CS215 Assignment 3

Aadish Jain - 190050001
Deepanshu - 190050032
Gaurang Dev - 19D070024

September 27, 2020

Contents

1 Q1 **1**

2 Q2 **6**

3 Q3 **8**

4 Q4 **12**

1 Q1

a) Since the first book will always be of some colour, we can say

$$\boxed{X_1 = 1}$$

When books with $i-1$ distinct types of colors have been collected, the probability of picking a book with a different color will be number of different colours possible for ith pick divided by the total number of books which can be picked. Therefore,

$$\boxed{P(X_i = 1) = \frac{n-i+1}{n}}$$

b) To find $P(X_i = k)$, $X_i = k$ means for $k-1$ picks, no new book was picked and in the k^{th} pick, new book was picked. Since the events $X_i = k$ for $k = 1$ to n are independent, we may write :

$$P(X_i = k) = P(X_i \neq 1)^{k-1} P(X_i = 1)$$

$$\implies P(X_i = k) = \left(1 - \frac{n-i+1}{n}\right)^{k-1} \frac{n-i+1}{n}$$

Comparing it with the formula for geometric random variable, the parameter p for $P(X_i = k)$ is

$$\boxed{p = \frac{n-i+1}{n}}$$

c) To find $E(Z)$, where Z is a geometric random variable with parameter p, we may write

$$E(Z) = \sum_{k=1}^{\infty} kP(Z = k)$$

$$\implies E(Z) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p$$

$$\implies E(Z) = p + 2p(1-p) + 3p(1-p)^2 ... \tag{1}$$

Multiply the above equation by $(1-p)$ both the sides, we get

$$\implies (1-p)E(Z) = 0 + p(1-p) + 2p(1-p)^2 + 3p(1-p)^3 ... \tag{2}$$

Subtracting equation 2 from 1, i.e. $(1) - (2)$, we get

$$\implies (1-(1-p))E(Z) = (p-0) + (2p(1-p) - p(1-p)) + (3p(1-p)^2 - 2p(1-p)^2) ...$$

$$\implies pE(Z) = p + p(1-p) + p(1-p)^2 + p(1-p)^3...$$

Cancelling p from both the sides, we get

$$\implies E(Z) = 1 + (1-p) + (1-p)^2 + (1-p)^3...$$

Therefore, the RHS becomes an infinite G.P. with first term 1 and common ratio $(1-p)$, we can write the sum of G.P. as first term divided by 1 minus common ratio. Hence,

$$E(Z) = \frac{1}{(1-(1-p))}$$

$$\implies \boxed{E(Z) = \frac{1}{p}}$$

To find $Var(Z)$, with geometric random variable Z, we can write :

$$Var(Z) = E(Z^2) - (E(Z))^2$$

We have already calculated $E(Z) = \frac{1}{p}$, now we need to calculate $E(Z^2)$. For this, we write :

$$E(Z^2) = \sum_{k=1}^{\infty} k^2 P(Z = k)$$

$$\implies E(Z^2) = \sum_{k=1}^{\infty} k^2 (1-p)^{k-1} p$$

$$\implies E(Z^2) = p + 4p(1-p) + 9p(1-p)^2 + 16p(1-p)^3... \tag{3}$$

Multiply the above equation by $(1-p)$ both the sides, we get

$$\implies (1-p)E(Z^2) = 0 + p(1-p) + 4p(1-p)^2 + 9p(1-p)^3 + 16p(1-p)^4... \tag{4}$$

Subtracting equation 4 from 3, i.e. $(3) - (4)$, we get

$$pE(Z^2) = (p-0) + 3p(1-p) + 5p(1-p)^2 + 7p(1-p)^3... \tag{5}$$

Again multiplying the above equation by $(1-p)$ both the sides, we get

$$\implies p(1-p)E(Z^2) = 0 + p(1-p) + 3p(1-p)^2 + 5p(1-p)^3... \tag{6}$$

Subtracting equation 6 from 5, i.e. $(5) - (6)$, we get

$$\implies (1-(1-p))pE(Z^2) = (p-0) + 2p(1-p) + 2p(1-p)^2 + 2p(1-p)^3...$$

$$\implies p^2E(Z^2) = p + 2p(1-p) + 2p(1-p)^2 + 2p(1-p)^3...$$

Cancelling p from both the sides, we get

$$pE(Z^2) = 1 + 2(1-p) + 2(1-p)^2 + 2(1-p)^3...$$

$$\implies pE(Z^2) - 1 = 2(1-p) + 2(1-p)^2 + 2(1-p)^3...$$

Therefore, the RHS becomes an infinite G.P. with first term $2(1-p)$ and common ratio $(1-p)$, we can write the sum of G.P. as first term divided by 1 minus common ratio. Hence,

$$pE(Z^2) - 1 = \frac{2(1-p)}{(1-(1-p))}$$

$$\implies pE(Z^2) = \frac{2-p}{p}$$

$$\implies E(Z^2) = \frac{2-p}{p^2}$$

Therefore, using formula of variance :

$$Var(Z) = E(Z^2) - (E(Z))^2$$

$$\implies Var(Z) = \frac{2-p}{p^2} - \left(\frac{1}{p}\right)^2$$

$$\implies \boxed{Var(Z) = \frac{1-p}{p^2}}$$

d) Since $X^{(n)}$ is sum of geometric random variables, we may write

$$E(X^{(n)}) = E(X_1 + X_2 + X_3 + ...X_n)$$

$$E(X^{(n)}) = E(X_1) + E(X_2) + E(X_3) + ...E(X_n)$$

Since, X_i for $i = 1$ to n is a geometric random variable with parameter $p_i = \dfrac{n-i+1}{n}$ and expectation value $1/p_i$, we can write

$$E(X^{(n)}) = \sum_{i=1}^{n} \frac{1}{p_i}$$

$$\implies \boxed{E(X^{(n)}) = \sum_{i=1}^{n} \frac{n}{n-i+1}}$$

e) To calculate $Var(X^{(n)})$, since $X^{(n)}$ is sum of n independent random variables, we may write

$$Var(X^{(n)}) = Var(X_1) + Var(X_2)...Var(X_n)$$

$$\implies Var(X^{(n)}) = \sum_{i=1}^{n} Var(X_i)$$

Since, X_i for $i = 1$ to n is a geometric random variable with parameter $p_i = \frac{n-i+1}{n}$ and variance $\frac{1-p_i}{p_i^2}$, we can write

$$Var(X^{(n)}) = \sum_{i=1}^{n} \frac{1-p_i}{p_i^2}$$

Since p_i is always greater than 0, we may write the inequality

$$Var(X^{(n)}) \leq \sum_{i=1}^{n} \frac{1}{p_i^2}$$

Substituting value of p_i, we get

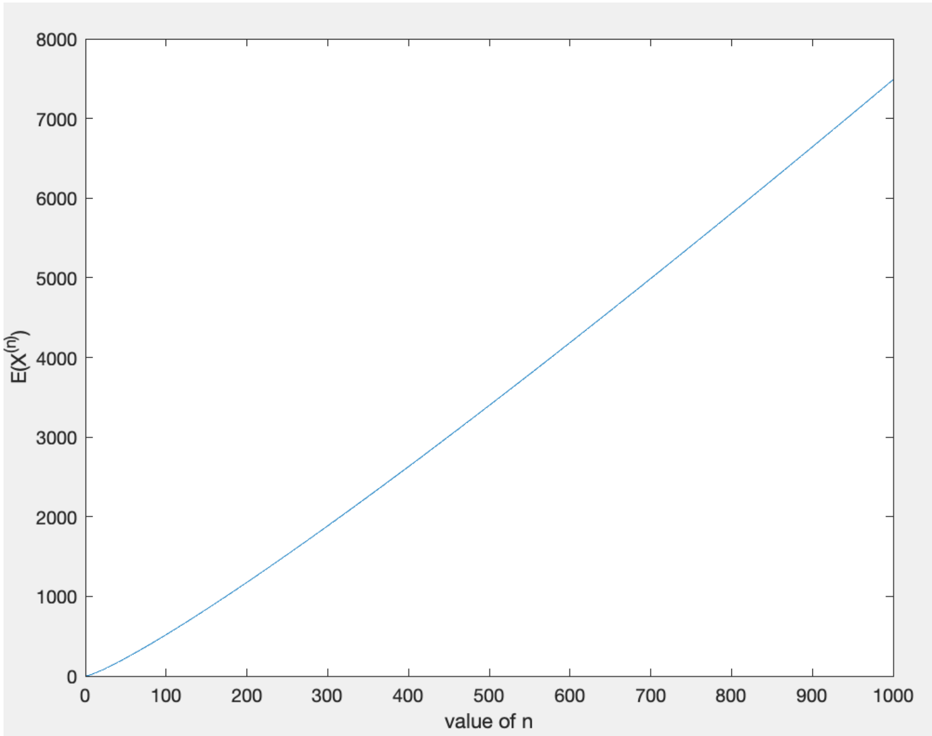
$$Var(X^{(n)}) \leq \sum_{i=1}^{n} \frac{n^2}{(n-i+1)^2}$$

$$\implies Var(X^{(n)}) \leq n^2 \sum_{i=1}^{n} \frac{1}{(n-i+1)^2}$$

$$\implies Var(X^{(n)}) \leq n^2 \left(\frac{1}{n^2} + \frac{1}{(n-1)^2} + \frac{1}{(n-2)^2} + ... + \frac{1}{1^2} \right)$$

Given that sum of reciprocals of squares of positive integers is upper bounded by $\frac{\pi^2}{6}$, and substituting it in the above inequality, it becomes

$$\implies \boxed{Var(X^{(n)}) < \frac{n^2\pi^2}{6}}$$



f) We know that $E(X^{(n)}) = \sum_{i=1}^{n} \dfrac{n}{n-i+1} = \sum_{i=1}^{n} \dfrac{n}{i}$

$$E(X^{(n)}) = \frac{n}{1}+\frac{n}{2}+\frac{n}{3}+\frac{n}{4}+\frac{n}{5}+\frac{n}{6}+\frac{n}{7}... < \frac{n}{1}+\frac{n}{2}+\frac{n}{2}+\frac{n}{4}+\frac{n}{4}+\frac{n}{4}+\frac{n}{4}... < nlog(n)$$

$$E(X^{(n)}) = \frac{n}{1}+\frac{n}{2}+\frac{n}{3}+\frac{n}{4}+\frac{n}{5}+\frac{n}{6}+\frac{n}{7}... > \frac{n}{1}+\frac{n}{2}+\frac{n}{4}+\frac{n}{4}+\frac{n}{8}+\frac{n}{8}+\frac{n}{8}... > n+\frac{nlog(n)}{2}$$

Using the above two inequalities, we can say that

$$E(X^n) = \Theta(nlog(n))$$

$$\implies \boxed{f(n) = nlog(n)}$$

2 Q2

(a)

Since F is a distribution function, it is an **increasing** function.
Since F is also inverible. It must be **strictly increasing**.
Let u be a random variable which can take values from [0,1] uniform distribution.
Let the values $\{v_i = F^{-1}(u_i)\}_{i=1}^{n}$ be denoted as v;

$$v = F^{-1}(u)$$

which equals:

$$F(v) = u$$

Let distribution function of v be $F_v(x)$.
Therefore:

$$F_v(x) = P(v \leq x)$$

As F is a strictly function, we can take F on both sides of inequality and the inequality will not change.
Hence:

$$F_v(x) = P(v \leq x) = P(F(v) \leq F(x)) = P(u \leq F(x))$$

Since u belongs to uniform distribution:

$$P(u \leq F(x)) = F(x)$$

which gives:

$$\boxed{F_v(x) = F(x)}$$

Hence we can say that the values $\{v_i\}_{i=1}^{n}$ follow the distribution F.

(b)

Let U be a random variable which can take values from [0,1] uniform distribution.
And let Y be a random variable which can take values from a continuous distribution F.
To show Distribution of E is same as Distribution of D.
Since F is a continuous function and monotonically increasing it will take all values from [0,1];
Now,Since y can take any value from [0,1], and F(x) can also take any value from [0,1], y can be safely substituted by F(x) in the expression for E.
Therefore:

$$E = max_{0 \leq F(x) \leq 1} \left| \frac{\sum_{i=1}^{n} 1(U_i \leq F(x))}{n} - F(x) \right|$$

Since F(x) depends only on x and if $F(x) \in [0,1]$ then $x \in R$ Therefore:

$$E = max_x \left| \frac{\sum_{i=1}^{n} 1(U_i \le F(x))}{n} - F(x) \right|$$

Assuming F is an invertible function, Since it is a distribution function it must be strictly increasing and F^{-1} will also be strictly increasing.
Therefore taking F^{-1} on both sides of inequality will not change the inequality and hence:

$$E = max_x \left| \frac{\sum_{i=1}^{n} 1(F^{-1}(U_i) \le x)}{n} - F(x) \right|$$

From part (a) we can say that if U belongs to [0,1] Uniform Distribution. $F^{-1}(U)$ follows the same distribution F.
Since Y is also a random variable which belongs to the same distribution F
Therefore we can say that: E and D belongs to the same distribution. which we can write as:

$$F_E(x) = F_D(x)$$

And therefore:

$$\boxed{P(E \ge d) = 1 - P(E < d) = 1 - F_E(d) = 1 - F_D(d) = 1 - P(D < d) = P(D \ge d) \quad \square}$$

Practical Significance:

For any given arbitrary Continuous Distribution F, Suppose we need to find the number of Data points to be taken so that the empirical distribution F_e lies within d(whatever we want) radius of True distribution F,

For this, we can safely replace the given distribution by [0,1] Uniform distribution as

$$P(E \le d) = P(D \le d)$$

.The number of sample points required for the [0,1] uniform distribution will be equal to number of data points required for the original distribution.

3 Q3

Part (a):

X, Y $\rightarrow$ exactly known
Z $\rightarrow$ know but corrupted

$$z_i = ax_i + by_i + c + \epsilon_i$$

where, $\epsilon \sim N(0, \sigma^2)$

$$P(z; a, b, c, x, y) = \frac{e^{\frac{-(z-(ax+by+c))^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$P(z_{i=1}^n) = \Pi_{i=1}^n \frac{e^{\frac{-(z_i-(ax_i+by_i+c))^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$\mathcal{L} = \log(P(z_{i=1}^n)) = \sum_{i=1}^n \frac{-(z_i-(ax_i+by_i+c))^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})$$

So now to maximize,

$$\frac{\partial \mathcal{L}}{\partial a} = 0$$
$$\frac{\partial \mathcal{L}}{\partial b} = 0$$
$$\frac{\partial \mathcal{L}}{\partial c} = 0$$

with respect to a,

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial a} &= \frac{-2}{2\sigma^2}\sum_{i=1}^n (z_i-(ax_i+by_i+c))(-x_i) = 0 \\ &= \sum (x_iz_i - ax_i^2 - bx_iy_i - cx_i) = 0 \\ &= \sum x_iz_i - a\sum x_i^2 - b\sum x_iy_i - cn\bar{x} = 0\end{aligned} \tag{7}$$

similarly with respect to b and c,

$$\frac{\partial \mathcal{L}}{\partial b} = \sum y_iz_i - a\sum x_iy_i - b\sum y_i^2 - cn\bar{y} = 0 \tag{8}$$

$$\frac{\partial \mathcal{L}}{\partial c} = n\bar{z} - an\bar{x} - bn\bar{y} - nc = 0 \tag{9}$$

from above 3 equation we can make the corresponding matrix and vector form,

$$\begin{bmatrix} \sum x_i^2 & \sum x_i y_i & n\bar{x} \\ \sum x_i y_i & \sum y_i^2 & n\bar{y} \\ n\bar{x} & n\bar{y} & n \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ n\bar{z} \end{bmatrix}$$

Part (b):

X, Y $\rightarrow$ exactly known
Z $\rightarrow$ know but corrupted

$$z_i = a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6 + \epsilon_i$$

where, $\epsilon \sim N(0, \sigma^2)$

$$P(z; a_1, a_2, a_3, a_4, a_5, a_6, x, y) = \frac{e^{\frac{-(z-(a_1 x^2 + a_2 y^2 + a_3 xy + a_4 x + a_5 y + a_6))^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$P(z_{i=1}^n) = \Pi_{i=1}^n \frac{e^{\frac{-(z_i-(a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6))^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$\mathcal{L} = \log(P(z_{i=1}^n)) = \sum_{i=1}^{n} \frac{-(z_i - (a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6))^2}{2\sigma^2} - \log(\sigma\sqrt{2\pi})$$

now to maximize,

$$\frac{\partial \mathcal{L}}{\partial a} = 0$$
$$\frac{\partial \mathcal{L}}{\partial b} = 0$$
$$\frac{\partial \mathcal{L}}{\partial c} = 0$$

with respect to a_1,

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial a_1} &= \frac{-2}{2\sigma^2} \sum_{i=1}^{n} (z_i - (a_1 x^2 + a_2 y^2 + a_3 xy + a_4 x + a_5 y + a_6))(-x_i^2) = 0 \\ &= \sum (z_i x_i^2 - a_1 x_i^4 - a_2 x_i^2 y_i^2 - a_3 x_i^3 y_i - a_4 x_i^3 - a_5 x_i^2 y_i - a_6 x_i^2) = 0 \\ &= \sum z_i x_i^2 - a_1 \sum x_i^4 - a_2 \sum x_i^2 y_i^2 - a_3 \sum x_i^3 y_i - a_4 \sum x_i^3 - a_5 \sum x_i^2 y_i - a_6 \sum x_i^2 = 0 \end{aligned} \tag{10}$$

similarly with respect to a_2,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_2} &= \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(z_i - (a_1x^2 + a_2y^2 + a_3xy + a_4x + a_5y + a_6))(-y_i^2) = 0 \\
&= \sum(z_iy_i^2 - a_1x_i^2y_i^2 - a_2y_i^4 - a_3x_iy_i^3 - a_4x_iy_i^2 - a_5y_i^2 - a_6y_i^2) = 0 \\
&= \sum z_iy_i^2 - a_1\sum x_i^2y_i^2 - a_2\sum y_i^4 - a_3\sum x_iy_i^3 - a_4\sum x_iy_i^2 - a_5\sum y_i^3 - a_6\sum y_i^2 = 0
\end{aligned} \tag{11}$$

with respect to a_3,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_3} &= \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(z_i - (a_1x^2 + a_2y^2 + a_3xy + a_4x + a_5y + a_6))(-x_iy_i) = 0 \\
&= \sum(z_ix_iy_i - a_1x_i^3y_i - a_2x_iy_i^3 - a_3x_i^2y_i^2 - a_4x_i^2y_i - a_5x_iy_i^2 - a_6x_iy_i) = 0 \\
&= \sum z_ix_iy_i - a_1\sum x_i^3y_i - a_2\sum x_iy_i^3 - a_3\sum x_i^2y_i^2 - a_4\sum x_i^2y_i - a_5\sum x_iy_i^2 - a_6\sum x_iy_i = 0
\end{aligned} \tag{12}$$

with respect to a_4,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_3} &= \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(z_i - (a_1x^2 + a_2y^2 + a_3xy + a_4x + a_5y + a_6))(-x_i) = 0 \\
&= \sum(z_ix_i - a_1x_i^3 - a_2x_iy_i^2 - a_3x_i^2y_i - a_4x_i^2 - a_5x_iy_i - a_6x_i) = 0 \\
&= \sum z_ix_i - a_1\sum x_i^3 - a_2\sum x_iy_i^2 - a_3\sum x_i^2y_i - a_4\sum x_i^2 - a_5\sum x_iy_i - a_6n\bar{x} = 0
\end{aligned} \tag{13}$$

with respect to a_5,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_3} &= \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(z_i - (a_1x^2 + a_2y^2 + a_3xy + a_4x + a_5y + a_6))(-y_i) = 0 \\
&= \sum(z_iy_i - a_1x_i^2y_i - a_2y_i^3 - a_3x_iy_i^2 - a_4x_iy_i - a_5y_i^2 - a_6y_i) = 0 \\
&= \sum z_iy_i - a_1\sum x_i^2y_i - a_2\sum y_i^3 - a_3\sum x_iy_i^2 - a_4\sum x_iy_i - a_5\sum y_i^2 - a_6n\bar{y} = 0
\end{aligned} \tag{14}$$

with respect to a_6,

$$\begin{aligned}
\frac{\partial \mathcal{L}}{\partial a_3} &= \frac{-2}{2\sigma^2}\sum_{i=1}^{n}(z_i - (a_1x^2 + a_2y^2 + a_3xy + a_4x + a_5y + a_6))(-1) = 0 \\
&= \sum(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \\
&= n\bar{z} - a_1\sum x_i^2 - a_2\sum y_i^2 - a_3\sum x_iy_i - a_4n\bar{x} - a_5n\bar{y} - a_6n = 0
\end{aligned} \tag{15}$$

from the above 6 equations, we can form the matrix and vector form,

$$\begin{bmatrix} \sum x_i^4 & \sum x_i^2 y_i^2 & \sum x_i^3 y_i & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i^2 \\ \sum x_i^2 y_i^2 & \sum y_i^4 & \sum x_i y_i^3 & \sum x_i y_i^2 & \sum y_i^3 & \sum y_i^2 \\ \sum x_i^3 y_i & \sum x_i y_i^3 & \sum x_i^2 y_i^2 & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i y_i \\ \sum x_i^3 & \sum x_i y_i^2 & \sum x_i^2 y_i & \sum x_i^2 & \sum x_i y_i & n\bar{x} \\ \sum x_i^2 y_i & \sum y_i^3 & \sum x_i y_i^2 & \sum x_i y_i & \sum y_i^2 & n\bar{y} \\ \sum x_i^2 & \sum y_i^2 & \sum x_i y_i & n\bar{x} & n\bar{y} & n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} \sum z_i x_i^2 \\ \sum z_i y_i^2 \\ \sum z_i x_i y_i \\ \sum z_i x_i \\ \sum z_i y_i \\ n\bar{z} \end{bmatrix}$$

Part (c):

for part (a)
a = 10.0022
b = 19.9980
c = 29.9516
So, Equation of Plane
z = 10.0022x + 19.998y + 29.9516
Variance of noise = 23.0685

for part(b)
a1 = 57191e-05
a2 = -5,1551e-05
a3 = -2.1305e-05
a4 = 9.9975
a5 = 20.0042
a6 = 29.9133
So, equation of Plane
z = 57191e-05x^2 -5,1551e-05y^2 -2.1305e-05xy + 9.9975x + 20.0042y +29.9133
Variance of noise = 23.0649

4 Q4

b) To find joint likelihood of the samples in V(with size k), based on the estimate of the PDF built from T(with size n) with bandwidth parameter σ, we can write($v_j for j = 1$ to k are the elements of the vector V)

$$JL = \prod_{j=1}^{k} \hat{p}_n(v_j; \sigma)$$

Using Kernel Density Estimation,($t_i for j = 1$ to n are the elements of the vector T)

$$\hat{p}_n(v_j; \sigma) = \frac{\sum_{i=1}^{n} exp^{(-(v_j - t_i)^2/(2\sigma^2))}}{n\sigma\sqrt{2\pi}}$$

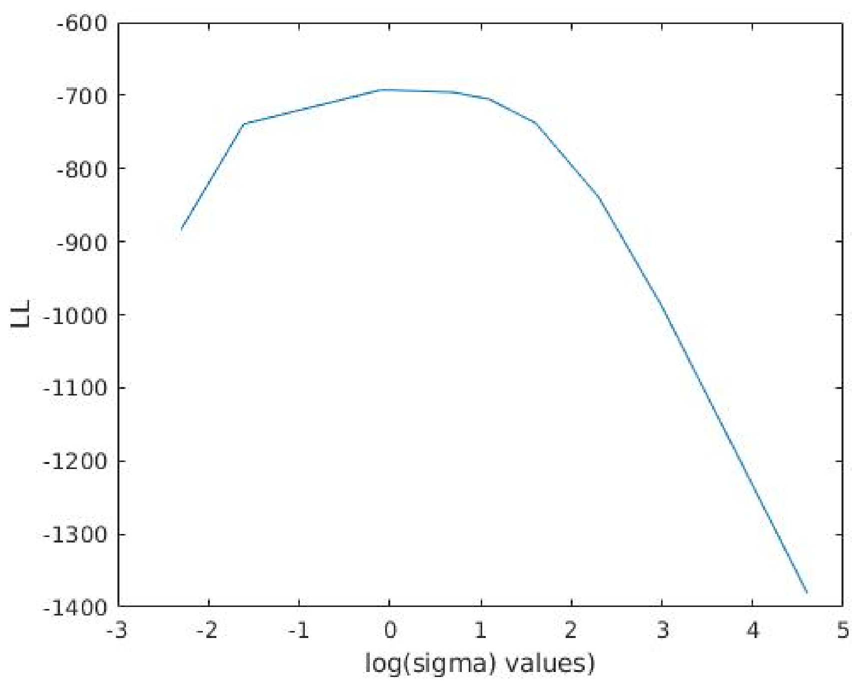
On substituting, we get

$$\boxed{JL = \prod_{j=1}^{k} \frac{\sum_{i=1}^{n} exp^{(-(v_j - t_i)^2/(2\sigma^2))}}{n\sigma\sqrt{2\pi}}}$$

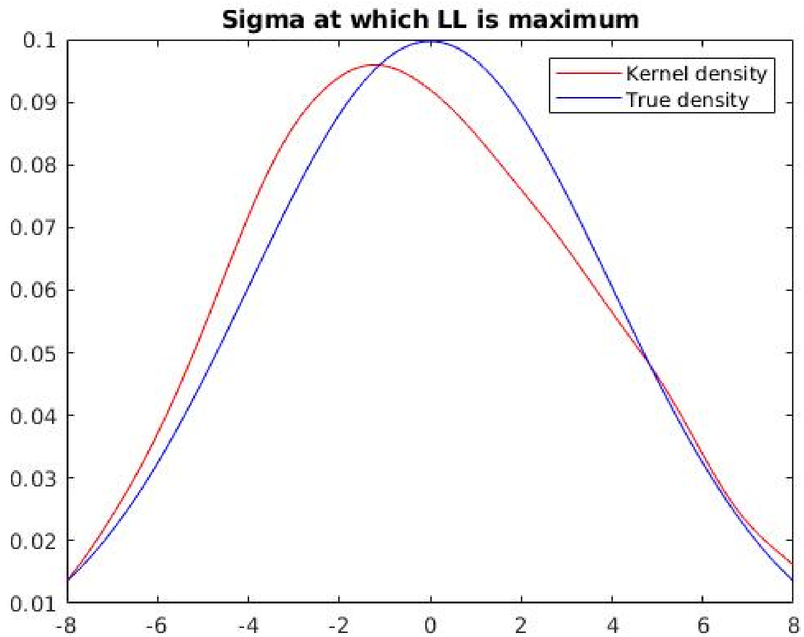
c) The value of sigma at which the maximum value of joint likelihood was attained came out to be

$$\boxed{\hat{\sigma} = 1}$$

Plot of LL vs. $log(\sigma)$:



Plot of $\hat{p}_n(x;\sigma)$ vs. x :



d)

The value of sigma at which the minimum value of D was attained came out to be

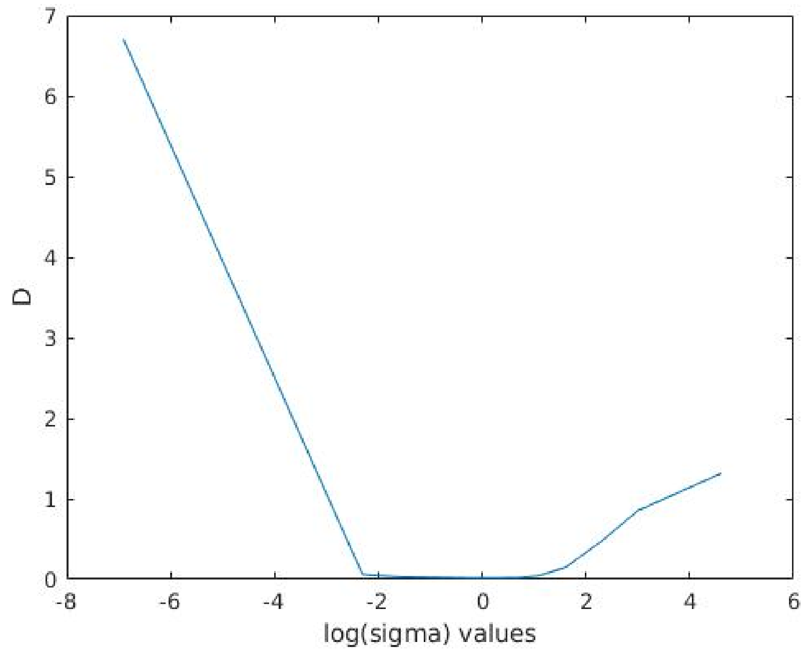
$$\boxed{\hat{\sigma} = 1}$$

The D value for the parameter which yielded the best LL came out to be

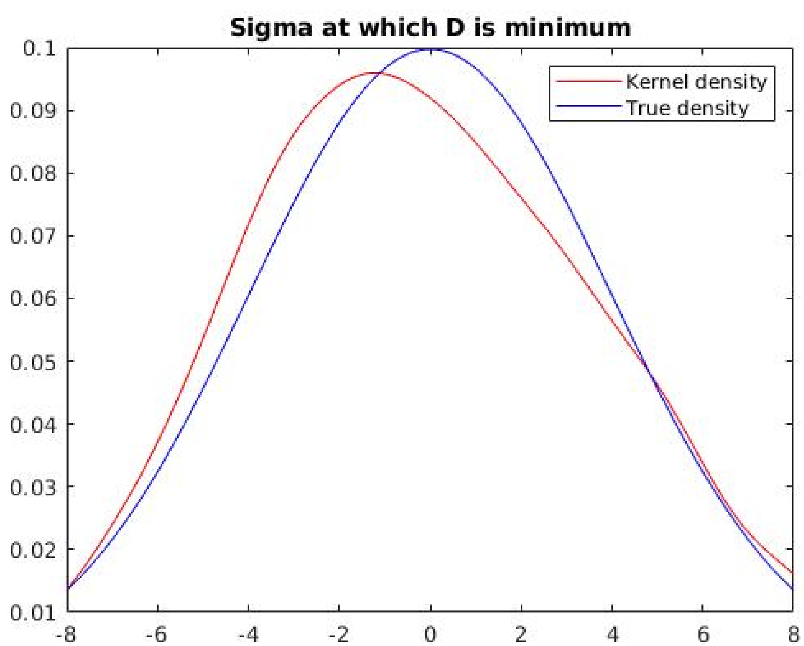
$$\boxed{D = 0.016}$$

Here we will take the minima of D as the best case, since we need to minimise D so that the Kernel density esitmate is closest possible to the true estimate.

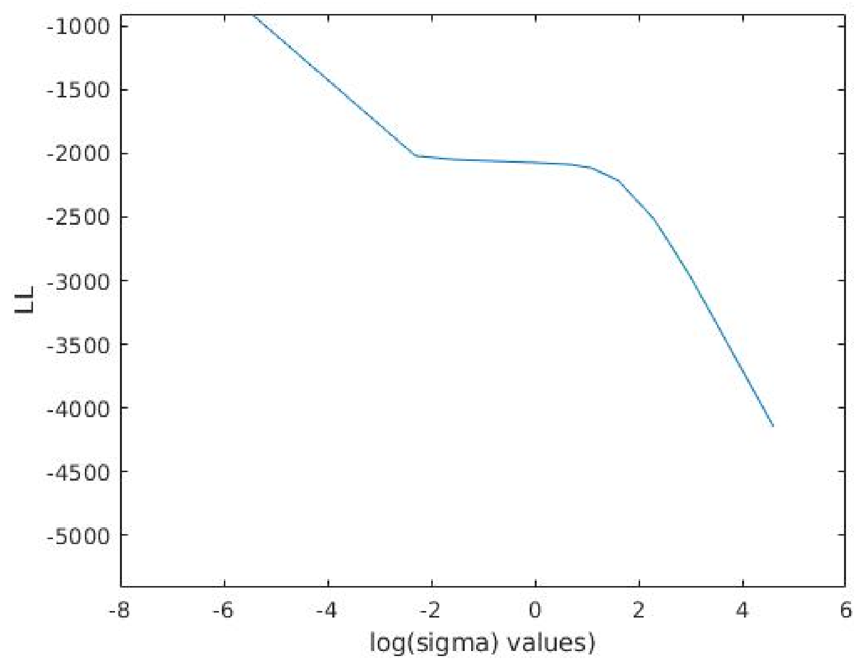
Plot of D vs. $log(\sigma)$:



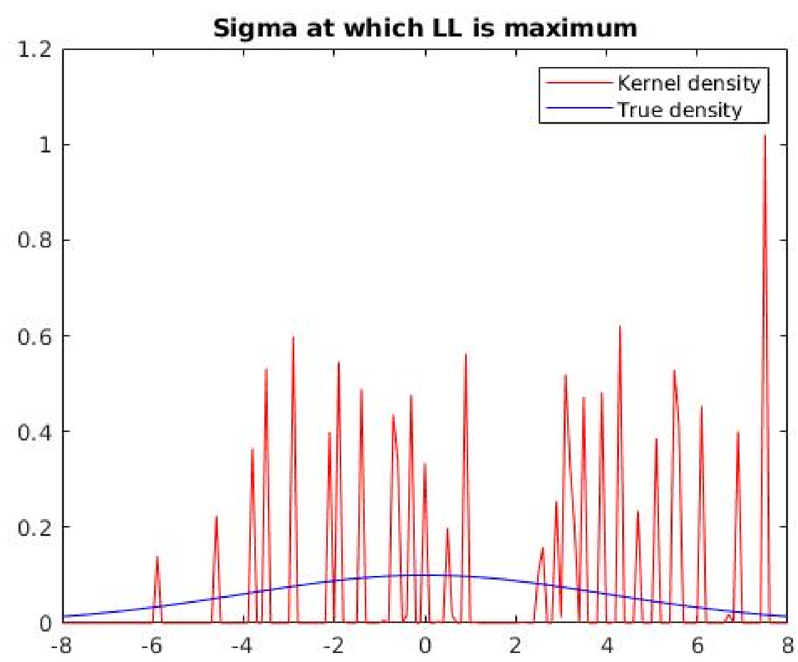
Plot of $\hat{p}_n(x; \sigma)$ vs. x :



e) Plot of LL vs $log(\sigma)$:



Plot of $\hat{p}_n(x;\sigma)$ vs. x :



Explaination:

Since every term of $\hat{p}_n(x_i;\sigma)$ will have 1 in the summation term (given in numerator) which will dominate all the other terms. (Since all other terms will be of order $\exp -100$ and therefore neglegible compared to 1). Therefore $\hat{p}_n(x_i;\sigma)$ will be inversely proportional to σ (given in the denominator). And hence it's value decreases with increase in σ and therefore value of σ at which LL is maximum will always be 0.001 (least σ value) which is not the σ we desired.
Therefore, this cross validation procedure fails when $V = T$.