CSE 544, Spring 2020, Probability and Statistics for Data Science

Assignment 3: Non Parametric Inference Due: 3/4, in class

(7 questions, 75 points total)

I/We understand and agree to the following:
(a) Academic dishonesty will result in an 'F' grade and referral to the Academic Judiciary.
(b) Late submission, beyond the 'due' date/time, will result in a score of 0 on this assignment.
(write down the name of all collaborating students on the line below)

---

1. (5 pts)

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

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

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

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

$$Bias^2\left(\hat{\theta}\right) = E^2\left[\hat{\theta}\right] + \theta^2 - 2\theta E\left[\hat{\theta}\right] - (3)$$

$From\ (3) + (2), we\ get$

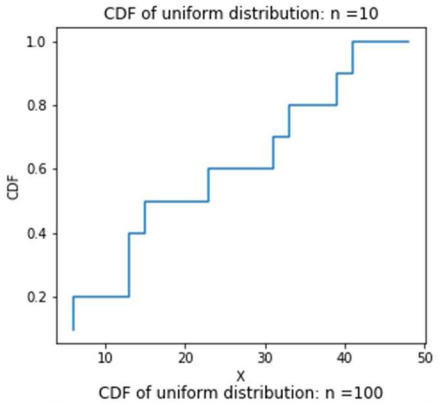
$$Var\left(\hat{\theta}\right) + Bias^2\left(\hat{\theta}\right) = E\left[\widehat{\theta^2}\right] + \theta^2 - 2\theta E\left[\hat{\theta}\right] - (4)$$

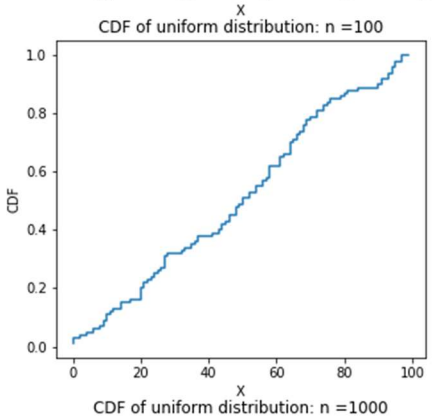
$From\ (1)\ and\ (4), we\ get$

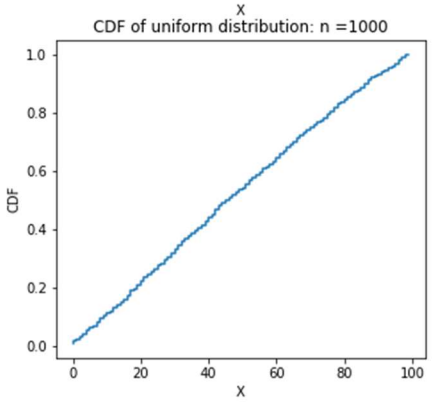
$$MSE = Var\left(\hat{\theta}\right) + Bias^2\left(\hat{\theta}\right)$$

2. (3+2+3+2+3+4 pts)

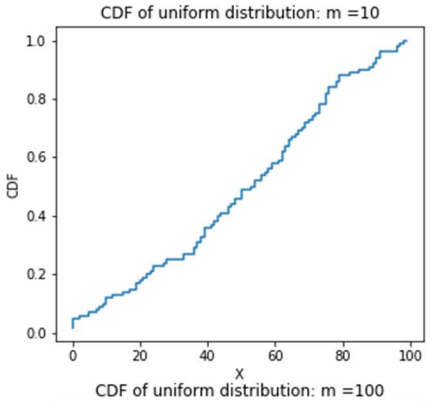
2. b

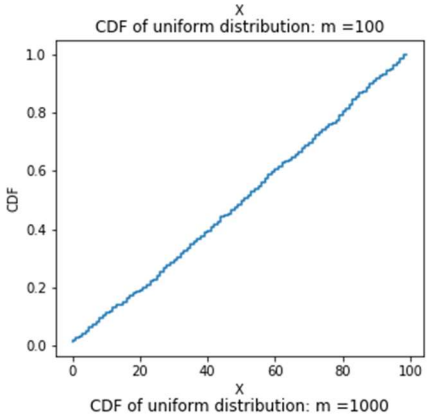


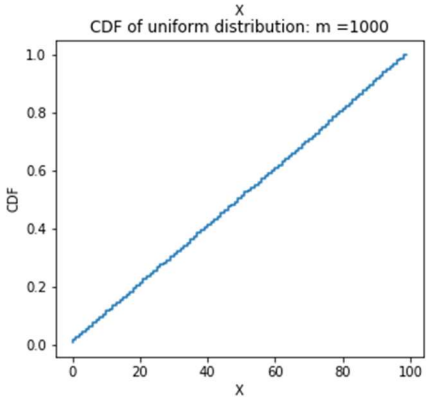




2. d





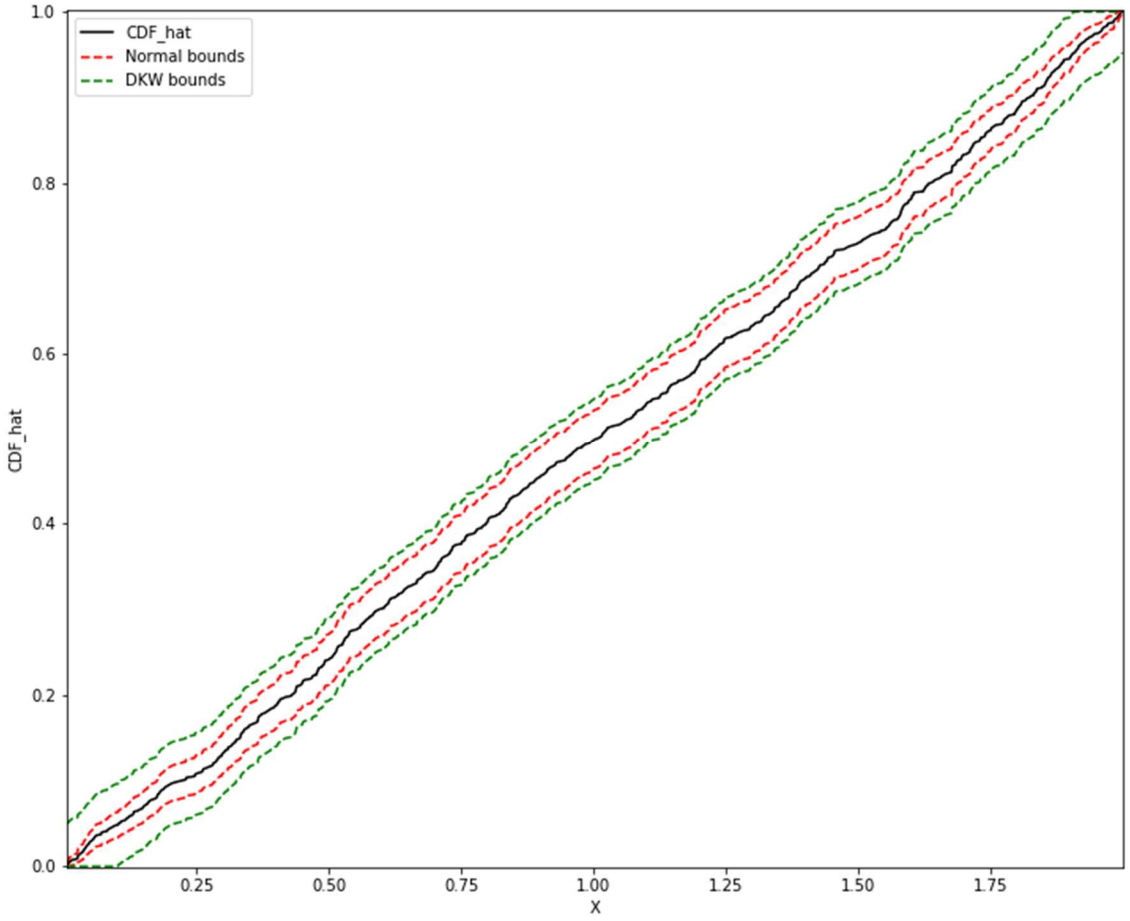


OBSERVATIONS

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

2.d: 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.

2.e, f



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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \sigma^2 - \frac{1}{n^2}\sum_{i}Var(X_i) = \sigma^2 - \frac{1}{n}\cdot\sigma^2$$

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

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

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

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

From ①

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

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

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

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

4.a

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

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

$$E\left[\hat{F}(\alpha)\right] = \frac{\sum_{i=1}^{n} E[I(X_i < \alpha)]}{n} \quad By\ L.O.E$$

$$E\left[\hat{F}(\alpha)\right] = \frac{n.E[I(X_i < \alpha)]}{n} \qquad \because X_i s\ are\ iid$$

$$E\left[\hat{F}(\alpha)\right] = E[I(X_i < \alpha)]$$

$$E\left[\hat{F}(\alpha)\right] = \Pr(X_i < \alpha)$$

$$E\left[\hat{F}(\alpha)\right] = F(\alpha)\ -(1)$$

4.b

$$Bias\left(\hat{F}(\alpha)\right) = E\left[\hat{F}(\alpha)\right] - F(\alpha) - (2)$$

From $(1)\ and\ (2)$, we get $Bias\left(\hat{F}(\alpha)\right) = 0$

4.c

$$Se\left(\hat{F}(\alpha)\right) = \sqrt{Var\left(\hat{F}(\alpha)\right)}$$

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

$$Var\left(\hat{F}(\alpha)\right) = \frac{1}{n^2} Var\left(\sum_{i=1}^{n} I(X_i < \alpha)\right)$$

$$Var\left(\hat{F}(\alpha)\right) = \frac{n}{n^2} Var(I(X_i < \alpha)) \because X_i s\ are\ iid$$

$$Var\left(\hat{F}(\alpha)\right) = \frac{1}{n} Var(I(X_i < \alpha))$$

$$Var\left(\hat{F}(\alpha)\right) = \frac{\left(1 - Pr\ (X_i < \alpha)\right).Pr\ (X_i < \alpha)}{n}$$

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

$$Se\left(\hat{F}(\alpha)\right) = \sqrt{\frac{F(\alpha)\left(1 - F(\alpha)\right)}{n}}$$

4.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.

5. (3+4+3+3)

a. Let x_i be the start point of the bin B_i which is of size b.

Let p_i be the probability that a pint x lies in B_i , i.e $p_i = Pr(x \in (x_i, x_i + b))$

$$\therefore p_i = \frac{1}{n}\sum_{j=1}^{n} I\big(x_j \in (x_i, x_i + b)\big) \; - (1)$$

b. Given $x \in B_j = (x_j, x_j + b)$

From (1), histogram estimate of $x\ is$

$$\hat{h}(x) = \frac{\hat{p}_j}{b} = \frac{1}{nb}\sum_{k=1}^{n} I\big(x_k \in (x_j, x_j + b)\big)$$

$$E\big[\hat{h}(x)\big] = \frac{1}{nb}\sum_{k=1}^{n} E\big[I\big(x_k \in (x_j, x_j + b)\big)\big] \;\; \{By\ L.O.E\}$$

$$E\big[\hat{h}(x)\big] = \frac{1}{nb}\sum_{k=1}^{n} Pr\left(x_k \in (x_j, x_j + b)\right)$$

$$E\big[\hat{h}(x)\big] = \frac{1}{nb}\sum_{k=1}^{n} Pr\left(x_k \in (x_j, x_j + b)\right)$$

$$E\big[\hat{h}(x)\big] = \frac{1}{nb}\sum_{k=1}^{n} \big[F\big(x_j + b\big) - F\big(x_j\big)\big]$$

$$E\big[\hat{h}(x)\big] = \frac{\big[F\big(x_j + b\big) - F\big(x_j\big)\big]}{b}$$

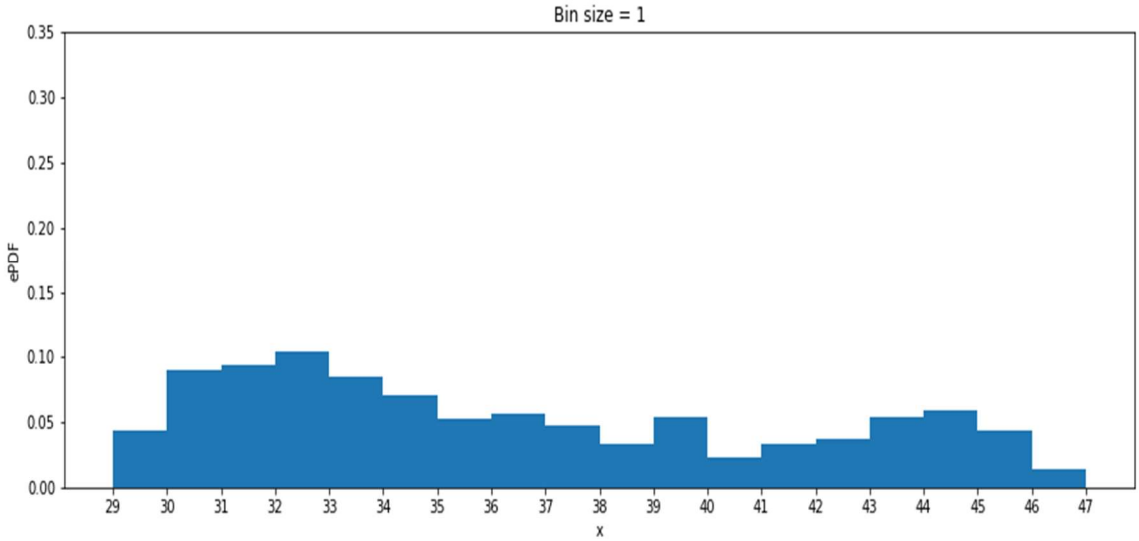
As $b \to 0$, we can say

$$E\big[\hat{h}(x)\big] = \lim_{b \to 0} \frac{\big[F\big(x_j + b\big) - F\big(x_j\big)\big]}{b} = f(x_j)$$

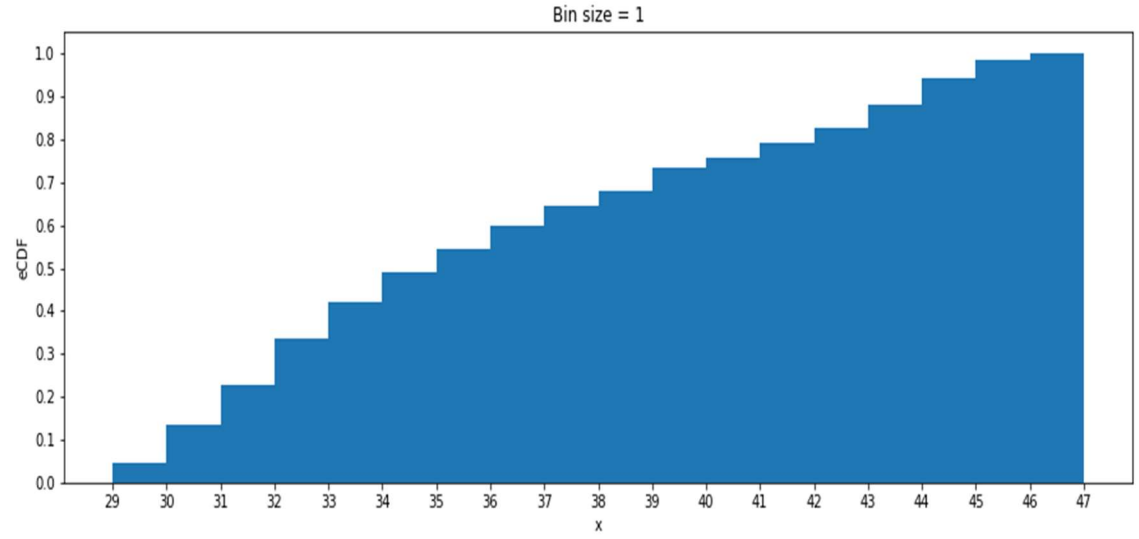
As PDF is derivative of CDF and $x \in \big(x_j, x_j + b\big)$, we can say that $f\big(x_j\big) \to f(x)$

$$\therefore E\big[\hat{h}(x)\big] = f(x)$$

c


Bin size = 1
ePDF
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
x

d


Bin size = 1
eCDF
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
x

6. (5 pts)

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

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

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

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

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

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

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

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

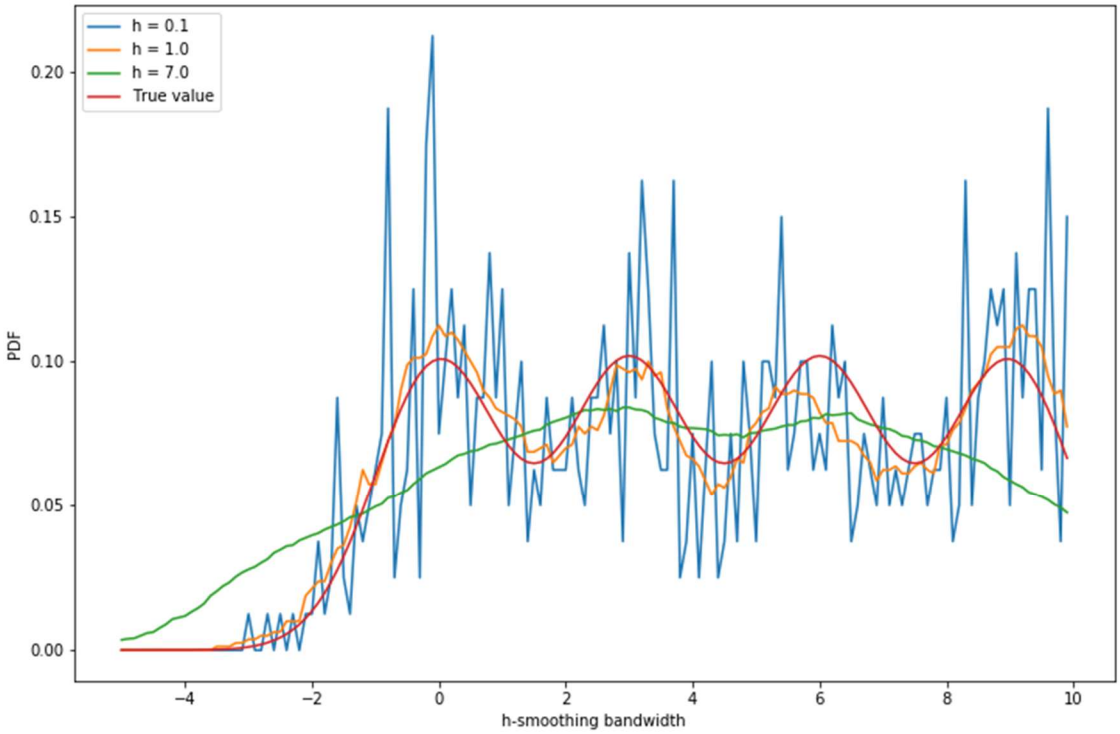
$$\Rightarrow Se(\hat{\theta}) = \sqrt{Var(\hat{\theta})} = \sqrt{\frac{\theta}{n}}$$

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

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

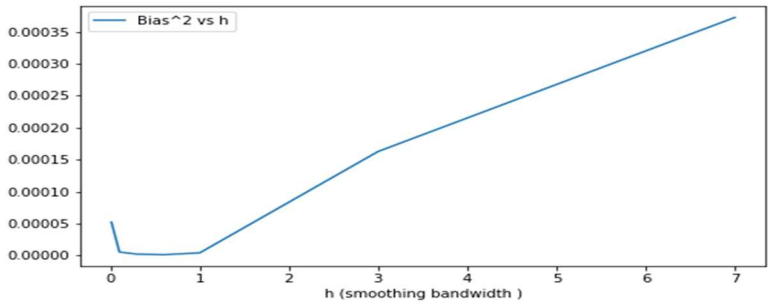
7. (8+ (5+2) pts)

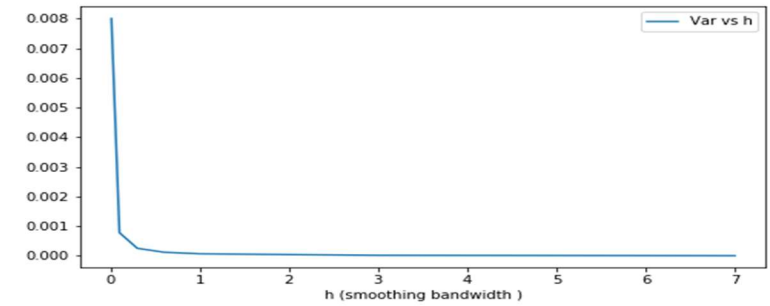
a.

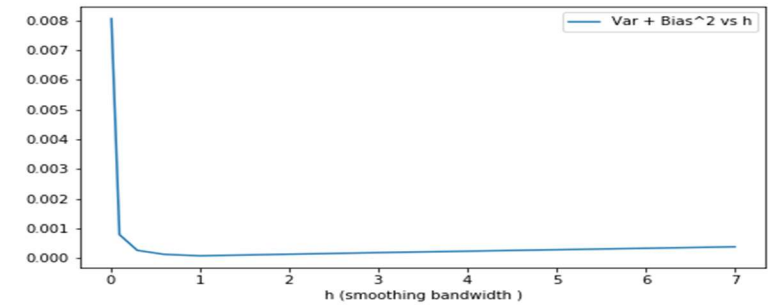


Observations:

1. Size of the smoothing bandwidth is inversely proportional to degree of smoothness of density estimate.
2. Smaller window leads to density estimate based on local configuration of data points.
3. In ideal case, when # of data points $n \to \infty$, window size $h \to 0$ the density estimate asymptotically converges to the true distribution.







7.b Observation: Size of the smoothing window is directly proportional to Bias and inversely proportional to the variance of the density estimate.

Small window size is highly sensitive to noise whereas large window size does not take into account the fine-grained local information.

h	0.01	0.1	0.3	0.6	**1.0**	3.0	7.0
$Bias^2$	5.23E-06	5.61E-06	2.27E-06	1.49E-06	**4.27E-06**	1.63E-04	3.72E-04
Variance	8.01E-03	7.83E-04	2.56E-04	1.23E-04	**7.07E-05**	1.89E-05	5.23E-06

Based on the given measure of $Bias^2$+Variance, the **optimal value of h is 1.0.**