CS 215 - Assignment 3

Niraj Mahajan
180050069

Raaghav Raaj
180050082

September 11, 2019

1. ~~Soln~~ X_i denotes additional number of times we have to pick a book such that we move from having picked books of $i-1$ distinct colours to i distinct colours.

(a) X_1 : Clearly, $X_1 = 1$.

Books with $i-1$ distinct colours have been collected.

$\Rightarrow$ the no. of books with distinct colours left,

$$n_i = n - i + 1$$

Thus, $P = \frac{n_i}{n} = \frac{n-i+1}{n}$

(b) We have that,

$$P[X_i \neq] = \frac{n-i+1}{n}.$$

Hence, the parameter of the geometric variable X_i, $p = \frac{n-i+1}{n}$

(c) For a ~~Be~~ geometric variable, we have that

$$P(Z=k) = (1-p)^{k-1} \cdot p \quad , k = 1, 2, 3 \ldots$$

• Now, expected value,

$$E(Z) = \sum_{i=1}^{\infty} i \cdot P(Z=i)$$

$$\Rightarrow E(Z) = \Sigma_i \cdot i \cdot (1-p)^{i-1} \cdot p.$$

$$\Rightarrow E(Z) = 1 \cdot (1-p)^0 \cdot p + 2 \cdot (1-p) \cdot p + 3 \cdot (1-p)^2 \cdot p + \cdots \infty \quad -①$$

$$\Rightarrow (1-p) \cdot E(Z) = 1 \cdot (1-p)^1 \cdot p + 2 \cdot (1-p)^2 p + 3 \cdot (1-p)^3 \cdot p + \cdots \infty \quad -②$$

• subtracting ② from ①,

$$p \cdot E(Z) = 1 \cdot (1-p)^0 \cdot p + 1 \cdot (1-p)^1 \cdot p + 1 \cdot (1-p)^2 \cdot p + \cdots \infty$$

$$= p \left\{ \sum_{i=0}^{\infty} (1-p)^i \right\} = p \cdot \frac{1}{1-(1-p)}$$

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

Now, for variance, we must calculate $E(z^2)$.

$$E(z^2) = \sum_{i=1}^{\infty} i^2 \cdot P(z=i).$$

$$\Rightarrow E[z^2] = \Sigma_i \, i^2 \cdot (1-p)^{i-1} \cdot p = p\left[\Sigma_i \, i^2 (1-p)^{i-1}\right].$$

~~Let $S = \sum_{i=1}^{\infty} i^2 (1-p)^{i-1}$~~

- Let $x = (1-p)$ and, $S = \Sigma_{i=0}^{i=\infty} i^2 \cdot x^{i-1}$

$$\Rightarrow S = 1 \cdot x^0 + 4 \cdot x^1 + 9 \cdot x^2 + \cdots \infty \quad -③$$

$$\Rightarrow x \cdot S = 1^2 \cdot x^1 + 2^2 \cdot x^2 + 3^2 \cdot x^3 + \cdots \infty \quad -④.$$

- Subtracting ④ from ③, we get

$$(1-x) \cdot S = 1 \cdot x^0 + 3 \cdot x^1 + 5 \cdot x^2 + \cdots \infty \quad -⑤.$$

$$\Rightarrow x(1-x)S = 1 \cdot x^1 + 3 \cdot x^2 + \cdots \infty \quad -⑥.$$

- Subtracting ⑥ from ⑤, we get.

$$(1-x)^2 \cdot S = 1 \cdot x^0 + 2\left[\Sigma_{i=1}^{\infty} x^i\right]$$

$$\Rightarrow (1-x)^2 \cdot S = 1 + \frac{2x}{1-x}$$

$$\Rightarrow S = \frac{1+x}{(1-x)^3} = \frac{2-p}{p^3}$$

- Hence, $\boxed{E(z^2) = p \cdot \frac{2-p}{p^3} = \frac{2-p}{p^2}}$

- Therefore,

$$Var(z) = E(z^2) - [E(z)]^2$$

$$\Rightarrow \boxed{Var(z) = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}}$$

(d) $E[X^{(n)}] = E[\sum_i X^{(i)}]$

$$\Rightarrow E[X^{(n)}] = \sum_{i=1}^{i=n} E[X_i^{(i)}]$$

$$\Rightarrow \boxed{E[X^{(n)}] = \sum_{i=1}^{n} \frac{n}{n+1-i} = \sum_{k=1}^{k=n} \frac{n}{k}}.$$

(e) $Var[X^{(n)}] = Var[\sum_{i=1}^{i=n} X_i]$

Since, $X_1, X_2, \ldots X_n$ are all independent.

$$\Rightarrow Var[X^{(n)}] = \sum_{i=1}^{i=n} Var[X_i]$$

$$\Rightarrow Var[X^{(n)}] = \sum_{i=1}^{i=n} \frac{1-p_i}{p_i^2}, \text{ where } p_i = \frac{n-i+1}{n}$$

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

$$\Rightarrow Var[X^{(n)}] = n \sum_{i=1}^{i=n} \frac{i-1}{(n-i+1)^2} = n \sum_{k=1}^{k=n} \frac{n-k}{k^2}$$

- Let $S = \sum_{k=1}^{k=n} \frac{n-k}{k^2}$

$$\Rightarrow S = \sum_{k=1}^{k=n} \frac{n}{k^2} - \sum_{k=1}^{k=n} \frac{1}{k}$$

- Hence,

$$Var[X^{(n)}] = \sum_{k=1}^{k=n} \frac{n^2}{k^2} - \sum_{k=1}^{k=n} \frac{n}{k}$$

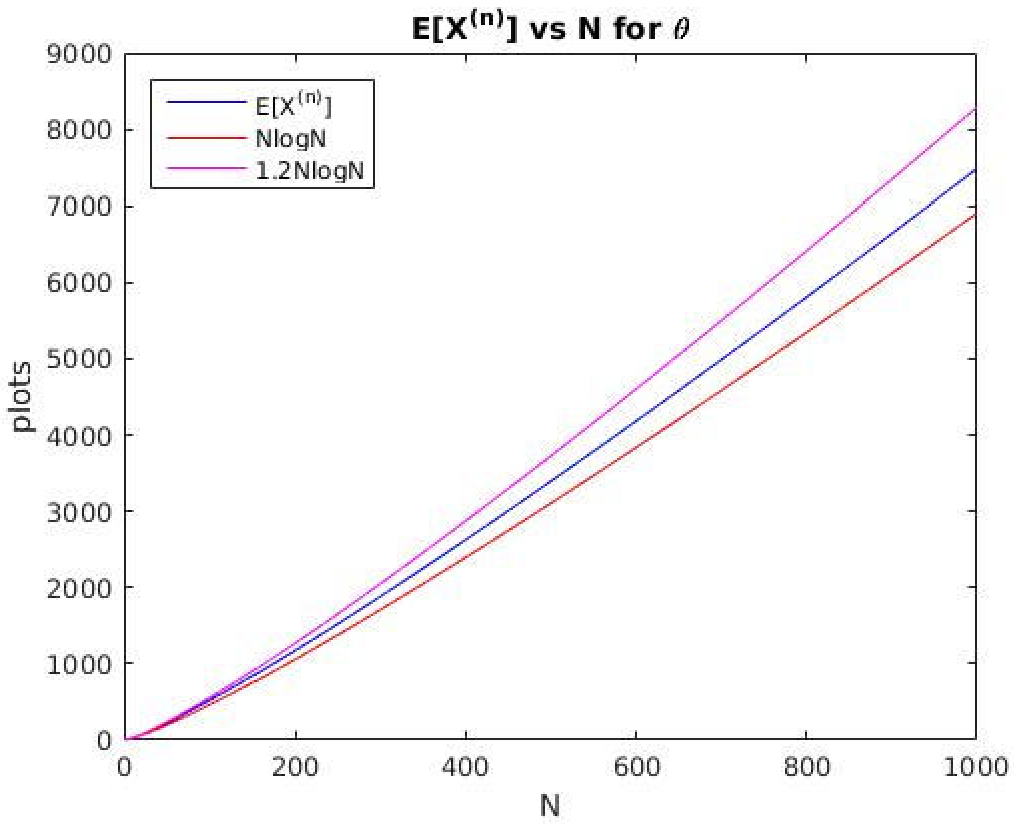
- Thus,

$$Var[X^{(n)}] \leq \sum_{k=1}^{k=n} \frac{n^2}{k^2}$$

$$\Rightarrow Var[X^{(n)}] \leq \sum_{k=1}^{k\to\infty} \frac{n^2}{k^2} = \frac{n^2\pi^2}{6}$$

$$\boxed{\therefore Var[X^{(n)}] \leq \frac{n^2\pi^2}{6}}$$

Plot $E[X^{(n)}]$ vs n:



Hence, $E[X^{(n)}] = \Theta n\log(n)$

2.

(a) The function be defined as,

$$F : D \longrightarrow [0,1]$$

and $F^{-1} : [0,1] \longrightarrow D$

- The function is invertible and thus, is monotic.

$$\Rightarrow \forall a, b \in D$$
$$a \leq b \leftrightarrow F(a) \leq F(b)$$

- Now, for a sample drawn from the ~~distribution~~ uniform $[0,1]$.

$$\forall z \in [0,1], P(u \leq z) = z$$

Now, consider a sample $u_i \in [0,1]$, we have that

$$v_i = F^{-1}(u_i)$$
$$\Rightarrow F(v_i) = u_i,$$

where $v_i \in D$.

Now, for any arbitrary $m \in D$,

$$P(v_i \leq m) = P(F(v_i) \leq F(m))$$

$$\Rightarrow P(v_i \leq m) = P(u \leq F(m)).$$

since, $u \in [0,1]$, and $F(m) \longrightarrow [0,1]$, we have that,

$$P(v_i \leq m) = F(m).$$

- Thus, $\{v_i\}_{i=1}^{n}$ follows the distribution F.

(b) We have that,

$$F_e(x) = \frac{\sum_{i=1}^{n} \mathbf{1}(Y_i \leq x)}{n}, \text{ where}$$

$\mathbf{1}(z) = 1$, if z is True.
$= 0$, if z is False.

- Hence, $\forall\ 0 \leq k \leq n,\ k \in \mathbb{Z}$,

$F_e(x) = \frac{k}{n}$, if and only if

exactly, k of $\{Y_i\}_{i=1}^{n} \leq x$.

- Since, $\{Y_i\}_{i=1}^{n}$ are independent and identically distributed variables, we find that

~~$F_e(x = \frac{k}{n}) = {}^nC_k (F(k))$~~

$$P\left[F_e(x) = k/n\right] = {}^nC_k \cdot [F(x)]^k \cdot [1 - F(x)]^{n-k} \quad \text{—— } (*)$$

- We have been given that,

$$D = \max_x \left(|F_e(x) - F(x)|\right)$$

→ Now, we draw $\{U_i\}_{i=1}^{n}$ from uniform $[0,1]$.

$\Rightarrow \forall y \in [0,1]$,

$\underline{P[U_i \leq y] = y}$.

- We have also defined,

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

- Let $G_e(y) = \frac{\sum_{i=1}^{n} \mathbf{1}(U_i \leq y)}{n}$.

Therefore, again,

$$P\left[G_e(y) = \frac{k}{n}\right] = {}^nC_k \cdot y^k \cdot (1-y)^{n-k} \quad \forall\ k \in \mathbb{Z}, k \in [0, n]$$

- We are provided with x, y and F, whereas $F_e(x)$ and $G_e(x)$ are random variables.

Thus, D and E are random variables.

- The function F maps to [0,1], and is a continuous distribution.

Hence, $\forall z \in [0,1]$,

$\exists x$, such that $F(x) = z$.

- We saw that,

$$P\left[F_e(x) = \frac{k}{n}\right] = {}^nC_k (F(x))^k (1 - F(x))^{n-k}$$

i.e the result depends on the value x is mapped to and not on x.

- Hence, the random variable D can be defined again as,

$$D = \max_{0 \le F(x) \le 1} |F_e(x) - F(x)|$$

- Clearly, the relation between $F_e(x)$ and $F(x)$, is identical to $G_e(y)$ and y, as can be seen above.

- Thus, using the expressions for D and E, we have that D and E are identically distributed random variables.

Also, D and E are independent.

$$\Rightarrow P(D \ge d) = P(E \ge d)$$

- Hence, proved.

→ D is a measure of deviation of the distribution of any arbitrary random variable from its continuous distribution for n.

And,

→ E is a measure of the deviation of the distribution in [0,1] from its true distribution. for same number of points n.

We proved that D and E are identically distributed.

- Therefore, we can say that, the empirical distributions of all random variables which also have continuous distribution, converge to the true distribution similarly, for the given number of points.

3. (a) We have that,

$$z = ax + by + c + \varepsilon$$

$$\forall \varepsilon \in \mathcal{N}(0, \sigma^2)$$

We conclude that,

$$z \in \mathcal{N}(ax + by + c, \sigma^2).$$

Hence, the likelihood function, is given as

$$lf = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2}\right)$$

$$\Rightarrow \log\text{-}lf = \left\{\sum_{i=1}^{n} -\frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2}\right\} - n\ln(\sigma\sqrt{2\pi})$$

Now,

- $$\frac{\partial(\log\text{-}lf)}{\partial a} = \sum_{i=1}^{n} \frac{(z_i - ax_i - by_i - c) \cdot x_i}{\sigma^2} = 0;$$

for maximizing wrt a

$$\Rightarrow \boxed{\sum z_i x_i = a\sum x_i^2 + b\sum y_i x_i + c\sum x_i}$$

- $$\frac{\partial(\log\text{-}lf)}{\partial b} = \sum_{i=1}^{n} \frac{(z_i - ax_i - by_i - c)\, y_i}{\sigma^2} = 0,$$

for maximizing wrt b.

$$\Rightarrow \boxed{\sum z_i y_i = a\sum x_i y_i + b\sum y_i^2 + c\sum y_i}$$

- $$\frac{\partial(\log\text{-}lf)}{\partial c} = \sum_{i=1}^{n} \frac{(z_i - ax_i - by_i - c)}{\sigma^2} = 0,$$

for maximizing wrt c

$$\Rightarrow \boxed{\sum z_i = a\sum x_i + b\sum y_i + nc}$$

Hence, the above equations can represented in matrix form as shown below.

$$\left[\begin{array}{ccc|c} \sum x_i^2 & \sum x_i y_i & \sum x_i & a \\ \sum y_i x_i & \sum y_i^2 & \sum y_i & b \\ \sum x_i & \sum y_i & \sum(1) & c \end{array}\right] = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}$$

- Let $\vec{X}, \vec{Y}, \vec{Z}$ be n dimensional vectors represented by sets $\{x_i\}_{i=1}^{n}$, $\{y_i\}_{i=1}^{n}$, $\{z_i\}_{i=1}^{n}$.

Let $\vec{I}$ be n dimensional identity vector $\Rightarrow \vec{I} = \{1\}_{i=1}^{n}$

$\Rightarrow$ Vector Form:

$$a(\vec{X}\cdot\vec{X}) + b(\vec{X}\cdot\vec{Y}) + c(\vec{X}\cdot\vec{I}) = \vec{X}\cdot\vec{Z}$$

$$a(\vec{X}\cdot\vec{Y}) + b(\vec{Y}\cdot\vec{Y}) + c(\vec{Y}\cdot\vec{I}) = \vec{Y}\cdot\vec{Z}$$

$$\cancel{b}a(\vec{X}\cdot\vec{I}) + b(\vec{Y}\cdot\vec{I}) + c(\vec{I}\cdot\vec{I}) = \vec{I}\cdot\vec{Z}$$

b) using MATLAB, we find that:

the predicted equation of the plane is

$$a_0 x + b_0 y + c_0 = z,$$

where

$a_0 = 10.00208$

$b_0 = 19.998022$

$c_0 = 29.951579$

and, predicted noise variance $= 23.068503$

***

For expectation of a, b and c,

we use the matrix form of the equations.

$$\begin{bmatrix} \sum x_i^2 & \sum y_i x_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & \sum(1) \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}$$

As we know, x_i and y_i with accuracy,

$$E(x_i) = x_i \text{ and } E(y_i) = y_i$$

Thus, using linearity of expectation operator, we get →

$$\begin{bmatrix} \Sigma x_i^2 & \Sigma x_i y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma y_i^2 & \Sigma y_i \\ \Sigma x_i & \Sigma y_i & \Sigma(1) \end{bmatrix} \begin{bmatrix} E(\hat{a}) \\ E(\hat{b}) \\ E(\hat{c}) \end{bmatrix} = \begin{bmatrix} E(\Sigma x_i z_i) \\ E(\Sigma y_i z_i) \\ E(\Sigma z_i) \end{bmatrix}.$$

$$\Rightarrow \begin{bmatrix} \Sigma x_i^2 & \Sigma x_i y_i & \Sigma x_i \\ \Sigma x_i y_i & \Sigma y_i^2 & \Sigma y_i \\ \Sigma x_i & \Sigma y_i & \Sigma(1) \end{bmatrix} \begin{bmatrix} E(\hat{a}) \\ E(\hat{b}) \\ E(\hat{c}) \end{bmatrix} = \begin{bmatrix} \Sigma x_i E(z_i) \\ \Sigma y_i E(z_i) \\ \Sigma E(z_i) \end{bmatrix} \quad \text{—} (*)$$

Now, we have that,

$$z = ax + by + c + \varepsilon$$

$$\forall\, \varepsilon \in \mathcal{N}(0,16) \Rightarrow E(\varepsilon) = 0$$

Hence, $E(z) = E(ax) + E(by) + E(c) + \cancel{E(\varepsilon)}^{0}$

$$\Rightarrow \boxed{E(z) = ax + by + c} \quad \text{———} (**)$$

as only z and ε are random variables

Simplifying the RHS using eq^n $(**)$

$$\begin{bmatrix} \Sigma x_i E(z_i) \\ \Sigma y_i E(z_i) \\ \Sigma E(z_i) \end{bmatrix} = \begin{bmatrix} a\Sigma x_i^2 + b\Sigma x_i y_i + c\Sigma x_i \\ a\Sigma x_i y_i + b\Sigma y_i^2 + c\Sigma y_i \\ a\Sigma x_i + b\Sigma y_i + c\Sigma 1 \end{bmatrix} = M \text{ (say)}.$$

$$\Rightarrow \quad M \cdot \begin{bmatrix} E(\hat{a}) \\ E(\hat{b}) \\ E(\hat{c}) \end{bmatrix} = M \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

Therefore,

$$\boxed{\begin{aligned} E(\hat{a}) &= a, \\ E(\hat{b}) &= b \\ E(\hat{c}) &= c \end{aligned}}$$

Usage of Matlab code for Q3

1) Find the code in the file 'Matlab_Code/Q3/q3.m'
2) Run the code, which will output the predicted equation of the plane and the predicted noise variance

4.

b) Joint likelihood =

$$\prod_{j=1}^{250} \hat{p}_{750}(x_j;\sigma)$$

$$= \prod_{j=1}^{250} \frac{\sum_{i=1}^{750} \exp\left(-\frac{(x_j-x_i)^2}{2\sigma^2}\right)}{750\sigma\sqrt{2\pi}}$$

c) and d) :

For maximum log likelihood,

$\sigma = 1$, log likelihood $= -691.090487$

For minimum D,

$\sigma = 1$, $D = 0.004875$

The value of D, for σ at which log likelihood was maximum is

$D = 0.004875$

e) The Kernel density estimation determined by the points in set T has peaks at each $x_i \in T$.

~~The~~ A low σ implies that the elements of T are indeed very likely to be drawn i.e high probability.

- As σ increases, the peaks flatten : the elements in T are picked by pure chance i.e completely randomly.

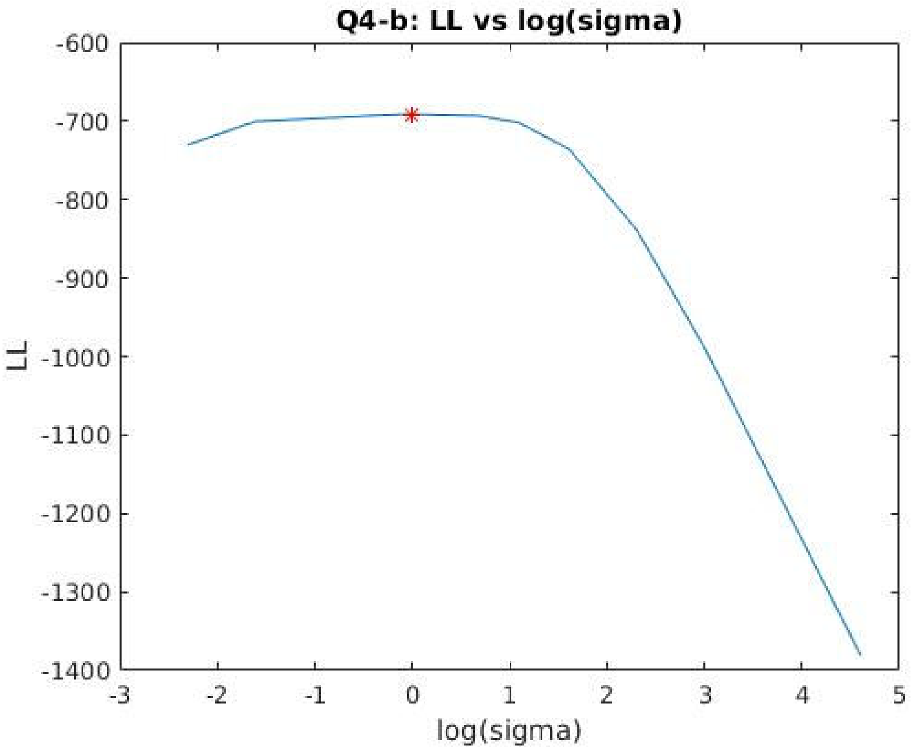
Now, we use V as the validation set. to find σ, and we will ~~f~~ observe the extent to which the peaks ~~of~~ of the two sets match.

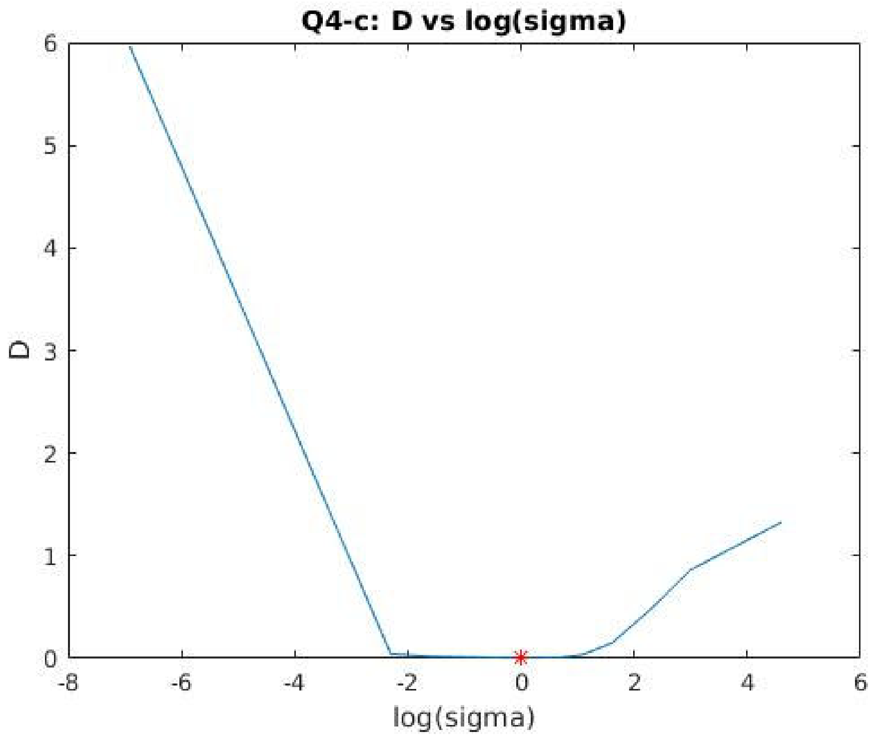
Now, if V = T, $\Rightarrow$ the elements of T have special ~~signifance~~ significance.

$\Rightarrow$ We can now choose the smallest σ, so that peaks are sharpest.

$\therefore$ For T=V, the graph is monotonically decreasing.

$\hookrightarrow$ LL vs σ





Usage of Matlab code for Q4

1) Find the code in the file 'Matlab_Code/Q4/q4.m'
2) Run the code, which will output the best values of sigma for maximum log likelihood and minimum D respetively, along with the values of D, LL
3) This will also produce two plots, one for Log likelihood vs log(sigma) and the other for D vs log(sigma)