question 1)
abcd

۲۶۰-۱۰۵

(1)

a) $\hat{\theta} = \underset{\theta}{\text{argmax}} \left(\ln \prod_{k=1}^{n} x_k^3 \frac{e^{-x_k/\theta}}{4\theta^4} \right) = \underset{\theta}{\text{argmax}} \sum_{k=1}^{n} \ln P(x_k|\theta)$

$= \underset{\theta}{\text{argmax}} \sum_{k=1}^{n} \left(3 \ln x_k - \frac{x_k}{\theta} - 4 \ln \theta - \ln 6 \right) = \underset{\theta}{\text{argmax}} \sum_{k=1}^{n} -\frac{x_k}{\theta} - 4\ln$

$\frac{\partial}{\partial \theta} = 0 \longrightarrow \sum_{k=1}^{n} \frac{x_k}{\theta^2} - \frac{4}{\theta} = 0 \longrightarrow \theta = \frac{\sum_{k=1}^{n} x_k}{4n}$

b) $-\frac{24}{\theta} + \frac{6.8 + 7.2 + 4.7 + 7.9 + 9.5 + 6.1}{\theta^2} = 0 \longrightarrow 42.2 - 24\theta = 0$

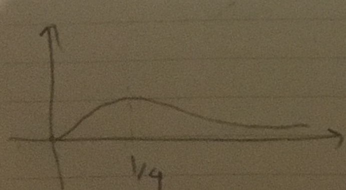
09

$\longrightarrow \theta = \frac{42.2}{24} = 1.75$

c) $\left(\frac{1}{\theta+1}\right)^4 \times \frac{\theta}{\theta+1} = \frac{\theta}{(1+\theta)^5}$

d) $\frac{(1+\theta)^5 - 5\theta(1+\theta)^4}{(1+\theta)^{10}} = 0 \longrightarrow (1+\theta-5\theta)(1+\theta)^4 = 0 \begin{cases} \theta = -1 \\ \theta = 1/4 \end{cases}$

باتوجه به شکل تقریبی داده مقدار بیشینه به ازای θ = ۱ اتفاق می‌افتد



M	T	W	T	F	S	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

efgh)

e) $f_{5}(\theta \mid R_{k}=4) = \dfrac{f_5(R=4 \mid \theta) \times f(\theta)}{\sum_{\theta} f(R=4 \mid \theta) \times p(\theta)} = \dfrac{\dfrac{\theta}{(\theta+1)^5}}{(\frac{1}{2})^5 + (\frac{2}{3})^5 + (\frac{3}{4})^5 + (\frac{4}{5})^5}$

f) $\lambda^2 \times 3e^{-3\lambda} \times \lambda^2 \times 5e^{-5\lambda} \times \lambda^2 \times e^{-\lambda} \times \lambda^2 \times 4e^{-4\lambda} \times \lambda^2 \times 7e^{-7\lambda}$

g) $\sum_i \log \lambda^2 x_i e^{-\lambda x_i}$ ← madimization

$\sum_i \log \lambda^2 + \sum -\lambda x_i \rightarrow \partial/\partial\lambda \rightarrow \sum_i 2 \times \frac{1}{\lambda} - \sum x_i = 0$

$10 \times 1/\lambda - 20 = 0 \rightarrow \lambda = 1/2 = 0{,}5$

h) $f_n(\lambda \mid x) = \dfrac{f(x \mid \lambda) \times 1}{\sum_{\lambda_i} f(x \mid \lambda_i) \times \cancel{f(\lambda_i)}^{=1}} = \dfrac{\lambda^{10} \times e^{-20\lambda} \times \cancel{420}}{\cancel{420} \int_0^1 \lambda^{10} e^{-20\lambda}\, d\lambda}$

$f(\lambda \mid x) = \dfrac{f(x \mid \lambda)}{\int_0^1 f(x \mid \lambda)\, d\lambda} = \dfrac{\lambda^2 x e^{-\lambda x}}{\int_0^1 \lambda^2 x e^{-\lambda x}\, d\lambda}$

i)

١٨|7 الخميس Thu.

٢٦٢-١٠٣ هفته ٣٨

i) $E[\lambda|x] = \int_0^1 \lambda \times f(\lambda|x)\, d\lambda$

question 2)

(a) $$\prod_i \frac{(n_i x_i \lambda)^{k_i} e^{-(n_i x_i \lambda)}}{(k_i)!} = \prod_i \underbrace{\frac{(n_i x_i)^{k_i} e^{-(n_i x_i)}}{(k_i)!}}_{T_i} \times \lambda^{k_i} e^{-\lambda}$$

$$= \prod_i T_i \lambda^{k_i} e^{-\lambda} = e^{-S\lambda} \times \lambda^{(k_1+\cdots+k_S)} \times \prod T_i$$

(b) $$\frac{\partial}{\partial \lambda} \to \log\left(e^{-S\lambda} \times \lambda^{(k_1+\cdots+k_S)} \times T_1 \times \cdots \times T_S\right)$$

$$= -S\lambda + (k_1+\cdots+k_S) \times \log \lambda + \sum_{i=1}^{S} \log T_i$$

$$\to \partial/\partial\lambda = -S + (k_1+\cdots+k_S) \times \frac{1}{\lambda} + 0 = 0$$

$$\to \lambda = \frac{k_1+k_2+\cdots+k_S}{S}$$

۲۷۲-۰۹۳

c) $E(\hat{\lambda}) = \dfrac{\sum_{i=1}^{n} E(k_i)}{n_1x_1+n_2x_2+\cdots+n_sx_s} = \dfrac{\lambda_1+\lambda_2+\cdots+\lambda_s}{n_1x_1+n_2x_2+\cdots+n_sx_s}$

$= \dfrac{n_1x_1\lambda+n_2x_2\lambda+\cdots+n_sx_s\lambda}{n_1x_1+n_2x_2+\cdots+n_sx_s} = \lambda \leftarrow$ unbiased

d) $var(\hat{\lambda}) = var\left(\dfrac{\sum_{i=1}^{n} k_i}{n_1x_1+n_2x_2+\cdots+n_sx_s}\right) = \dfrac{\sum_{i=1}^{n} var(n_ix_i\lambda)}{n_1x_1+n_2x_2+\cdots+n_sx_s}$

$= \lambda$

e) $E(\lambda-\hat{\lambda})^2 = var(\hat{\lambda}) + \underbrace{(E(\lambda-\hat{\lambda}))^2}_{\text{صفر}} = \lambda$

f) $\Rightarrow \hat{\lambda} = \dfrac{1+۲+۱۵}{۳+۵+۵۰} = \dfrac{۱۹}{۵۸}$

g) $X \sim gamma(\alpha, \beta)$

$P(x) = \dfrac{x^{\alpha-1}e^{-x/\beta}}{(\alpha-1)!\,\beta^{k}}$

question 3)
part A)

ML estimation for N is 50, as we have observed a serial number of 50. the probability of N being smaller than 50 is zero.

Part A1)
As 30 is smaller than 50 our estimation won't change. And again ML estimation of N is 50

Part A2)
Now ML estimation is 60.

Part A3)
Now ML estimation is 70

Part A4)
observing number 5 doesn't effect our estimation because we used to know that N is larger than 70

B)
ML estimation for series of {y1, y2, …, ym} is:

mk = max [y1, y2, …, ym]

C)
We define the CDF of mk as follows:

Pr(mk <= r) = Pr(y1 <= r, y2 <= r , y3<= r, …, ym <= r)
= Prob of choosing m values from sequence of 1 to r / Prob of choosing m values from sequence of 1 to N

C)

Sample : 1, 2, .. , N   seq of observed values: $\{y_1, y_2, \ldots, y_k\}$

ML estimation : $\max\{y_1, y_2, \ldots, y_k\} = m_k$

density function: $F = P(m_k \le r) \rightarrow P(y_1 \le r, y_2 \le r, \ldots, y_k \le r)$

$$F(r) \rightarrow = \begin{cases} \dfrac{\binom{r}{k}}{\binom{n}{k}} & \text{if } k \le r \le n \\ 0 & \text{if } r \le k-1 \\ 1 & \text{if } r \geq n \end{cases}$$

$$f(r) = P(M_k = r) = P(M_k \le r) - P(M_k \le r-1) =$$

$$f(r) = \frac{\binom{r}{k}}{\binom{n}{k}} - \frac{\binom{r-1}{k}}{\binom{n}{k}} = \frac{\binom{r-1}{k-1}}{\binom{n}{k}}$$

bias: $E[M_k] - n \rightarrow$ we are going to calculate $E[M_k]$

$$E[M_k] = \sum_{r=k}^{n} r \times f(r) = \frac{\sum_{r=k}^{n} r \times \binom{r-1}{k-1}}{\binom{k}{n}} =$$

$$= \sum_{r=k}^{n} \frac{r! \times k}{(k-1)!\,(r-k)! \times k} \times \frac{1}{\binom{n}{k}} = \sum_{r=k}^{n} \frac{k \cdot r!}{k! \times (r-k)!} \times \frac{1}{\binom{n}{k}} =$$

$$\sum_{r=k}^{n} \frac{k}{\binom{n}{k}} \times \frac{r!}{k!\,(r-k)!} = \frac{k}{\binom{n}{k}} \times \sum_{r=k}^{n} \binom{r}{k} = \frac{k}{\binom{n}{k}} \times \binom{n+1}{k+1}$$

$$= \frac{\frac{k \cdot (n+1)!}{(k+1)!\,(n-k)!}}{\frac{n!}{k!\,(n-k)!}} = \frac{k\,(n+1)!\,k!}{(k+1)!\,n!} = \frac{k\,(n+1)}{(k+1)}$$

$$\rightarrow E\{M_k\} - n = \frac{k(n+1)}{k+1} - \frac{n(k+1)}{k+1} = \frac{k-n}{k+1} \leftarrow$$ بایاس

بنابراین تخمین ML، بایاس است.

D)

d) $E[\max(x_1, \ldots, x_k)] = \frac{k(N+1)}{k+1}$

این را در قسمت قبل اثبات کردیم:

$M_k = \max(x_1, \ldots, x_k)$

تخمین گر مقابل را معرفی می کنیم:

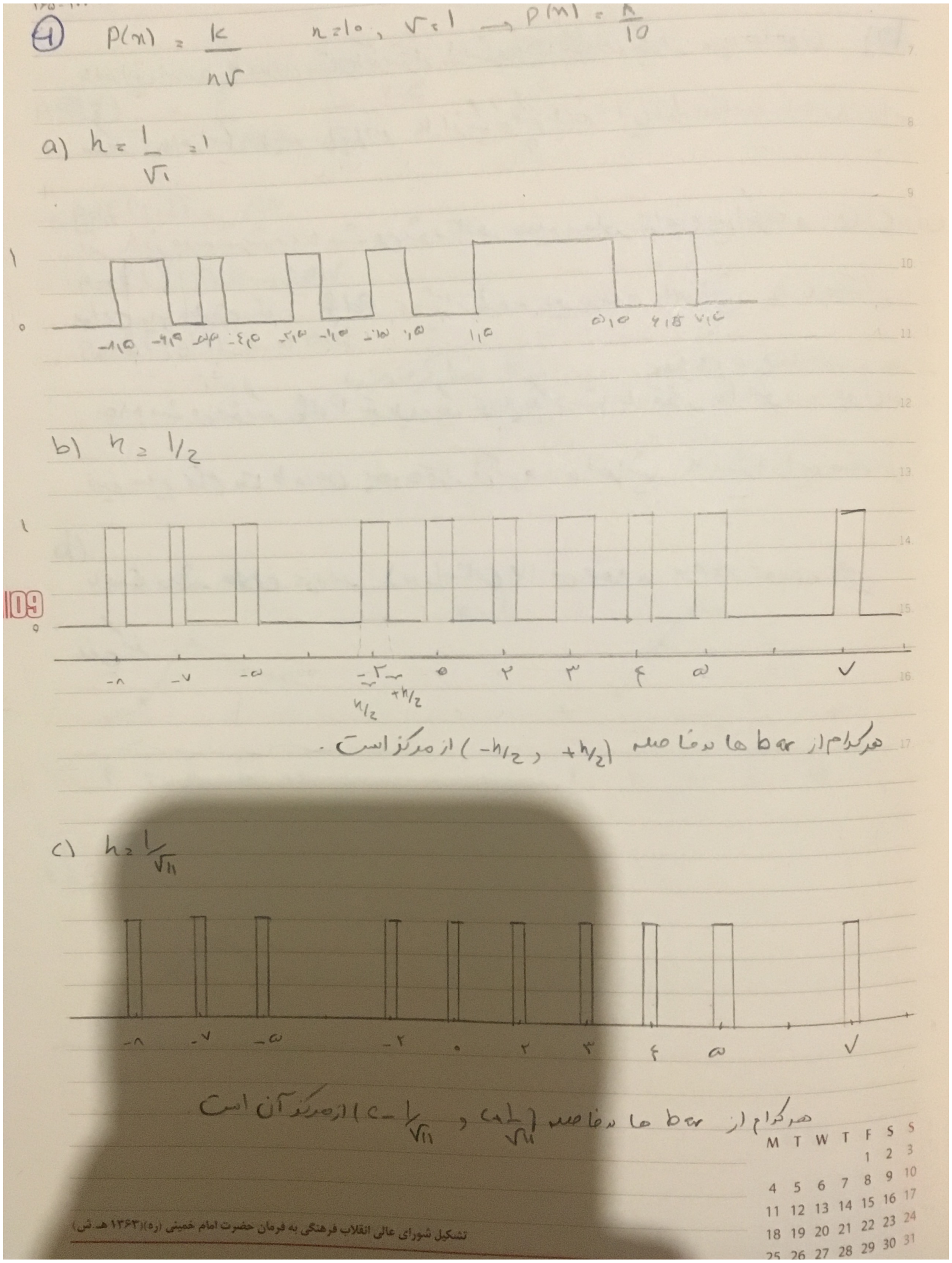
$\hat{N}_k = \frac{k+1}{k} \times M_k - 1$

$\hookrightarrow E[\hat{N}_k] = \frac{k+1}{k} E[M_k] - 1 = \frac{k+1}{k} \times \frac{k}{k+1} \times (N+1) - 1$

$= N+1-1 = N \leftarrow$ این تخمین گر بایاس است.

question 4)

A)



4 p(x) = k / nV   n = 10, V = 1 → p(x) = k / 10
a) h = 1/√1 = 1
b) h = 1/2
هرکدام از bar ها به فاصله (-h/2 , +h/2) از مرکز است.
c) h = 1/√11
هرکدام از bar ها به فاصله (+1/√11 , -1/√11) از مرکز آن است
تشکیل شورای عالی انقلاب فرهنگی به فرمان حضرت امام خمینی (ره)

B)

هفته ۲۹

۰۹۹ - ۲۶۶

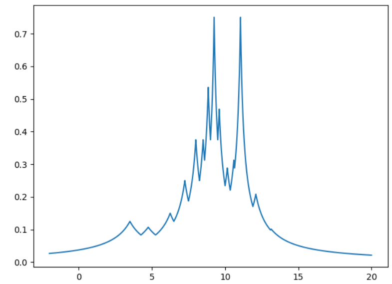
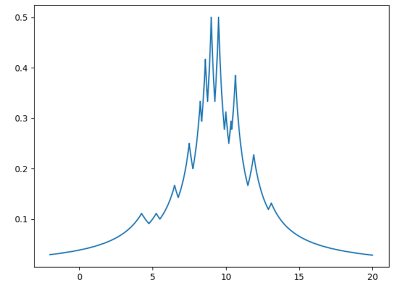
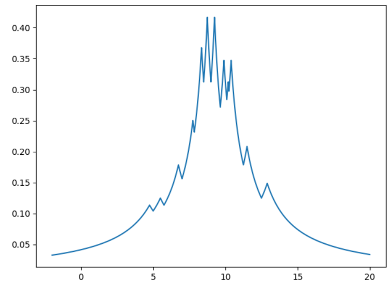
b) با افزایش اندازه h، برای مثال اگر بیشتر از 1 شود، فاصله هر یک از پیک‌ها همون مرکز آن افزایش می‌یابد، اما اگر h از یک کوچکتر باشد، این اندازه فاصله کاهش می‌یابد.

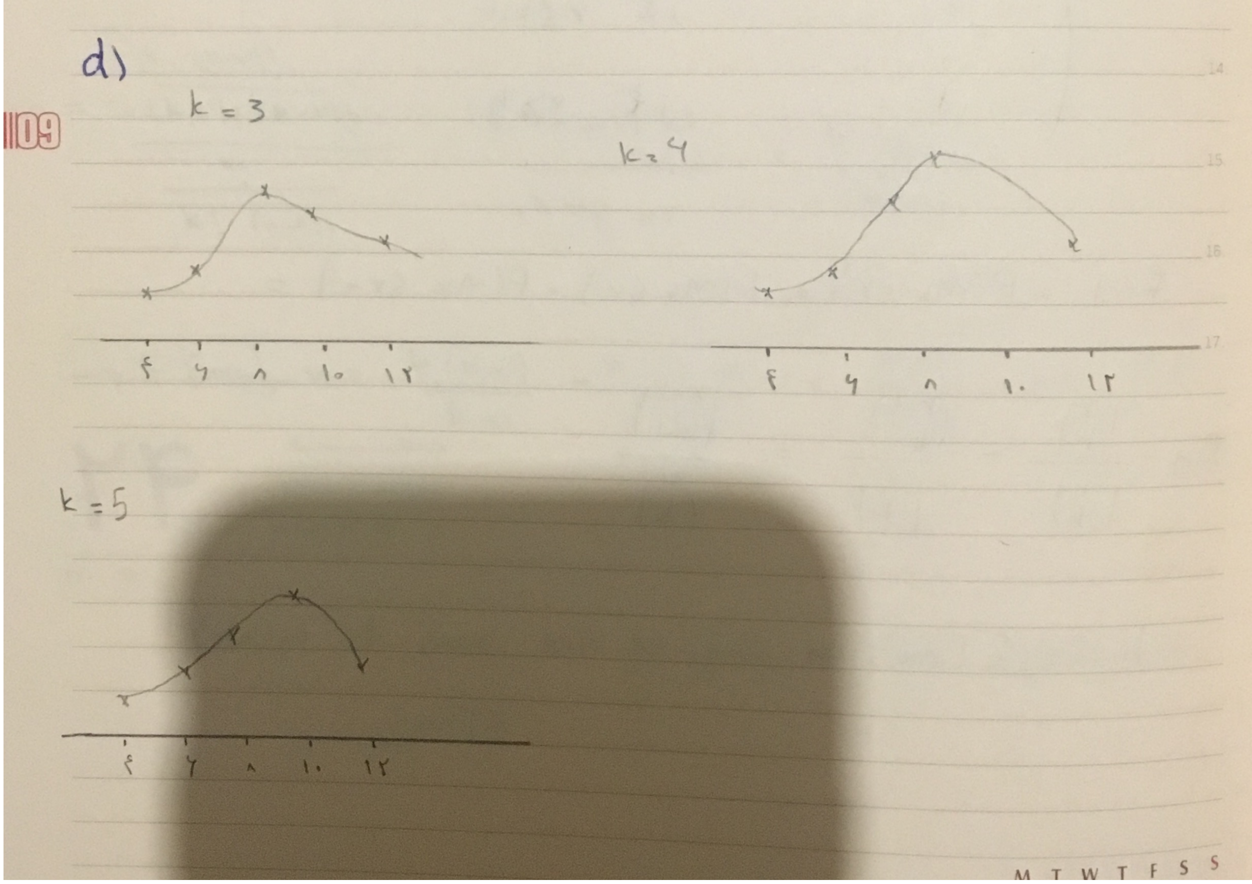
اگر h از حدی بیشتر شود، باعث می‌شود تا دو پیک مجاور تداخل پیدا کنند و امکان دارد آن نواحی یا به عبارتی مقدار pdf، بیشتر از 1 و به دو برسد، افزایش h تا مقدار 15 باعث می‌شود که pdf تبدیل به یک توزیع یکدست با مقدار 0.1 شود، چون که این سایز تمام نقاط داخل پنجره قرار می‌گیرد، و افزایش h بیشتر از این مقدار باعث رشد طولی عرضی (در راستای x) می‌شود و محور عمودی تغییر نمی‌کند.

C)

X	K	V	P(X)
4	3	2	0.09375
4	4	2.5	1
4	5	3.5	0.08928
6	3	1.5	0.125
6	4	2	0.125
6	5	2.5	0.125
8	3	0.5	0.375
8	4	1.0	0.25
8	5	1.1999	0.260416
10	3	0.8	0.2343749
10	4	0.8	0.312499
10	5	1.0	0.3125
12	3	1.0	0.1875
12	4	1.1999	0.2083
12	5	2.0	0.15625

D)

K = 3	K = 4	K = 5
		



E)

۳۶۷-۰۹۸

e)

$$pdf(y) = \sum_{x \in X} \frac{1}{\sqrt{2\pi}} \times e^{-\frac{(|x-y|)^2}{2}}$$

→ pdf(y=۴) = ۰٫۹۵

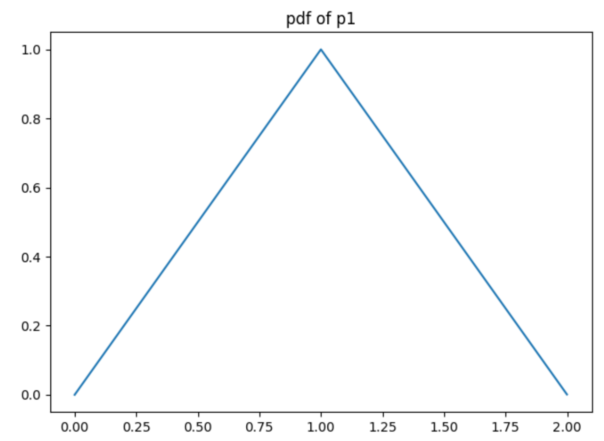
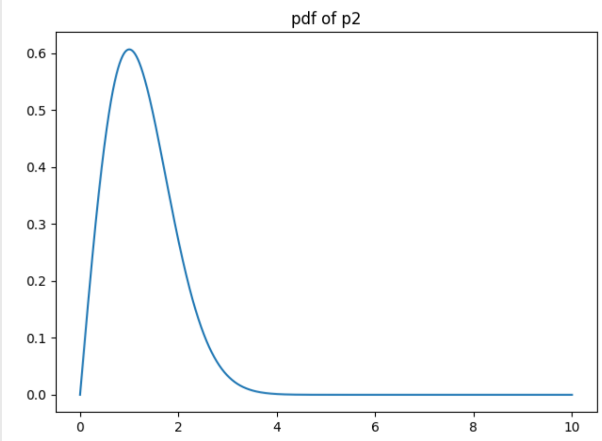
pdf(y=۹) = ۰٫۸۵۱۷

pdf(y=۱۴) = ۰٫۵۱۱۴

* در محاسبات سایز ویندو h = ۱ در نظر گرفته شده

question 5)

A)

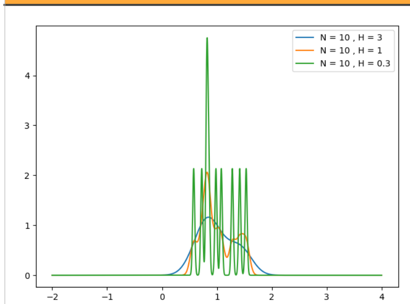
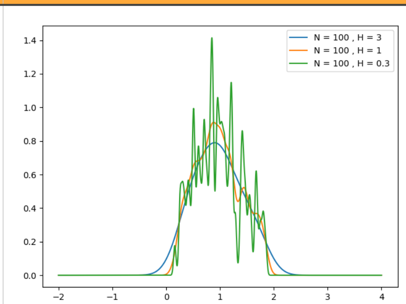
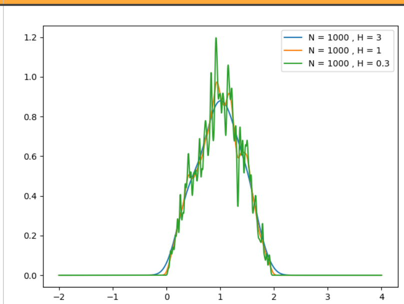
P1	P2
	

B)
samples are generated for following parts!

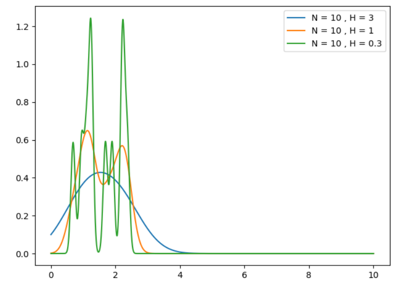
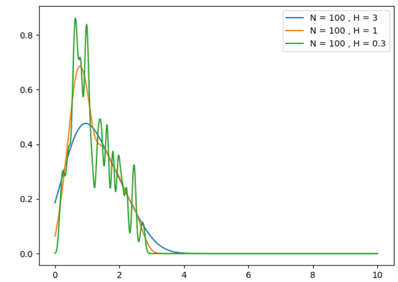
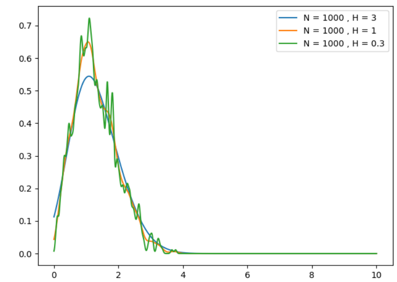
C)

	N = 10	N = 100	N = 1000
P1	0.17084992486258985	0.17818509550840606	0.10789754358648916
P2	0.3791096185216779	0.2711647969041991	0.17591413360373356

D)
for P1:

10	100	1000
		

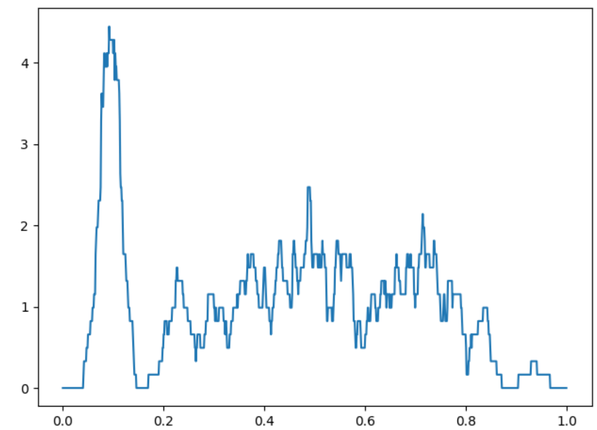
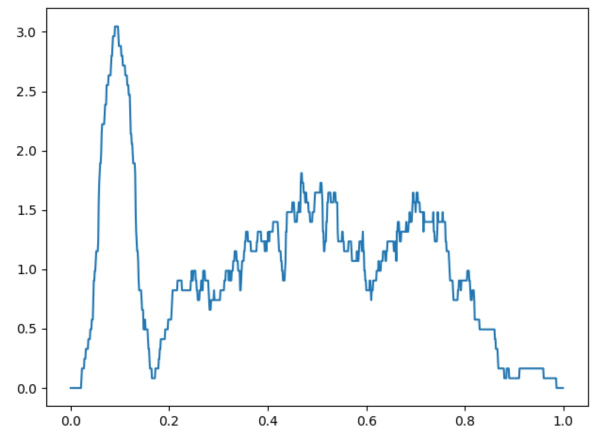
for P2:

10	100	1000
		

As wee see in results increase in number of samples makes the predicted density more similar to the original density function.
And larger values for bandwidth makes the estimated density smoother than smaller bandwidth.

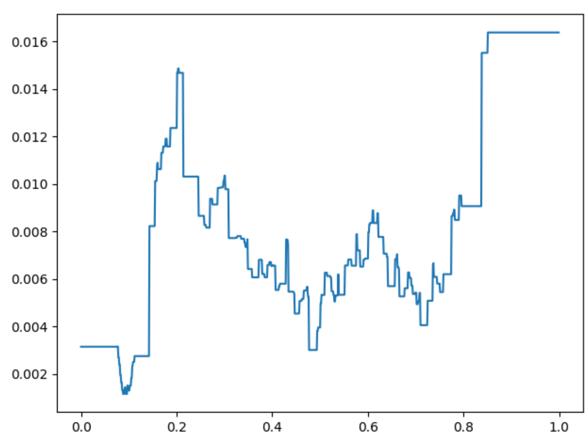
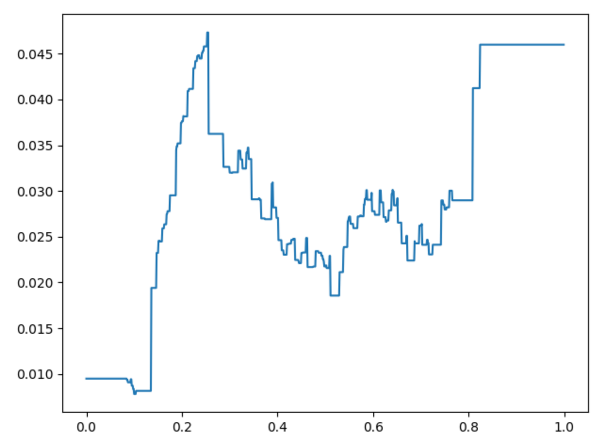
question 6)

A)

h1 = 0.0375	**h = 0.075**
	

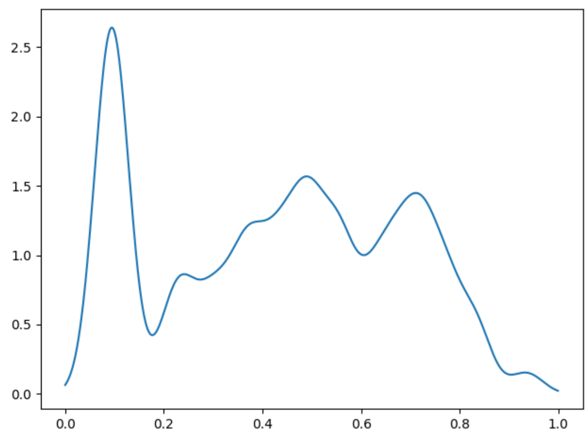
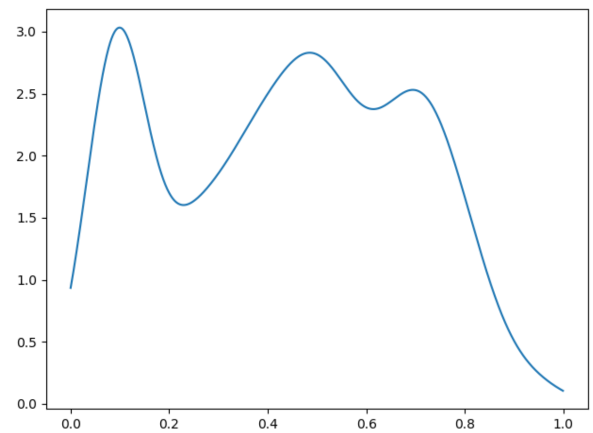
As it can be seen in results larger value for window size results in smoother estimation than smaller window sizes. while smaller window size results in more variance estimation.

B)

K = 15	**K = 30**
	

According to the results larger value of K results in smoother and more biased estimation.

C)

sigma = 0.03	sigma = 0.06
	

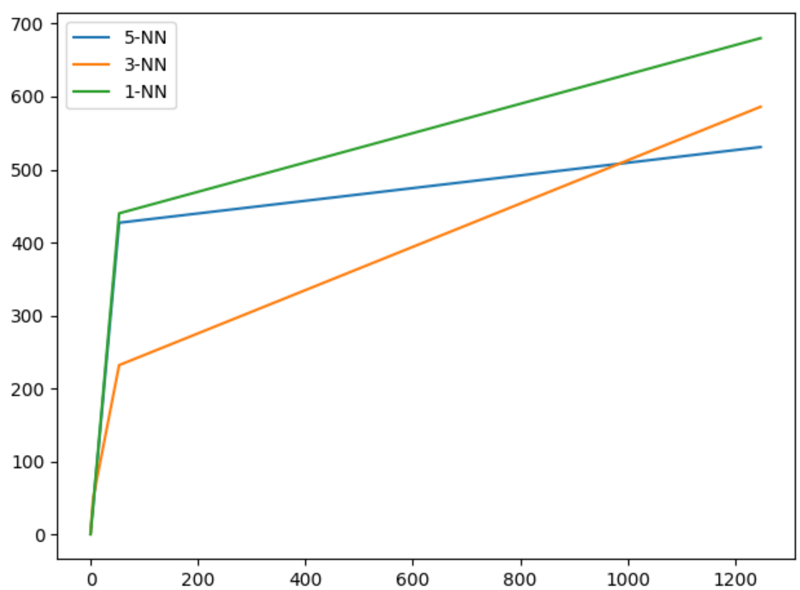
According to the results larger value for sigma results in smoother and more biased estimation.

question 7)

A)

K NN	prediction
5 NN	[0.45600000000000007, 8.02, 37.12, 427.1, 530.8]
3 NN	[0.3433333333333333, 8.1, 49.199999999999996, 232.0, 586.0]
1 NN	[0.33, 2.4, 39.2, 440.0, 680.0]

B)

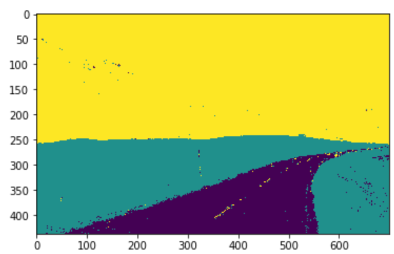
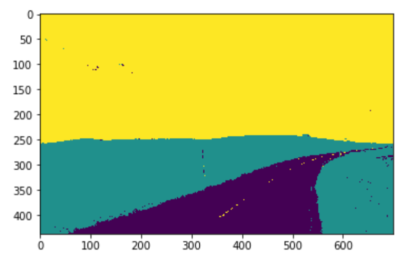
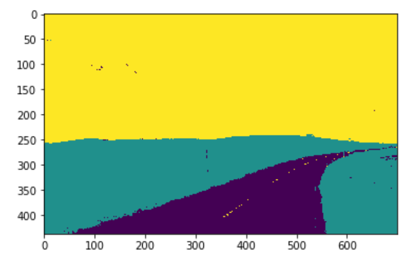


C)

K NN	prediction
5 NN	[6.0, 5.0, 6.0, 5.0, 6.0]
3 NN	[6.0, 5.0, 6.0, 6.0, 6.0]
1 NN	[6.0, 5.0, 6.0, 6.0, 6.0]

question 8)

A)

1-NN	5-NN	9-NN
		

B) Knn can be used for lane detection if lane is uniformly shown in picture, like the above test case. That means there is no other object inside the surface of road and neighbor of any point in the lane is an other point from lane with a similar color. But if there were a car or an other object in the image, or the road didn’t has a uniform color, It could mislead KNN in predicting the lane.

question 9)

A) In this case numerical methods or Sub-gradient methodology can be used for finding the optimal value of log likelihood function.

B) The only difference between MAP and MLE is that: MAP estimator considers prior probability whereas MLE doesn't. When prior belief is weak it is better to use MLE.

C) Here are some reasons:

1) When likelihood function is product of set of probability functions, using log functions makes calculations easier. Because it converts products to sum.
2) In the case of gaussian distribution computational cost of exponential function is omitted by using log function.
3) Ln(X) function is a monotonically increasing function, thus log-likelihoods have the same relations of order as the likelihoods. maximizing likelihood is equivalent to maximizing log likelihood.
4) summing is less expensive than multiplication and likelihoods would become very small and you will run out of your floating point precision very quickly, yielding an underflow.

D) KNN is a lazy algorithm that means it has no training phase and just gathering the train dataset is done in this phase and all of the computations are done in test time.

E) A good choice depends on the data itself, the nature of the distribution, and the sample size. If a value is chosen that is too small, this can overemphasize the data variability. Therefore in most cases the more structure a graph has the smaller the value and the flatter the graph the larger the value. The best choice of the smoothing parameter hinges on the sample size and the following three factors as described in Hart (1997): The smoothness of the density function; the distribution of the design points; and the amount of variability among the data.
different approaches for selection of H are proposed in this article:
https://pdfs.semanticscholar.org/6836/334793939c6d29890aba1b33216f58f69734.pdf

F) by large value of K, KNN is a biased estimator and by smaller value of K, KNN is a high variance estimator. For large value of K volume in each point is a large number for including the K samples. this leads to an approximately identical Volume size for most of the points which results in a biased estimation. For smaller values of K the volume V for each point would be a non-steady number which leads to a high variance estimation.

G)
test data: (1,1)

train data:
A = (0, 1)
B = (3, 1)

cosine distance of test point with:
A -> 1- (1/ radical(2)) = 1- 0.70 = 0.3
B -> 1 - (4/ radical (20)) = 1 - 0.89 = 0.11

manhattan distance:
A -> 1
B -> 2

As we see different distance measures in the above example for a 1-NN classifier leads to different predictions for the test case.