

1. (a)  $x_i$  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,

$$\textcircled{1} - \quad n - n_i = n - i + 1 = 2$$

$$\textcircled{2} - \quad \text{Thus, } P = \frac{n_i}{n} = \frac{n - i + 1}{n} = 2$$

(b) We have that,

$$\textcircled{3} - \quad P[x_i \neq] = \frac{n - i + 1}{n}$$

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

(c) For a geometric variable, we have

$$\text{that } P(z=k) = (1-p)^{k-1} \cdot p, k=1,2,3,\dots$$

Now, expected value,

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

$$\Rightarrow E(z) = \sum i \cdot i \cdot (1-p)^{i-1} \cdot p$$

$$\Rightarrow E(z) = 1 \cdot (1-p)^0 \cdot p + 2 \cdot (1-p)^1 \cdot p + 3 \cdot (1-p)^2 \cdot p + \dots \infty - \textcircled{1}$$

$$\Rightarrow (1-p) \cdot E(z) = 1 \cdot (1-p)^1 \cdot p + 2 \cdot (1-p)^2 \cdot p + 3 \cdot (1-p)^3 \cdot p + \dots \infty - \textcircled{2}$$

Subtracting  $\textcircled{2}$  from  $\textcircled{1}$ ,

$$\begin{aligned} p \cdot E(z) &= 1 \cdot (1-p)^0 \cdot p + 1 \cdot (1-p)^1 \cdot p + 1 \cdot (1-p)^2 \cdot p + \dots \infty \\ &= p \left\{ \sum_{i=0}^{\infty} (1-p)^i \right\} = p \cdot \frac{1}{1-(1-p)} \end{aligned}$$

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

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

$$E(z^2) = \sum_{i=1}^{\infty} i^2 \cdot P(z=i)$$
$$\Rightarrow E[z^2] = \sum_i i^2 \cdot (1-p)^{i-1} \cdot p = p \left[ \sum_i i^2 (1-p)^{i-1} \right].$$

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

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

$$\Rightarrow S = 1 \cdot x^0 + 4 \cdot x^1 + 9 \cdot x^2 + \dots \infty \quad - \textcircled{3}$$

$$\Rightarrow x \cdot S = 1^2 \cdot x^1 + 2^2 \cdot x^2 + 3^2 \cdot x^3 + \dots \infty \quad - \textcircled{4}$$

• Subtracting  $\textcircled{4}$  from  $\textcircled{3}$ , we get

$$(1-x) \cdot S = 1 \cdot x^0 + 3 \cdot x^1 + 5 \cdot x^2 + \dots \infty \quad - \textcircled{5}$$

$$\Rightarrow x(1-x)S = -1 \cdot x^1 + 3 \cdot x^2 + \dots \infty \quad - \textcircled{6}$$

• Subtracting  $\textcircled{6}$  from  $\textcircled{5}$ , we get

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

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

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

• Hence,

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

• Therefore,

$$\text{Var}(z) = E(z^2) - [E(z)]^2$$

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

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

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

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

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

Since,  $(X_1, X_2, \dots, X_n)$  are all independent.

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

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

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

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

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

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

- Hence,

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

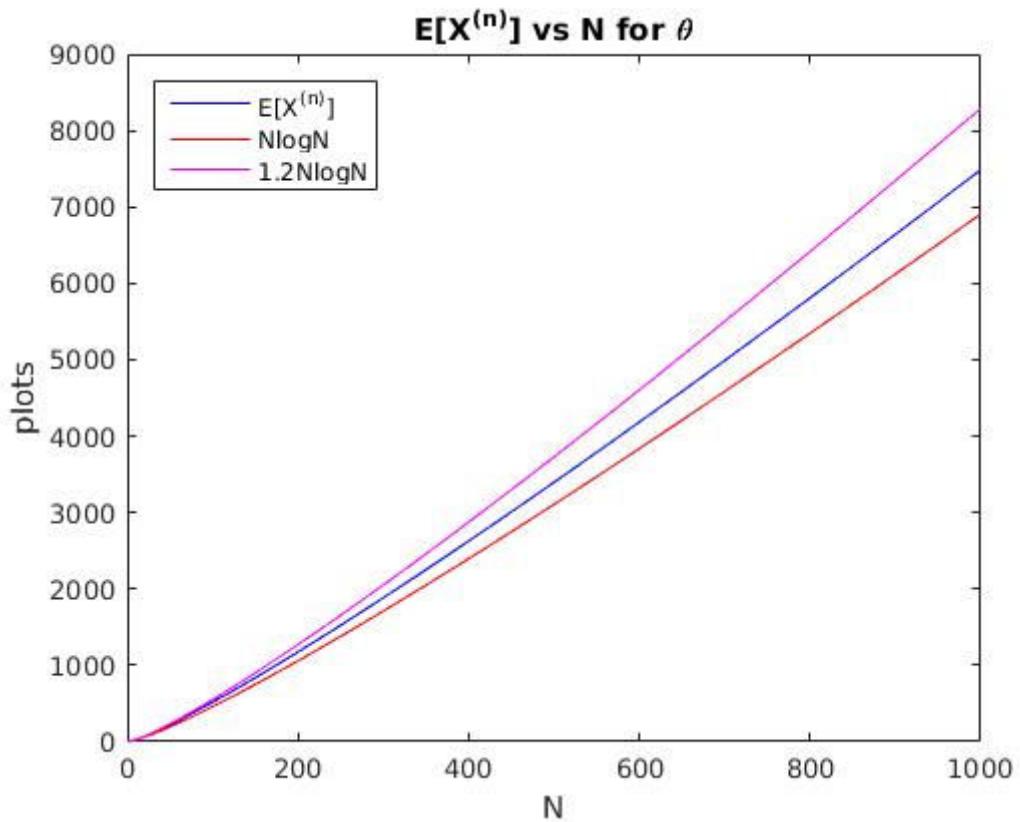
- Thus,

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

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

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

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



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

2:

(a) The function be defined as,

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

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

- The function is invertible and thus, is monotonic.

*arbitrary and also no  $\Rightarrow$*

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

- 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  $F(m) \in [0,1]$ ,

we have that,

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

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

(b)

We have that,

$$F_e(x) = \frac{\sum_{i=1}^n \mathbf{1}(Y_i \leq x)}{n}, \text{ where } \begin{cases} 1(z) = 1 & \text{if } z \text{ is True.} \\ 1(z) = 0 & \text{if } z \text{ is False.} \end{cases}$$

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

we have  $F_e(x) = \frac{k}{n}$ , 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

$$(F_e(x) = \frac{k}{n}) \iff \binom{n}{k} (F(x))^k \cdot (1-F(x))^{n-k}$$

$$\therefore P[F_e(x) = k/n] = \binom{n}{k} [F(x)]^k \cdot [1-F(x)]^{n-k} \quad (*)$$

- We have been given that,

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

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

$$\Rightarrow \forall y \in [0, 1], P[U_i \leq y] = y$$

- We have also defined,

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

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

$$\text{Therefore, again, } G_e(y) = \frac{\sum_{i=1}^n \mathbf{1}(U_i \leq y)}{n} = \binom{n}{k} y^k (1-y)^{n-k} \quad \forall k \in \mathbb{Z}, k \in [0, n]$$

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

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

- We saw that,

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

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,

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

- 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.

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

- 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 some 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.

30(a) We have that,

$$z = ax + by + c + \epsilon, \epsilon$$

$$\forall \epsilon \in N(0, \sigma^2)$$

We conclude that,

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

Hence, the likelihood function, is given as

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

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

$$\hat{\Sigma} = (\hat{X}^T \hat{X})^{-1} + (\hat{Y}^T \hat{Y})^{-1} + (\hat{Y}^T \hat{X})$$

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

for maximizing wrt. a

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

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

for maximizing wrt. b.

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

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

for maximizing wrt. c

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

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

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

- 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:

$$\left[ \begin{array}{l} a(\vec{x} \cdot \vec{x}) + b(\vec{x} \cdot \vec{y}) + c(\vec{x} \cdot \vec{I}) = \vec{x} \cdot \vec{z} \\ a(\vec{x} \cdot \vec{y}) + b(\vec{y} \cdot \vec{y}) + c(\vec{y} \cdot \vec{I}) = \vec{y} \cdot \vec{z} \\ b(\vec{x} \cdot \vec{I}) + b(\vec{y} \cdot \vec{I}) + c(\vec{I} \cdot \vec{I}) = \vec{I} \cdot \vec{z} \end{array} \right]$$

- b) Using MATLAB, we find that:  
the predicted equation of the plane is

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

where

$$a_0 = 10.00268$$

$$b_0 = 19.998022$$

$$c_0 = 29.951579$$

and, predicted noise variance = 23.068503

\*\*\* For expectation of  $a, b$  and  $c$ , we use the matrix form of the equations.

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

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  $\rightarrow$

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

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

Now, we have that,

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

$$\forall \epsilon \in N(0, \sigma^2) \Rightarrow E(\epsilon) = 0$$

$$\text{Hence, } E(z) = E(ax) + E(by) + E(c) + E(\epsilon)$$

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

as only  $z$  and  $\epsilon$  are random variables

Simplifying the RHS using eqn (\*\*)

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

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

Therefore,

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

$$M \cdot \begin{bmatrix} (a)I_3 \times 3 \\ (b)I_3 \times 3 \end{bmatrix} = \begin{bmatrix} (a)I_3 \\ (b)I_3 \end{bmatrix} \begin{bmatrix} I_3 \times 3 & I_3 \times 3 & I_3 \times 3 \\ I_3 \times 3 & I_3 \times 3 & I_3 \times 3 \\ I_3 \times 3 & I_3 \times 3 & I_3 \times 3 \end{bmatrix} \leftarrow$$

## 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

b) Joint likelihood =  $\prod_{i=1}^n p(x_i | \theta)$

b) Joint likelihood = maximum joint似然

$$\prod_{j=1}^{250} \hat{P}_{150}(x_j; \sigma) \text{, in rot manier aufwärts}$$

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

c) and d) :

For maximum log likelihood,  
 $\sigma = 1$ ,  $\log \text{likelihood} = -691.090487$

For minimum  $P$ ,  $\text{austerns}$   $\text{geitdritib}$   $\text{out}$

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

The value of  $D$ , for  $\sigma$  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  $x_i \in T$ .

in set I has peaks at even ...  
This implies that the ele

The Axiom 5 implies that the elements of  $T$  are

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

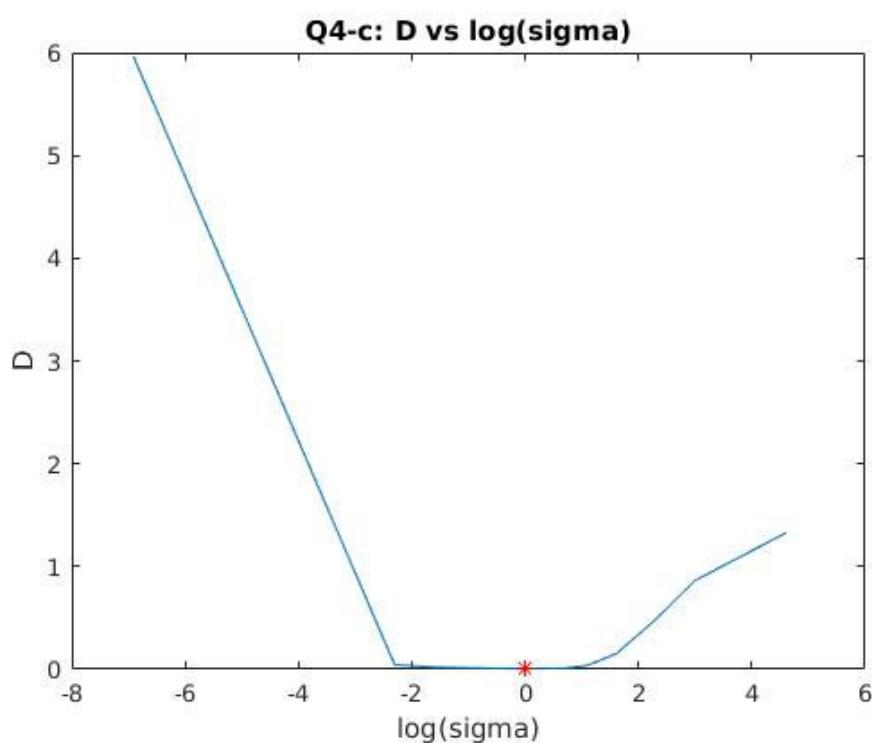
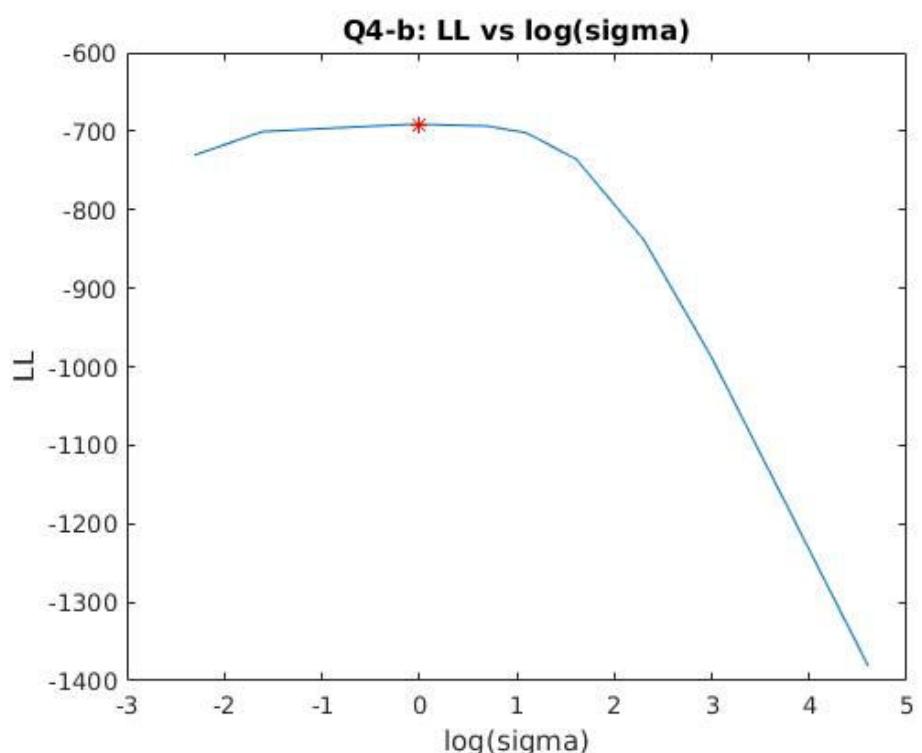
- As  $\sigma$  increases, the peaks flatten: the elements in T are picked by pure chance i.e completely randomly.

Now, we use  $V$  as the validation set. to find  $\sigma$ , and we will observe the extent to which the peaks of the two sets match.

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

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

$\therefore$  For  $T = V$ , the graph is monotonically decreasing.  
 $\hookrightarrow$  LL vs  $\sigma$



## 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 respectively, 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)