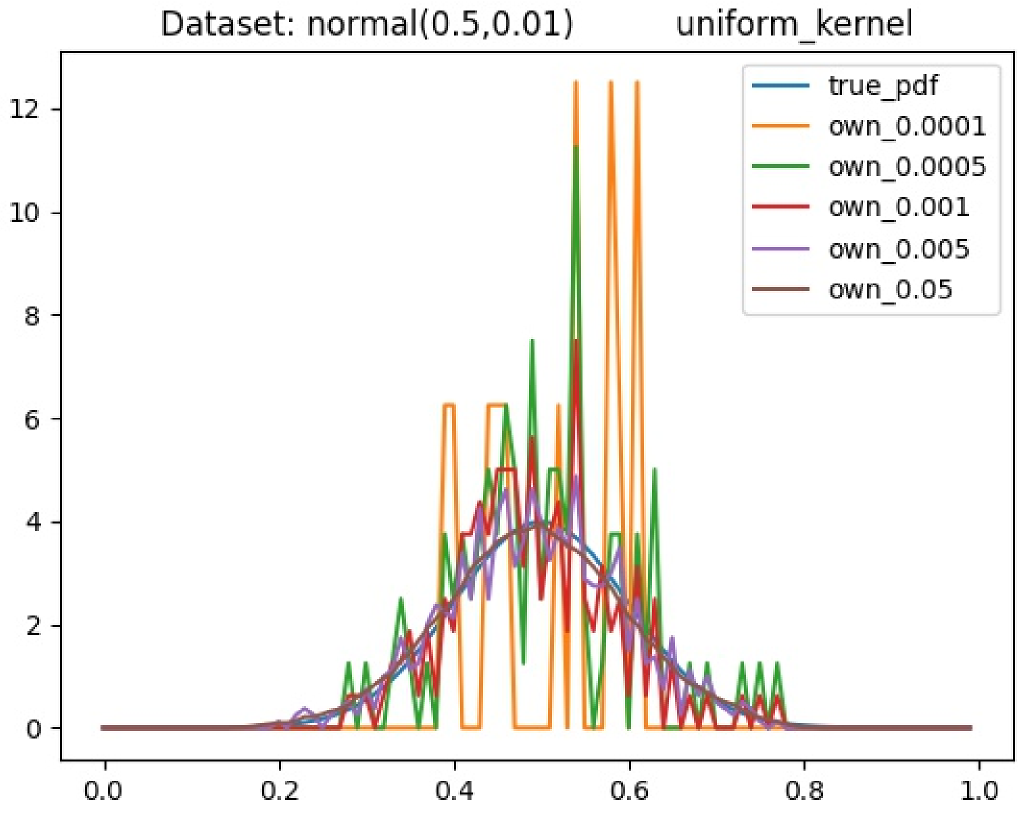
$$
\begin{aligned}
L(x) &= \theta - z_\alpha\sqrt{\frac{\theta(1-\theta)}{n}} \\
R(x) &= \theta + z_\alpha\sqrt{\frac{\theta(1-\theta)}{n}}
\end{aligned}
$$

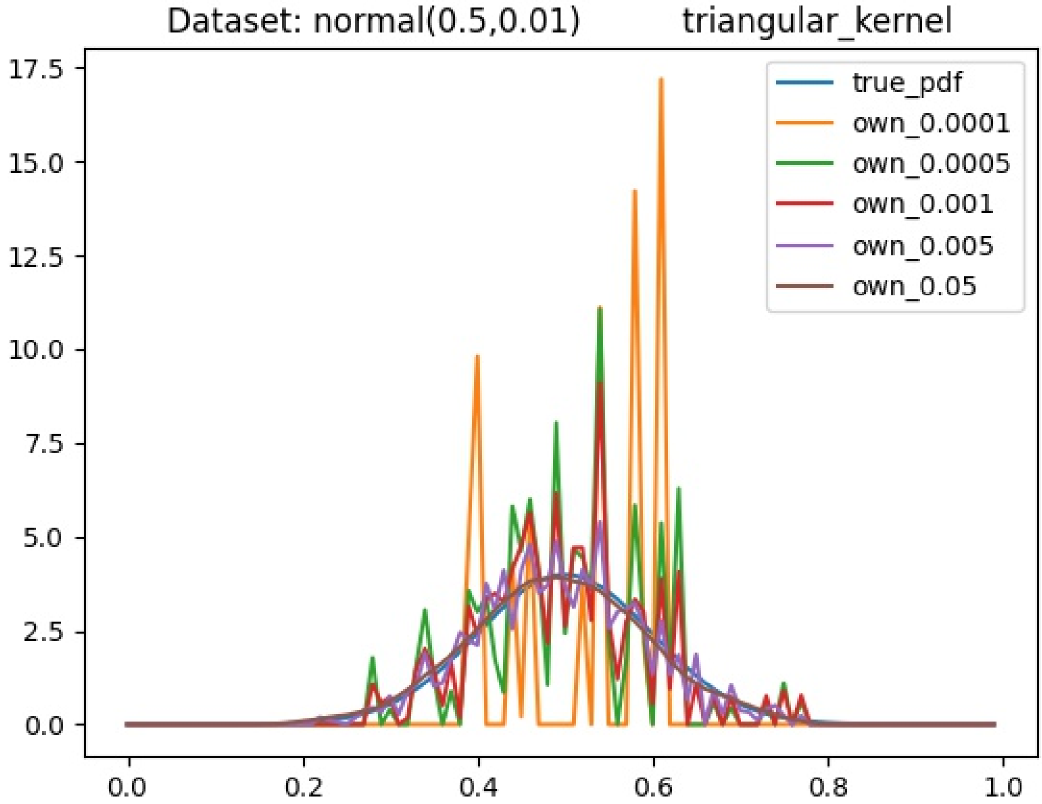
Problem 8

Dataset	Kernel	Bandwidth	%deviation mean	%deviation variance
normal(0.5,0.01)	normal_kernel	0.0001	13.53	-33.92
normal(0.5,0.01)	normal_kernel	0.0005	8.81	-19.03
normal(0.5,0.01)	normal_kernel	0.001	4.57	-8.44
normal(0.5,0.01)	**normal_kernel**	**0.005**	**0.16**	**-0.81**
normal(0.5,0.01)	normal_kernel	0.05	0.22	2.16
normal(0.5,0.01)	uniform_kernel	0.0001	17.03	-51.73
normal(0.5,0.01)	uniform_kernel	0.0005	11.52	-26.62
normal(0.5,0.01)	uniform_kernel	0.001	2.08	-2.9
normal(0.5,0.01)	uniform_kernel	0.005	0.13	-0.85
normal(0.5,0.01)	**uniform_kernel**	**0.05**	**0.16**	**0.24**
normal(0.5,0.01)	triangular_kernel	0.0001	5.55	-28.91
normal(0.5,0.01)	triangular_kernel	0.0005	14.76	-37.24
normal(0.5,0.01)	triangular_kernel	0.001	9.76	-21.72
normal(0.5,0.01)	triangular_kernel	0.005	1.45	-2.83
normal(0.5,0.01)	**triangular_kernel**	**0.05**	**0.15**	**-0.3**
a3_q8.csv	normal_kernel	0.0001	13.53	-33.92
a3_q8.csv	normal_kernel	0.0005	8.81	-19.03
a3_q8.csv	normal_kernel	0.001	4.57	-8.44
a3_q8.csv	**normal_kernel**	**0.005**	**0.16**	**-0.81**
a3_q8.csv	normal_kernel	0.05	0.22	2.16
a3_q8.csv	uniform_kernel	0.0001	17.03	-51.73
a3_q8.csv	uniform_kernel	0.0005	11.52	-26.62
a3_q8.csv	uniform_kernel	0.001	2.08	-2.9
a3_q8.csv	uniform_kernel	0.005	0.13	-0.85
a3_q8.csv	**uniform_kernel**	**0.05**	**0.16**	**0.24**
a3_q8.csv	triangular_kernel	0.0001	5.55	-28.91
a3_q8.csv	triangular_kernel	0.0005	14.76	-37.24
a3_q8.csv	triangular_kernel	0.001	9.76	-21.72
a3_q8.csv	triangular_kernel	0.005	1.45	-2.83
a3_q8.csv	**triangular_kernel**	**0.05**	**0.15**	**-0.3**

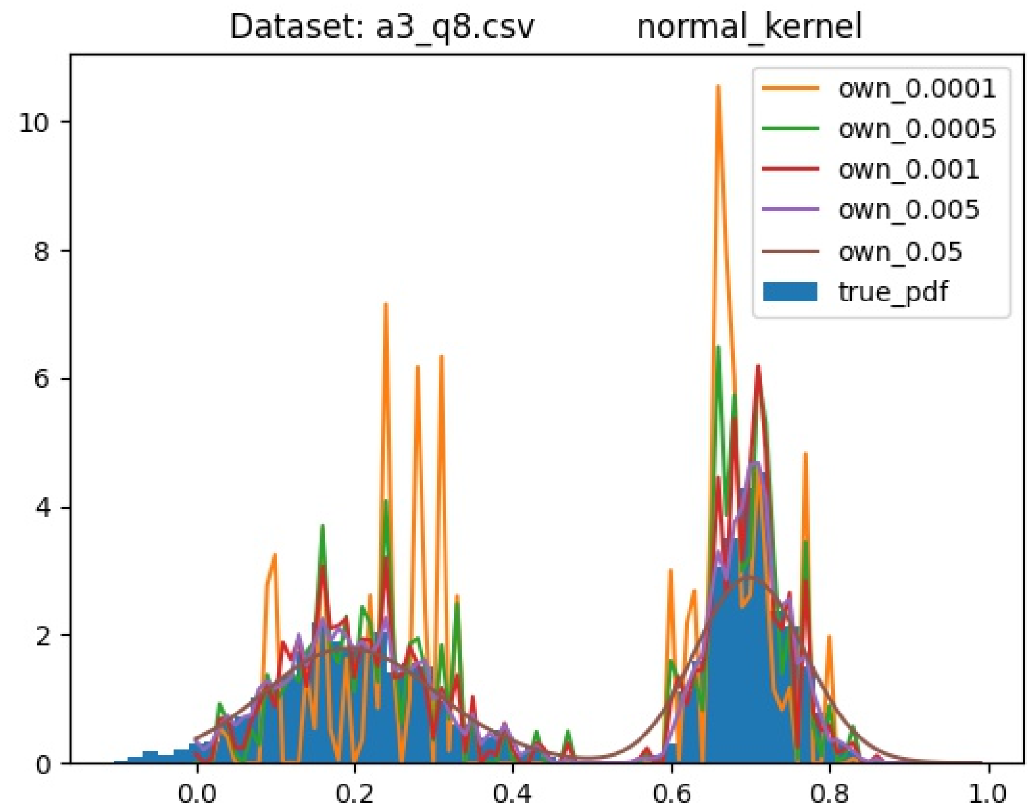
Best bandwidth and kernel combinations are highlighted in bold. However, the “best” selection is subjective.


Dataset: normal(0.5,0.01) normal_kernel
true_pdf
own_0.0001
own_0.0005
own_0.001
own_0.005
own_0.05
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0

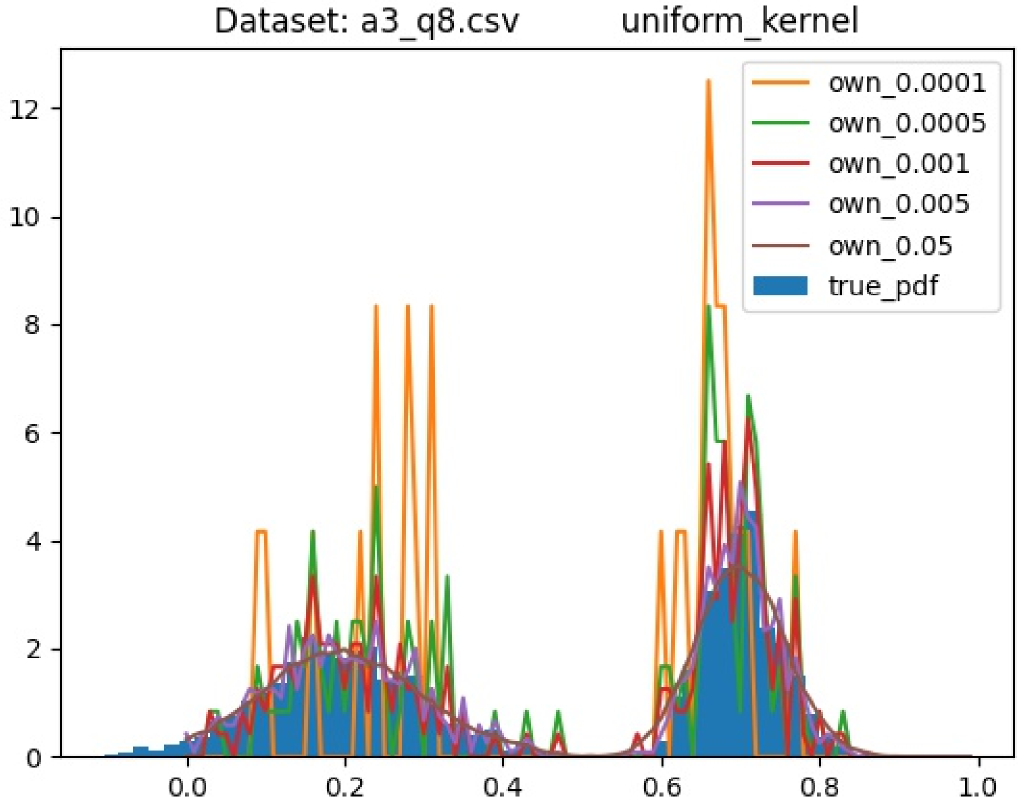

Dataset: normal(0.5,0.01) uniform_kernel
true_pdf
own_0.0001
own_0.0005
own_0.001
own_0.005
own_0.05
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0



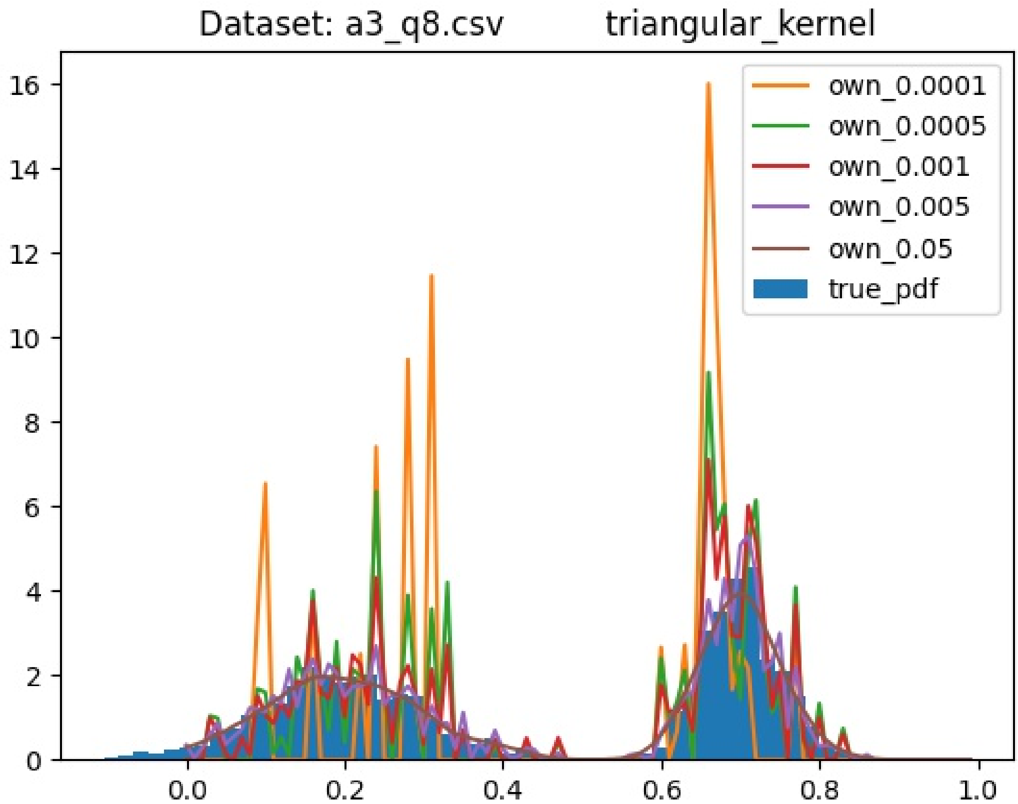
Dataset: normal(0.5,0.01) triangular_kernel
true_pdf
own_0.0001
own_0.0005
own_0.001
own_0.005
own_0.05
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
0.0
0.2
0.4
0.6
0.8
1.0



Dataset: a3_q8.csv normal_kernel
own_0.0001
own_0.0005
own_0.001
own_0.005
own_0.05
true_pdf
10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0



Dataset: a3_q8.csv uniform_kernel
own_0.0001
own_0.0005
own_0.001
own_0.005
own_0.05
true_pdf
0
2
4
6
8
10
12
0.0
0.2
0.4
0.6
0.8
1.0



Dataset: a3_q8.csv triangular_kernel
own_0.0001
own_0.0005
own_0.001
own_0.005
own_0.05
true_pdf
0
2
4
6
8
10
12
14
16
0.0
0.2
0.4
0.6
0.8
1.0