CS 215.

Details of students in the group for the assignment:

1.) Name: TATHAGAT VERMA Roll Number: 180050111.

2.) Name: NEEL ARYAN GUPTA Roll Number: 180050067.

ANSWERS: Follow from next page.

- Instructions for running the MATLAB code is specified in the INSTRUCIONS.txt file.

Question 1 :

(a) X_1 = no. of times additionaly a book has to be picked so that we move from 0 distinct colors to 1 distinct color

$= \underline{1}$. (as the 1st book itself serves as the 1st distinct color)

$\Rightarrow \boxed{X_1 = 1}$ (Ans.)

(b) $P(X_i = k)$ = P((k-1) books picked belonged to the (i-1) colors & the kth book belongs to the (n-i+1) other colors)

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

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

$$= (1-p)^{k-1} p.$$

$\Rightarrow X_i \sim$ Geometric random variable with parameter $= \boxed{\frac{n-i+1}{n}}$ (Ans.)

(C) $p(Z=k) = (1-p)^{k-1} p$ To prove : $\underline{E(Z) = 1/p}$.

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

Now consider $g(x) = \sum_{k=1}^{\infty} x(1-x)^k$

$$g(x) = \frac{x(1-x)}{1-(1-x)} = (1-x) \qquad \{ x \in (0,1) \}.$$

$$g'(x) = \sum_{k=1}^{\infty} (1-x)^k - \sum_{k=1}^{\infty} kx(1-x)^{k-1}$$

$$\Rightarrow \frac{d}{dx}\left((1-x)\right) = \frac{(1-x)}{x} - \sum_{1}^{\infty} kx(1-x)^{k-1}$$

$$\Rightarrow -1 = \frac{1}{x} - 1 - \sum_{1}^{\infty} kx(1-x)^{k-1}$$

$$\Rightarrow \boxed{\sum_{1}^{\infty} kx(1-x)^{k-1} = \frac{1}{x}} \qquad \{ x \in (0,1) \}$$

Put $x = p$; valid as $0 < p \leq 1$

$$\Rightarrow \sum_{1}^{\infty} kp(1-p)^{k-1} = \boxed{\frac{1}{p} = E(Z)}$$

Hence proved.

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

$$E(Z) = 1/p.$$

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

$$E(Z)$$

$$= \sum_1^{\infty} (k(k-1) + k)(1-p)^{k-1} p$$

$$= \sum_1^{\infty} p(1-p)\, k(k-1)(1-p)^{k-2} + \sum_1^{\infty} pk(1-p)^{k-1}$$

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

Now consider again $g(x) = \sum_{k=1}^{\infty} (1-x)^k$

$$g'(x) = \sum_1^{\infty} -k(1-x)^{k-1}$$

$$g''(x) = \sum_1^{\infty} k(k-1)(1-x)^{k-2}$$

$$g(x) = \frac{(1-x)}{1-(1-x)} = \frac{1-x}{x} = \frac{1}{x} - 1 = g(x).$$

$$g'(x) = -\frac{1}{x^2}. \Rightarrow \boxed{g''(x) = \frac{2}{x^3}}$$

$$\Rightarrow \boxed{\frac{2}{p^3} = \sum_1^{\infty} k(k-1)(1-p)^{k-2}}$$

$$\Rightarrow E(Z^2) = \frac{2p(1-p)}{p^3} + \frac{1}{p} = \frac{2}{p^2} - \frac{2}{p} + \frac{1}{p} = \boxed{\frac{2}{p^2} - \frac{1}{p}}$$

$\Rightarrow \text{Var}(Z) = E(Z^2) - E(Z)^2$

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

$\Rightarrow \boxed{\text{Var}(Z) = \frac{1-p}{p^2}}$ Variance of Geometric random variable.

(Ans)

(d) $E(X^{(n)}) = \sum_{1}^{n} E(X_i)$

$$= \sum_{i=1}^{n} \frac{1}{p_i} \qquad p_i = \frac{n-i+1}{n}.$$

$$= \sum_{i=1}^{n} \frac{n}{n-i+1}$$

(Ans)

$$\boxed{\begin{aligned} E(X^{(n)}) &= n\left(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + \frac{1}{1}\right) \\ &= \sum_{i=1}^{n} \left(\frac{n}{n-i+1}\right) \end{aligned}}$$

Scanned by CamScanner

(e.) $Var(X^{(n)}) = Var(\sum_{1}^{n} X_i)$

$$= \sum_{1}^{n} Var(X_i) + 2\sum\sum_{i<j} \cancel{Cov(X_i, X_j)}$$

0 (X_i, X_j) are independent.

$$= \sum Var(X_i)$$

$$= \sum_{1}^{n} \frac{1-p_i}{p_i^2} \qquad p_i = \frac{n-i+1}{n}.$$

$$= \sum_{1}^{n} \frac{\left(\frac{i-1}{\cancel{n}}\right) n^{\cancel{2}}}{(n-i+1)^2}$$

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

$$< n \times n \left(\sum_{1}^{n} \frac{1}{(n-i+1)^2}\right)$$

$$= n^2\left(\frac{1}{1^2} + \frac{1}{2^2} + \cdots + \frac{1}{n^2}\right)$$

$$< \frac{n^2 \pi^2}{6}$$

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

$\Rightarrow$ The required upper bound is $\boxed{\frac{n^2\pi^2}{6}}$ (Ans) f.) $E(X^{(n)}) = n\left(\frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}\right) = n \cdot S_n$

$S_n = \frac{1}{1}+\frac{1}{2}+\cdots+\frac{1}{n}$.

for large n;

$S_n \approx \int_{1/n}^{1} \frac{1}{x}\,dx$ $\quad \left\{ \sum \frac{1}{i/n}(1/n) = \sum \frac{1}{i} = \int_{\alpha}^{1} \frac{dx}{x} \quad \alpha = 1/n \right\}$

$= \ln x \Big|_{1/n}^{1}$

$= \log 1 - \log(1/n)$

$= \underline{\log n}$.

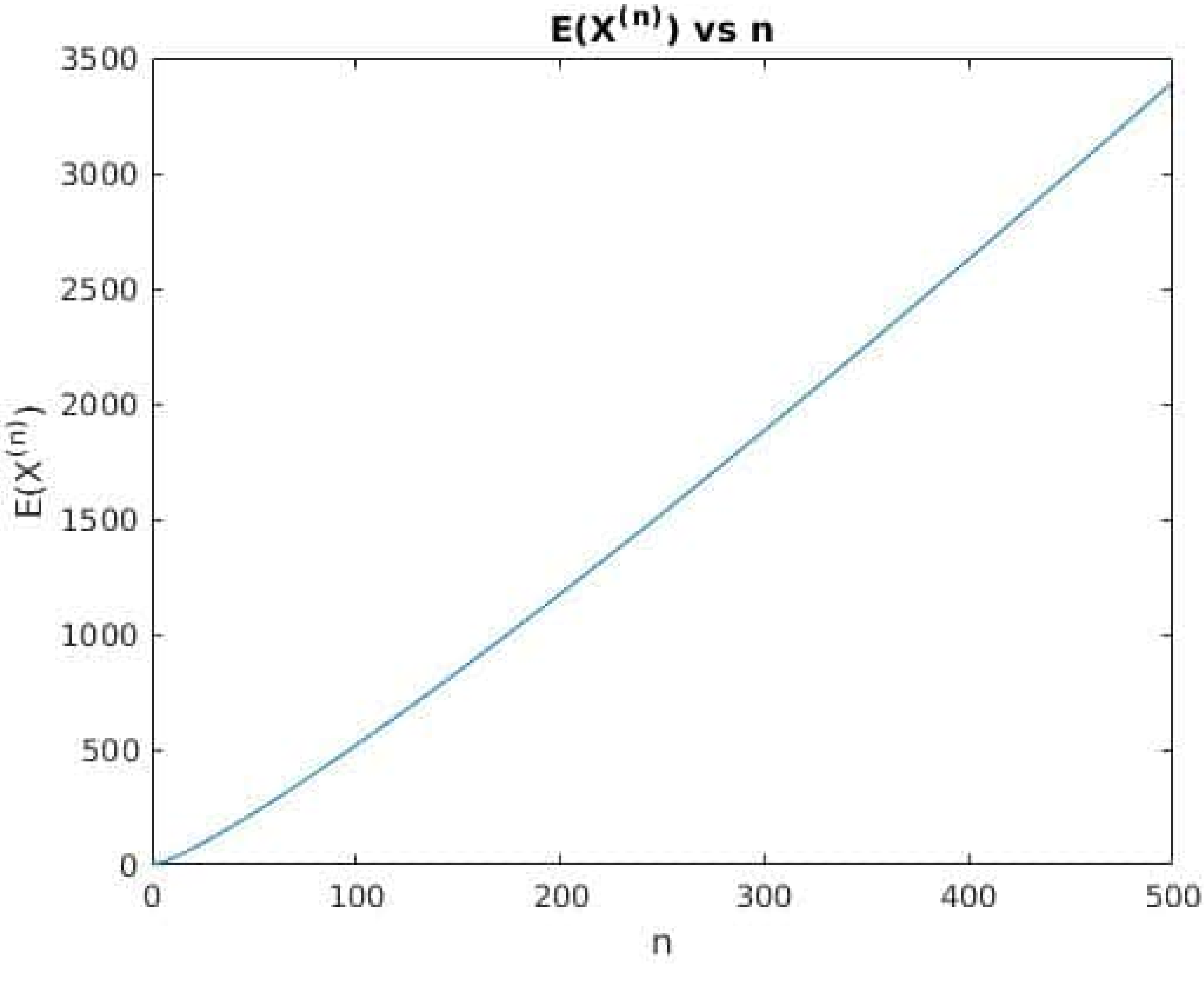
$\Rightarrow \lim_{n \to \infty} S_n = \log n$.

$\Rightarrow E(X^{(n)}) = n-1\,(n S_n)$

$= \underline{n-1\,(n \log(n))}$.

$\Rightarrow \boxed{f(n) = n \log(n)}$ $\quad$ <u>(Ans.)</u>

Plot of $E(X^{(n)})$ vs n is attached below.


E(X^(n)) vs n
3500
3000
2500
2000
1500
1000
500
0
E(X^(n))
0
100
200
300
400
500
n

Question - 2

(a) Given $u_i \sim$ uniform $(0,1)$

$\Rightarrow P(u_i \le k) = k$ for $k \in (0,1)$

Since F is a cdf and is given invertible

$\Rightarrow F$ is strictly increasing (since F is always non-decreasing as it is a cdf)

$\therefore v_i = F^{-1}(u_i)$ for $i = \{1, 2 \ldots n\}$

$\therefore F_v(x) = P(v_i \le x)$

$= P(F^{-1}(u_i) \le x)$

Since F is increasing

$F^{-1}(u_i) \le x \Rightarrow u_i \le F(x)$

$\therefore F_v(x) = P(F^{-1}(u_i) \le x)$

$= P(u_i \le F(x))$

$= F(x)$

$\Rightarrow v_i$ has the distribution defined by F.

Hence, proved.

(b) For simplicity, we assume that F is strictly increasing instead of being non-decreasing

ie. we assume that F^{-1} exists.

F^{-1} is also strictly increasing

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

Since we assumed F^{-1} exists, there exists y such that $y = F(x)$ which is unique

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

$= \max_{-\infty < x < \infty} |F_e(F^{-1}(y)) - F(F^{-1}(y))|$

$= \max_{0 < y < 1} |F_e(F^{-1}(y)) - y|$

The last equality comes from the fact that for every $\infty < x < \infty$, there is a $0 \le y \le 1$ such that $y = F(x)$. So we can always establish a map from x to y such that we can index them.

Note that,

$$F_e(F^{-1}(y)) = \frac{1}{n} \sum 1(Y_i \le F^{-1}(y))$$

$$= \frac{1}{n} \sum 1(F(Y_i) \le y)$$

Now $\{F(Y_i)\}$ are also i.i.d random variables.

Note that,

$$P(F(Y_i) \le y) = P(Y_i \le F^{-1}(y))$$

$\therefore$ $P(Y_i \le k) = F(k)$

$$\Rightarrow P(F(Y_i) \le y) = P(Y_i \le F^{-1}(y))$$
$$= F(F^{-1}(y))$$
$$= y$$

$\Rightarrow$ $F(Y_i) \sim$ Uniform $(0, 1)$ [Since $P(U_i \le k) = k$]

$\Rightarrow$ $F(Y_i) = U_j$ (for some j)

$$\therefore F_e(F^{-1}(y)) = \frac{1}{n} \sum 1(F(Y_i) \le y)$$

$$= \frac{1}{n} \sum 1(U_j \le y)$$

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

$$= \max_{0 \le y \le 1} \left| \frac{1}{n} \sum 1(U_j \le y) - y \right|$$

$\Rightarrow D = \max_{0 \le y \le 1} \left| \frac{1}{n} \sum 1(V_i \le y) - y \right|$

$= E$

$\Rightarrow$ So D & E have the same distribution

i.e. $P(D \ge x) = P(E \ge x)$

In the general case, where F^{-1} does not exist we can use another function G which is defined as follows-

$G(x) = \min\{y : F(x) > y\}$

Note that $G(x)$ is a one-one mapping & we can replace $F^{-1}(x)$ everywhere by $G(x)$ to get the same result.

The practical significance for this result is given on next page.

The interpretation of the preceding result can be as follows-
Any arbitrary empirical distribution will converge to the actual distribution to the same extent as the uniform distribution w.r.t n, the number of samples.
Since this relationship is transitive in nature, we can extend this relation to the empirical distribution of any two random variables provided their distribution functions are continuous.
The most remarkable property of this result is that the random variable D is independent of the underlying function F i.e. it is independent of the distribution of the original random variables.

Question - 3

Given N points (x_i, y_i), we have

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

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

Known: $\{x_i, y_i\}$ accurately

The $\{z_i\}$ have been corrupted by noise from $N(0, \sigma^2)$

Now, $z_i \sim N(ax_i + by_i + c, \sigma^2)$

$$P(z_i \mid x_i, y_i, a, b, c) = \frac{e^{-\frac{(z_i - (ax_i + by_i + c))^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-(z_i - (ax_i + by_i + c))^2 / 2\sigma^2\right)$$

$$\underset{(\forall i)}{\log(P(z_i \mid x_i, y_i, a, b, c))} = -\sum_{i=1}^{N}\left[\frac{(z_i - (ax_i + by_i + c))^2}{2\sigma^2}\right] - n\log\sqrt{2\pi} - n\log\sigma$$

$$\Rightarrow \mathcal{L}(x_i, y_i, z_i, a, b, c) = -\sum_{i=1}^{N}\left[\frac{(z_i - (ax_i + by_i + c))^2}{2\sigma^2}\right] - n\log\sqrt{2\pi} - n\log\sigma$$

$$\therefore \frac{\partial \mathcal{L}}{\partial a} = +\sum_{i=1}^{N} \frac{2(z_i - ax_i - by_i - c) \cdot x_i}{2\sigma^2} = 0$$

$$\Rightarrow \sum^{N} x_i z_i = a\sum^{N} x_i^2 + b\sum^{N} x_i y_i + c\sum^{N} x_i \quad \text{——— (1)}$$

$$\therefore \frac{\partial \mathcal{L}}{\partial b} = +\sum_{i=1}^{N} \frac{2(z_i - ax_i - by_i - c) \cdot y_i}{2\sigma^2} = 0$$

$$\Rightarrow \sum^{N} y_i z_i = a\sum^{N} x_i y_i + b\sum^{N} y_i^2 + c\sum^{N} y_i \quad \text{——— (2)}$$

$$\therefore \frac{\partial \mathcal{L}}{\partial c} = +\sum_{i=1}^{N} \frac{2(z_i - ax_i - by_i - c) \cdot 1}{2\sigma^2} = 0$$

$$\Rightarrow \sum^{N} z_i = a\sum^{N} x_i + b\sum^{N} y_i + Nc \quad \text{——— (3)}$$

In matrix / vector form.

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

Let $\bar{I}$ denote the identity vector: $(1, 1, \underbrace{\ldots}_{n \text{ times}}, 1)$

& $\bar{X}, \bar{Y}, \bar{Z}$ denote the $\{x_i\}, \{y_i\}, \{z_i\}$ vectors

So,
$$\begin{bmatrix} \bar{X}.\bar{X} & \bar{X}.\bar{Y} & \bar{X}.\bar{I} \\ \bar{X}.\bar{Y} & \bar{Y}.\bar{Y} & \bar{Y}.\bar{I} \\ \bar{X}.\bar{I} & \bar{Y}.\bar{I} & \bar{I}.\bar{I} \end{bmatrix} \begin{bmatrix} \hat{a} \\ \hat{b} \\ \hat{c} \end{bmatrix} = \begin{bmatrix} \bar{X}.\bar{Z} \\ \bar{Y}.\bar{Z} \\ \bar{Z}.\bar{I} \end{bmatrix}$$

Using Matlab for solving the above matrix

We get $a = 10.002208$

$b = 19.998022$

$c = 29.951579$

Expected noise variance = 23.068503

Eqn.: $z = 10.002208x + 19.998022y + 29.951579$

Matrix form: $AV = B$

Note that here A is fixed i.e. known with certainty

$\Rightarrow E(AV) = E(B)$

$\Rightarrow A\,E(V) = E(B)$

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

$z_i = \hat{a}x_i + \hat{b}y_i + \hat{c} + \varepsilon$ where $\varepsilon \sim N(0, \sigma^2)$

$\Rightarrow E(z_i) = ax_i + by_i + c + 0 = ax_i + by_i + c$

$$\Rightarrow \begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i \\ \sum x_i & \sum y_i & \sum 1 \end{bmatrix} \begin{bmatrix} E(\hat{a}) \\ E(\hat{b}) \\ E(\hat{c}) \end{bmatrix} = \begin{bmatrix} \sum x_i(ax_i+by_i+c) \\ \sum y_i(ax_i+by_i+c) \\ \sum(ax_i+by_i+c) \end{bmatrix}$$

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

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

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

Since A is invertible

$$\Rightarrow E(\hat{a}) = a$$
$$E(\hat{b}) = b$$
$$E(\hat{c}) = c$$

For variance,

$$Var(z_i) = Var(ax_i + by_i + c + N(0, \sigma^2))$$
$$= \sigma^2$$

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

$$= \begin{bmatrix} \sigma^2 \sum x_i^2 \\ \sigma^2 \sum y_i^2 \\ \sigma^2 \sum 1 \end{bmatrix}$$

Question 4 :

(b) $$JL = \prod_{j=1}^{250} \left(\frac{\sum_{i=1}^{750} \exp(-(v_j - t_i)^2 / 2\sigma^2)}{750\,\sigma\sqrt{2\pi}} \right)$$

where $\{v_1, v_2, .., v_{250}\}$ is set V.

$\{t_1, t_2, .., t_{750}\}$ is set T.

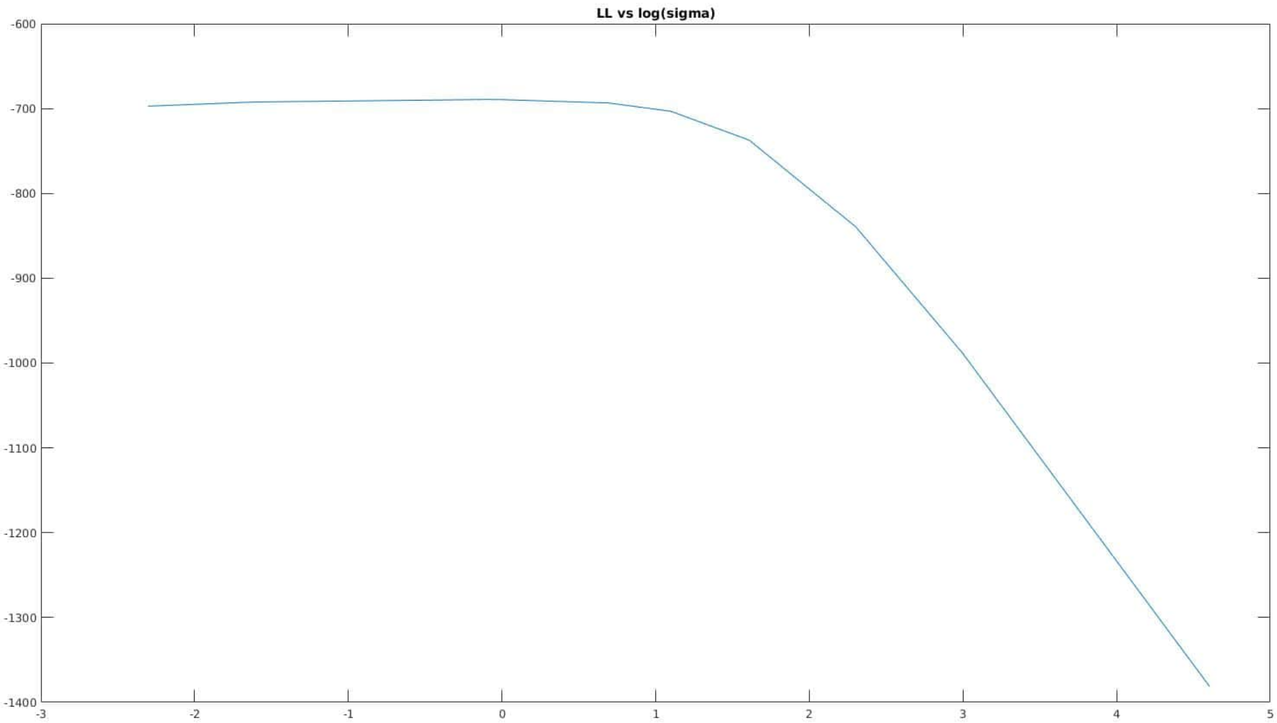
(c) $\sigma = 0.9$ yielded the maximum value ~~for~~ of LL (and hence the best value of LL)
(since T, V are randomly chosen, the best σ value changes on repeated iterations of the same code; but it remains one of 0.9, 1)

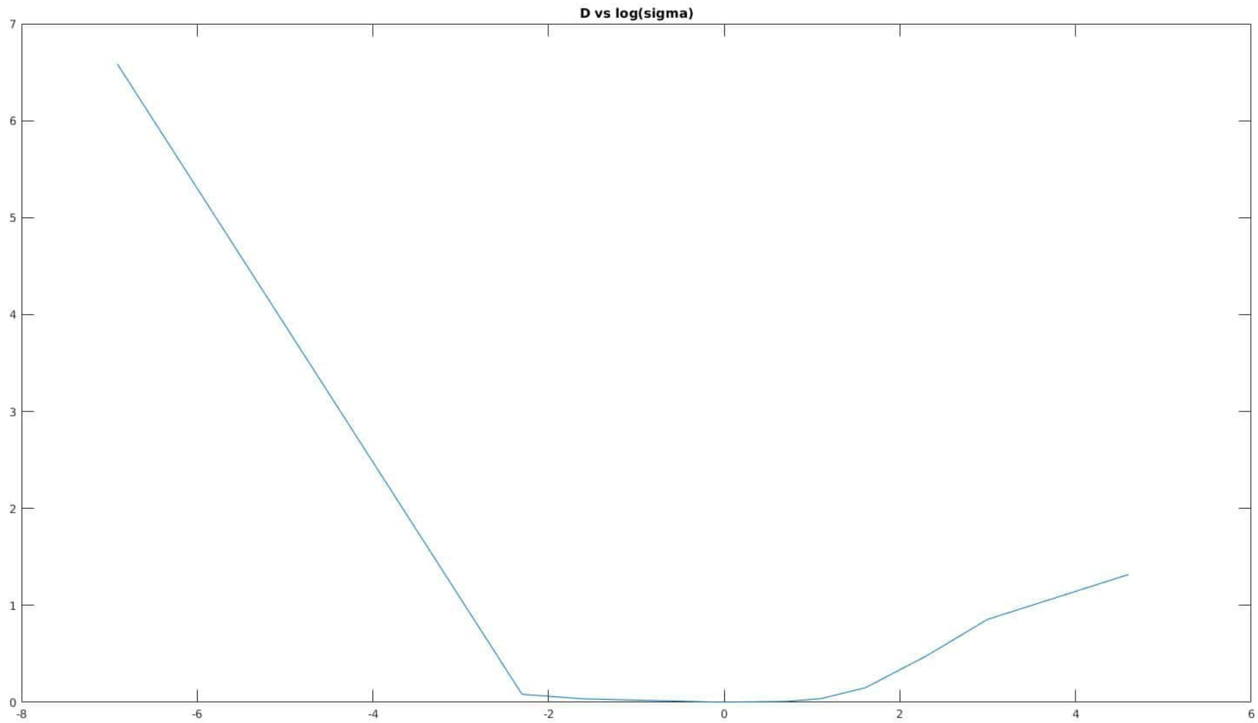
Plot of graph of LL versus $\log \sigma$ is attached below.

(d) $\sigma = 1.000$ yielded the minimum & hence best value for D

value of D for $\sigma = 0.9$ (best LL producer) is 0.00335 (Ans)

Scanned by CamScanner


LL vs log(sigma)
-600
-700
-800
-900
-1000
-1100
-1200
-1300
-1400
-3
-2
-1
0
1
2
3
4
5

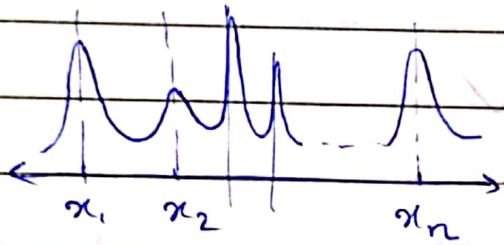


D vs log(sigma)
7
6
5
4
3
2
1
0
-8
-6
-4
-2
0
2
4
6

(e.) For this we first observe $\hat{p}_n$ & what σ represents.

$\hat{p}_n(x;\sigma)$ = average of probabilities of occurrence of x_i given a Gaussian centred at x with variance σ^2.

The graph of $\hat{p}_n(x;\sigma)$ will be somewhat like



peaks like these around (not exactly) at x_i's.

Now the σ is "bandwidth", this is related to the average (not precisely but somewhat) widths of the peaks.

When cross-validation is done, ~~or~~ if many values of set V are different (or have large differences with values of T). then the graph of $\hat{p}_n$ will widen out causing larger σ (larger sigma means that σ which maximises LL becomes larger)

So when T & V are identical,
the graph of $\hat{p}_n$ will get more concentrated

(will have sharper peaks) at around x_i's
as ~~are~~ the predictions which we had before are
just reinforced.

Hence for max LL, σ will reduces as
peaks became sharper.

Thus σ ~~red.~~ giving maximum LL
reduces (i.e. lesser than the ideal σ)

when set T & V are identical, in the
cross validation procedure.
Explanation given already!

{ Also using MATLAB plots, one can confirm
that LL is ~~max~~ indeed maximised for
lower σ }