CS215 : Assignment 3

Dion Reji (22B0029)
Rijul Bhat (22B0971)
Thomas Biju Cheeramvelil (22B1073)

Question 1

Question 1

(a) $X_1(\omega)$ = additional number of times you have to pick a book so that number of distinct colours of picked books changes from 0 to 1, in the sample point (ω)

$= \underline{\underline{1}}$

[∵ No books were picked intially, since initial number of distinct colours is 0. After choosing any 1 book, the number of distinct colours of picked books increases to 1, since all the books had unpicked colours.]

$$P\left(\begin{array}{l}\text{picking a book of different}\\ \text{colour, when } (i-1) \text{ distinct}\\ \text{colours have been picked}\end{array}\right) = \frac{\begin{array}{l}\text{Number of books}\\ \text{with colours different}\\ \text{than the } (i-1) \text{ collected}\\ \text{colours}\end{array}}{\text{Total number of books}}$$

$$= \frac{n-(i-1)}{n}$$

$$= \underline{\underline{\frac{n-i+1}{n}}}$$

[Since all the n books have different colours]

[Since the book is picked uniformly at random]

(b) $$P(X_i = 1) = P\left(\begin{array}{l}\text{picking a book of different}\\ \text{colour, when } (i-1) \text{ distinct}\\ \text{colours have been picked}\end{array}\right) = \frac{n-i+1}{n} = p$$

$$P(X_i = k) = P\left(\begin{array}{l}\text{picking } (k-1) \text{ books}\\ \text{of colours among the}\\ (i-1) \text{ colours picked}\end{array}\right) \times P\left(\begin{array}{l}\text{picking a book of}\\ \text{different colour than}\\ (i-1) \text{ distinct colours}\end{array}\right)$$

$$= \left[P\left(\begin{array}{l}\text{picking a book of}\\ \text{colours among the}\\ (i-1) \text{ colours picked}\end{array}\right)\right]^{k-1} \times P\left(\begin{array}{l}\text{picking a book of}\\ \text{different colour than}\\ (i-1) \text{ distinct colours}\end{array}\right)$$

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

[∵ The events are independent.]

∴ X_i is a geometric random variable with parameter

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

(c) Let Z be a geometric random variable with parameter p. The probability mass function of X is given by

$$P(Z=k) = (1-p)^{k-1} p \quad \text{where } k = 1, 2, \dots$$

Expectation of Z:

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

→ This is an AGP (Arithmetic-Geometric progression.)

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

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

[Replacing k by $(k-1)$, and changing lower limit of sum to 2]

$$\Rightarrow E(Z) - (1-p)E(Z) = 1(1-p)^0 p + \sum_{k=2}^{\infty} (1-p)^{k-1} p$$

$$= p\left(\sum_{k=1}^{\infty} (1-p)^{k-1}\right)$$

$$\Rightarrow \quad pE(Z) = \frac{p}{1-(1-p)} = 1$$

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

[Infinite Geometric series with common ratio $(1-p)$ [where $0 < 1-p < 1$]

$$\sum_{k=1}^{\infty} ar^{k-1} = \frac{a}{1-r},$$

(iff $|r| < 1$)]

Variance of Z:

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

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

Subtracting both equations,

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

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

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

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

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

(d) $E(X^{(n)}) = E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$ [Linearity of expectation operator]

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

$\therefore$ $\boxed{E(X^{(n)}) = n \sum_{i=1}^{n} \frac{1}{i} = n H_n}$ [where H_n is the n^{th} Harmonic number (partial sum of harmonic series)]

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

$= \sum_{i=1}^{n} \mathrm{Var}(X_i)$ [Since, X_i's are independent random variables]

$= \sum_{i=1}^{n} \frac{1-p_i}{p_i^2}$ [where $p_i = \frac{n-i+1}{n}$, parameter of geometric distribution of X_i]

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

$= n^2 \sum_{i=1}^{n} \frac{1}{i^2} - n \sum_{i=1}^{n} \frac{1}{i}$ [$\because \sum_{i=1}^{n} \frac{1}{i^2} \leq \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6}$]

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

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

Claim: $\log(n+1) < \sum_{i=1}^{n} \frac{1}{i} < \log n + 1$

$\frac{1}{i} \leq \frac{1}{x} \quad \forall x \in [i-1, i] \Rightarrow \int_{i-1}^{i} \frac{1}{i} dx \leq \int_{i-1}^{i} \frac{1}{x} dx \Rightarrow \frac{1}{i} \leq \int_{i-1}^{i} \frac{1}{x} dx$

$\Rightarrow \sum_{i=2}^{n} \frac{1}{i} \leq \sum_{i=2}^{n} \int_{i-1}^{i} \frac{1}{x} dx$

$\Rightarrow \sum_{i=1}^{n} \frac{1}{i} \leq 1 + \sum_{i=1}^{n} \frac{1}{i} \leq 1 + \int_{1}^{n} \frac{1}{x} dx$

$\Rightarrow \sum_{i=1}^{n} \frac{1}{i} \leq 1 + \log n$

$\frac{1}{i} \geq \frac{1}{x} \quad \forall x \in [i, i+1] \Rightarrow \int_{i}^{i+1} \frac{1}{i} dx \geq \int_{i}^{i+1} \frac{1}{x} dx$

$\Rightarrow \sum_{i=1}^{n} \frac{1}{i} dx \geq \int_{1}^{n+1} \frac{1}{x} dx = \log(n+1)$

$\therefore \; n \log(n+1) \leq E(X^{(n)}) \leq n[\log n + 1]$

We can show that $E(X^{(n)}) = \Theta(n \log n)$.

[$\because$ $E(X^{(n)}) \leq n[\log n + 1] \leq 2n \log n \quad \forall n \geq 2$

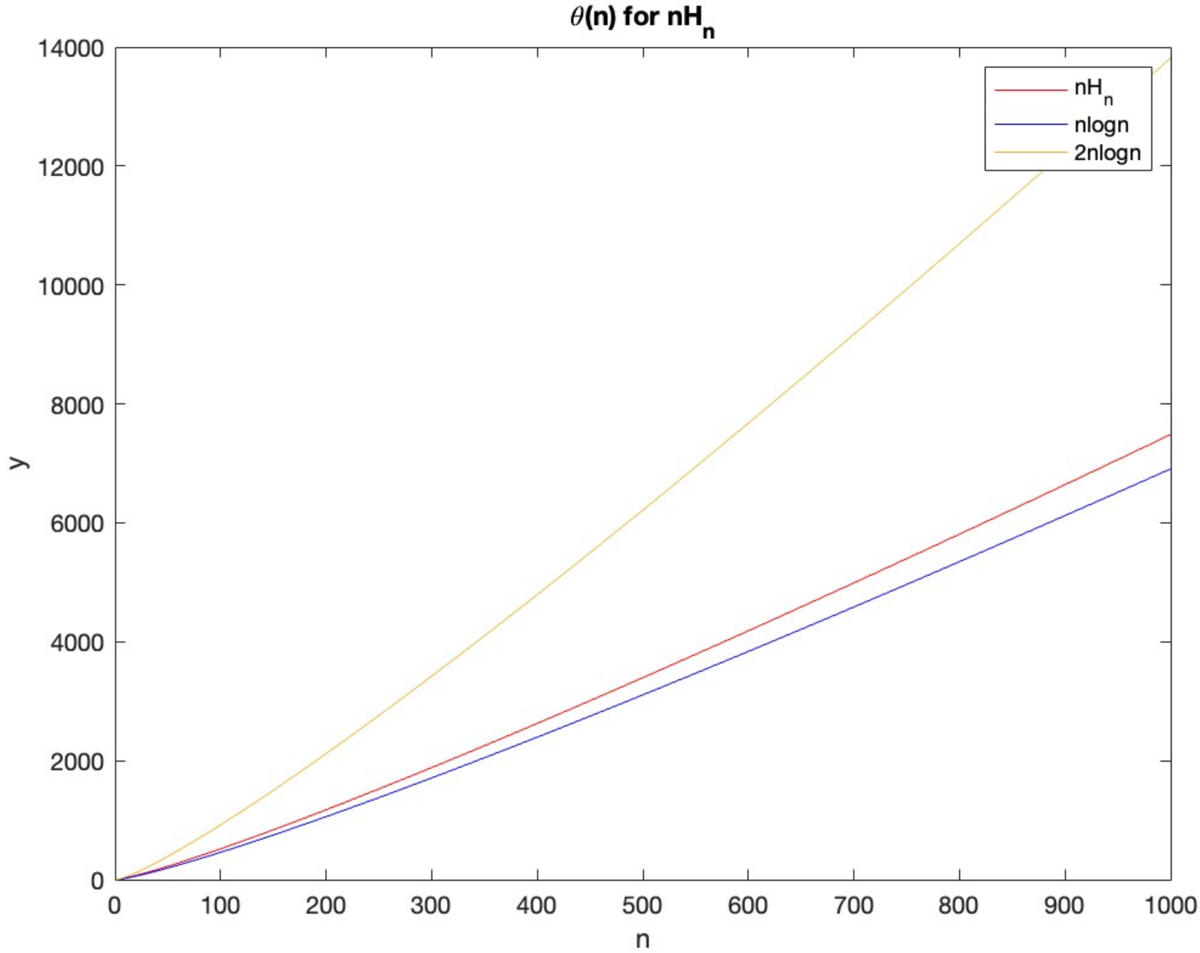
$E(X^{(n)}) \geq n \log(n+1) \geq n \log n \quad \forall n \geq 2$

$\therefore$ Taking $c_1 = 1$, $c_2 = 2$, $n_0 = 2$, we get

$c_1(n \log n) \leq E(X^{(n)}) \leq c_2(n \log n) \quad \forall n \geq n_0$

Hence, $E(X^{(n)}) = \Theta(n \log n)$ [by definition of Θ notation]]

$\therefore$ We can take $f(n) = n \log n$ or any function which is $\Theta(n \log n)$ so that $E(X^{(n)}) = \Theta(f(n))$.



θ(n) for nH_n
14000
12000
10000
8000
6000
4000
2000
0
y
0
100
200
300
400
500
600
700
800
900
1000
n
nH_n
nlogn
2nlogn

Question 2

(a) Let U be a random variable with a uniform distribution to $[0,1]$, i.e., its cumulative distribution function is given by

$$F_U(x) = \begin{cases} 0, & \text{if } x < 0 \\ x, & \text{if } 0 \leq x \leq 1 \\ 1, & \text{if } x > 1 \end{cases} \qquad \left[\because f_U(x) = \begin{cases} 1, & \text{if } x \in (0,1) \\ 0, & \text{otherwise} \end{cases}\right]$$

Consider the random variable $V = F^{-1}(U)$, i.e., $V(\omega) = F^{-1}(U(\omega))$.
The cdf of the random variable V is given by

$$F_V(y) = P(V \leq y) = P(F^{-1}(U) \leq y)$$

$$= P(U \leq F(y))$$ [Since F is invertible and is a distribution function, and hence, is strictly increasing]

$$= F_U(F(y))$$ [By definition of cdf]

$$= \underline{\underline{F(y)}}$$ [$\because$ Since U is a uniform random variable to $[0,1]$ and $0 \leq F(y) \leq 1$, since $F(y)$ is a distribution function.]

$\therefore$ The cdf of the random variable $V = F^{-1}(U)$ is F.
Since the values $\{v_i\}_{i=1}^{n}$ are independent random samples of the random variable V, they follow the distribution F.

(b) $Y_1, Y_2, \dots, Y_n \rightarrow$ identically distributed random variables with cdf F

$U_1, U_2, \dots, U_n \rightarrow$ identically distributed random variables with Uniform $[0,1]$ distribution

$$P(E \geqslant d) = P\left(\max_{0 \leq y \leq 1} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(U_i \leq y)}{n} - y \right| \geqslant d\right) \quad -①$$

Consider the random variable $V_i = F^{-1}(U_i)$, assuming F to be an invertible function, and hence a strictly increasing function.
Then, the random variable $\mathbb{1}(U_i \leq y)$ is same as
$\mathbb{1}(F^{-1}(U_i) \geqslant F^{-1}(y)) = \mathbb{1}(V_i \leq F^{-1}(y_i))$ since F is strictly increasing.

From part (a), we know that

$$\left.\begin{array}{l} V_i = F^{-1}(U_i) \\ U_i \sim \text{Uniform}[0,1] \\ F \text{ is invertible.} \end{array}\right\} \Rightarrow V \text{ has a cdf } F.$$

Since F is bijective,

$\forall\, y \in [0,1]$, $\exists\, x \in \mathbb{R}$ such that $F(x) = y$ and $F^{-1}(y) = x$

Taking the x corresponding to y, $\mathbb{1}(U_i \leq y) = \mathbb{1}(V_i \leq x)$

From ①,

$$\therefore\ P(E \geq d) = P\left(\max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(V_i \leq x)}{n} - F(x) \right| \leq d\right)$$

Now, since both V_i and Y_i are identically distributed and independent random variables with cdf F, for $i = 1, 2, \dots, n$, the random vectors $(V_1, V_2, \dots, V_n)$ and $(Y_1, Y_2, \dots, Y_n)$ are identically distributed.

Note that both $E = \max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(V_i \leq x)}{n} - F(x) \right|$ and

$D = \max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(Y_i \leq x)}{n} - F(x) \right|$ are random variables that are obtaining on applying the same function on the random vectors $(V_1, V_2, \dots, V_n)$ and $(Y_1, Y_2, \dots, Y_n)$. Since the random vectors are identically distributed, E and D are also identically distributed.

$$\therefore\ P(E \geq d) = P\left(\max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(V_i \leq x)}{n} - F(x) \right| \geq d\right)$$

$$= P\left(\max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(Y_i \leq x)}{n} - F(x) \right| \geq d\right) = P(D \geq d) \quad \forall\, d \in \mathbb{R}$$

$$\therefore\ \boxed{P(E \geq d) = P(D \geq d) \quad \forall\, d \in \mathbb{R}}$$

NOTE: If it is not given that Y_i's are independent and U_i's are independent, then the given statement may not be true.

eg: If $F(x)$ is the CDF of Gaussian distribution, then $V_i = F^{-1}(U_i)$ and Y_i are both normal variables with cdf F. Suppose U_i are independent (which implies $V_i = F^{-1}(U_i)$ are independent), and Y_i are dependent.

For example,

$$V_1, V_2 \sim N(\mu, \sigma^2) \rightarrow \text{independent}$$
$$Y_1 = Y_2 \sim N(\mu, \sigma^2) \rightarrow \text{dependent}$$

$$\mathbb{1}(V_i \leq x) \sim \text{Bernoulli}(F(x)) \quad \text{for } i=1,2$$

$$\mathbb{1}(V_1 \leq x) + \mathbb{1}(V_2 \leq x) \sim \text{Binomial}(2, F(x))$$

[Sum of n independent Bernoulli variables with parameter p is a Binomial r.v. with parameters n and p.]

i.e.,
$$P(\mathbb{1}(V_1 \leq x) + \mathbb{1}(V_2 \leq x) = 0) = (1-F(x))^2$$
$$P(\mathbb{1}(V_1 \leq x) + \mathbb{1}(V_2 \leq x) = 1) = 2F(x)[1-F(x)]$$
$$P(\mathbb{1}(V_1 \leq x) + \mathbb{1}(V_2 \leq x) = 2) = (F(x))^2$$

$$\mathbb{1}(Y_i \leq x) \sim \text{Bernoulli}(F(x)), \quad \text{for } i=1,2$$

$$\mathbb{1}(Y_1 \leq x) + \mathbb{1}(Y_2 \leq x) = 2 \times \mathbb{1}(Y \leq x)$$ [$\because$ we have taken $Y_1 = Y_2 = Y$]

i.e.,
$$P(2 \times \mathbb{1}(Y \leq x) = 0) = (1-F(x))$$
$$P(2 \times \mathbb{1}(Y \leq x) = 1) = 0$$
$$P(2 \times \mathbb{1}(Y \leq x) = 2) = F(x)$$

Hence, the distributions of $E = \max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(V_i \leq x)}{n} - F(x) \right|$

and $D = \max_{x \in \mathbb{R}} \left| \frac{\sum_{i=1}^{n} \mathbb{1}(Y_i \leq x)}{n} - F(x) \right|$ can be different,

if the Y_i's and U_i's are not both independent, as shown by this this example.

Question 3

(a) Let the coordinates of the n given points be $\{(x_i, y_i, z_i)\}_{i=1}^n$, and let the equation of the plane be $z = ax + by + c$

$z_i = ax_i + by_i + c + \varepsilon_i$ for $i = 1, 2, \dots, n$ where $\varepsilon_i \sim N(0, \sigma^2)$

[Since z_i is corrupted independently by noise ε_i from $N(0, \sigma^2)$]

The likelihood function for $\{(x_i, y_i, z_i)\}_{i=1}^n$ is given by

Likelihood function for $\{(x_i, y_i, z_i)\}_{i=1}^n$ $= p_2(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$ [where p_2 is the joint pmf for the random vector $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)$]

$$= \prod_{i=1}^n p(\varepsilon_i)$$

[since ε_i's are independent, where $p(x)$ is the pdf of the gaussian distribution of mean 0 and variance σ^2

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$]

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n e^{\frac{-\sum \varepsilon_i^2}{2\sigma^2}}$$

$\Rightarrow$ Log-likelihood function of $\{(x_i, y_i, z_i)\}_{i=1}^n$ $= \mathcal{L}(a,b,c) = -n\log(\sigma\sqrt{2\pi}) - \frac{\sum \varepsilon_i^2}{2\sigma^2}$

$$\boxed{\mathcal{L}(a,b,c) = -n\log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^n (z_i - ax_i - by_i - c)^2}$$

we have to find a, b, c for which $\mathcal{L}(a,b,c)$ is maximised.

$$\frac{\partial \mathcal{L}}{\partial a} = 0 \Rightarrow \sum_{i=1}^n x_i(z_i - ax_i - by_i - c) = 0$$

$$\Rightarrow a\sum_{i=1}^n x_i^2 + b\sum_{i=1}^n x_i y_i + c\sum_{i=1}^n x_i = \sum_{i=1}^n x_i z_i \quad -①$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^n y_i(z_i - ax_i - by_i - c) = 0$$

$$\Rightarrow a\sum_{i=1}^n x_i y_i + b\sum_{i=1}^n y_i^2 + c\sum_{i=1}^n y_i = \sum_{i=1}^n y_i z_i \quad -②$$

$$\frac{\partial \mathcal{L}}{\partial c} = 0 \Rightarrow \sum_{i=1}^n (z_i - ax_i - by_i - c) = 0$$

$$\Rightarrow a\sum_{i=1}^n x_i + b\sum_{i=1}^n y_i + cn = \sum_{i=1}^n z_i \quad -③$$

Vector form:

$$\vec{x}\cdot(\vec{z}-a\vec{x}-b\vec{y}-c\vec{1})=0 \quad -①$$
$$\vec{y}\cdot(\vec{z}-a\vec{x}-b\vec{y}-c\vec{1})=0 \quad -②$$
$$\vec{1}\cdot(\vec{z}-a\vec{x}-b\vec{y}-c\vec{1})=0 \quad -③$$

where
$\vec{x}=(x_1,x_2,\dots,x_n)$
$\vec{y}=(y_1,y_2,\dots,y_n)$
$\vec{1}=(1,1,\dots,1)$

Matrix form:

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \\ 1 & 1 & \cdots & 1 \end{bmatrix}\left(\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}-\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}\right)=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

which is same as

$$X^T(Z-XA)=0 \iff \boxed{X^TXA=X^TZ}$$ where $X=\begin{bmatrix} x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix}$, $A=\begin{bmatrix} a \\ b \\ c \end{bmatrix}$, $Z=\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$

$$\begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_iy_i & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_iy_i & \sum_{i=1}^n y_i^2 & \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n y_i & n \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^n x_iz_i \\ \sum_{i=1}^n y_iz_i \\ \sum_{i=1}^n z_i \end{bmatrix}$$

(b) Repeating the same steps as in part (a),

$$\varepsilon_i = z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6$$

Likelihood function for $\{(x_i,y_i,z_i)\}_{i=1}^n$ $= \prod_{i=1}^n p(\varepsilon_i) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left(-\frac{\sum_{i=1}^n \varepsilon_i^2}{2\sigma^2}\right)$

Log-likelihood function for $\{(x_i,y_i,z_i)\}_{i=1}^n$ $= \mathcal{L}(\{a_i\}_{i=1}^n) = -n\log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^n \varepsilon_i^2$

$$\boxed{\mathcal{L}(\{a_i\}_{i=1}^n) = -n\log(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6)^2}$$

We have to find $a_1, a_2, a_3, a_4, a_5, a_6$ so that $\mathcal{L}(\{a_i\}_{i=1}^n)$ is maximised.

$$\frac{\partial \mathcal{L}}{\partial a_1} = 0 \Rightarrow \sum_{i=1}^{n} x_i^2(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \quad \text{— ①}$$

$$\frac{\partial \mathcal{L}}{\partial a_2} = 0 \Rightarrow \sum_{i=1}^{n} y_i^2(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \quad \text{— ②}$$

$$\frac{\partial \mathcal{L}}{\partial a_3} = 0 \Rightarrow \sum_{i=1}^{n} x_iy_i(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \quad \text{— ③}$$

$$\frac{\partial \mathcal{L}}{\partial a_4} = 0 \Rightarrow \sum_{i=1}^{n} x_i(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \quad \text{— ④}$$

$$\frac{\partial \mathcal{L}}{\partial a_5} = 0 \Rightarrow \sum_{i=1}^{n} y_i(z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \quad \text{— ⑤}$$

$$\frac{\partial \mathcal{L}}{\partial a_6} = 0 \Rightarrow \sum_{i=1}^{n} (z_i - a_1x_i^2 - a_2y_i^2 - a_3x_iy_i - a_4x_i - a_5y_i - a_6) = 0 \quad \text{— ⑥}$$

<u>vector form:</u>

$$\overrightarrow{x^2}\cdot(\vec{z} - a_1\overrightarrow{x^2} - a_2\overrightarrow{y^2} - a_3\overrightarrow{xy} - a_4\vec{x} - a_5\vec{y} - a_6\vec{1}) = 0 \quad \text{— ①}$$

$$\overrightarrow{y^2}\cdot(\vec{z} - a_1\overrightarrow{x^2} - a_2\overrightarrow{y^2} - a_3\overrightarrow{xy} - a_4\vec{x} - a_5\vec{y} - a_6\vec{1}) = 0 \quad \text{— ②}$$

$$\overrightarrow{xy}\cdot(\vec{z} - a_1\overrightarrow{x^2} - a_2\overrightarrow{y^2} - a_3\overrightarrow{xy} - a_4\vec{x} - a_5\vec{y} - a_6\vec{1}) = 0 \quad \text{— ③}$$

$$\vec{x}\cdot(\vec{z} - a_1\overrightarrow{x^2} - a_2\overrightarrow{y^2} - a_3\overrightarrow{xy} - a_4\vec{x} - a_5\vec{y} - a_6\vec{1}) = 0 \quad \text{— ④}$$

$$\vec{y}\cdot(\vec{z} - a_1\overrightarrow{x^2} - a_2\overrightarrow{y^2} - a_3\overrightarrow{xy} - a_4\vec{x} - a_5\vec{y} - a_6\vec{1}) = 0 \quad \text{— ⑤}$$

$$\vec{1}\cdot(\vec{z} - a_1\overrightarrow{x^2} - a_2\overrightarrow{y^2} - a_3\overrightarrow{xy} - a_4\vec{x} - a_5\vec{y} - a_6\vec{1}) = 0 \quad \text{— ⑥}$$

$$\left[\begin{array}{lll} \text{where} & \vec{x} = (x_1, x_2, \cdots, x_n), & \overrightarrow{x^2} = (x_1^2, x_2^2, \cdots, x_n^2) \\ & \vec{y} = (y_1, y_2, \cdots, y_n), & \overrightarrow{y^2} = (y_1^2, y_2^2, \cdots, y_n^2) \\ & \vec{z} = (z_1, z_2, \cdots, z_n), & \overrightarrow{xy} = (x_1y_1, x_2y_2, \cdots, x_ny_n) \end{array}\right]$$

Matrix form

$$X^T(Z - XA) = 0$$

$$\Updownarrow$$

$$\boxed{X^T X A = X^T Z}$$

$$\left[\text{where } X = \begin{bmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 & 1 \\ x_2^2 & y_2^2 & x_2y_2 & x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n^2 & y_n^2 & x_ny_n & x_n & y_n & 1 \end{bmatrix}, \quad A = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_6 \end{bmatrix}, \quad Z = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}\right]$$

$$\begin{bmatrix} \sum x_i^4 & \sum x_i^2y_i^2 & \sum x_i^3y_i & \sum x_i^3 & \sum x_i^2y_i & \sum x_i^2 \\ \sum x_i^2y_i^2 & \sum y_i^4 & \sum x_iy_i^3 & \sum x_iy_i^2 & \sum y_i^3 & \sum y_i^2 \\ \sum x_i^3y_i & \sum x_iy_i^3 & \sum x_i^2y_i^2 & \sum x_i^2y_i & \sum x_iy_i^2 & \sum x_iy_i \\ \sum x_i^3 & \sum x_iy_i^2 & \sum x_i^2y_i & \sum x_i^2 & \sum x_iy_i & \sum x_i \\ \sum x_i^2y_i & \sum y_i^3 & \sum x_iy_i^2 & \sum x_iy_i & \sum y_i^2 & \sum y_i \\ \sum x_i^2 & \sum y_i^2 & \sum x_iy_i & \sum x_i & \sum y_i & n \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix} = \begin{bmatrix} \sum x_i^2z_i \\ \sum y_i^2z_i \\ \sum x_iy_iz_i \\ \sum x_iz_i \\ \sum y_iz_i \\ \sum z_i \end{bmatrix}$$

[where i goes from 1 to n in the sums]

(c) MATLAB code conclusions

The values of a, b, c obtained from the MATLAB code are $a = 10.0022, b = 19.9980, c = 29.9516$.

Thus, the predicted equation of the plane is $z = 10.0022\ x + 19.9980\ y + 29.9516$

The predicted noise variance is $\sigma^2_{predicted} = 23.0570$.

Question 4

(e)

WHAT HAPPENS AND WHY?
Cross-validation helps assess whether the estimator is overfitting about the points of T and not accurately predicting the distribution. However, this doesn't indicate that the estimator will estimate the other data accurately. Therefore, the ability to detect overfitting is compromised.

If validation and training sets become equal, the cross-validation procedure used in the question is no longer reliable. This is so because the expression $p_n(x_i;\sigma)$ now contains a term $\dfrac{e^{-(x_i-x_i^2)/(2\sigma^2)}}{(n\sigma\sqrt{2\pi})}$ which shoots up the expression for small values of σ, for instance the order of the expression for $\sigma = 10^{-3}$ is 10^6. This would evidently lead to misleading results. One can check the fact by running *q4_additional.m*. The primary purpose of cross-validation is to avoid issues like overfitting (where a model performs well on the training data but poorly on new data) and to provide a more reliable estimate of a model's performance.

```
clear;
clc;
rng("default");
% Parameters
mean_value = 0;
variance_value = 16;
num_points = 1000;

% Generate random data points following N(0, 1)
random_data = randn(1, num_points);

% Scale and shift to N(0, 16)
scaled_data = sqrt(variance_value) * random_data;

T = scaled_data(1:750);
V = scaled_data(751:1000);
V = T;
N = 750;
m = (V*V');
LL = zeros(1,11);
V = T;
N = 750;
x = -8:0.1:8;

sigma = [0.001, 0.1, 0.2, 0.9, 1, 2, 3, 5, 10, 20, 100];

% The term when p_n_x(V(i), x, sigma(j)) will overshoot when V(i) = x which
% is will happen if we set V = T.
for i = 1:11
    LL(i) = 1;
    for j = 1:N
        LL(i) = LL(i)+log(p_n_x(V(j), T, sigma(i)));
    end
end

[maxima, index] = max(LL);
sigma_best = sigma(index)

% This is the code in which we are not evaluating p_n_x if V(i) = x. Here
% the modified function p_n_updated_x is used so it is providing values of
% sigma closer to our expectation so the modified cross-validation
% procedure is working.

for i = 1:11
    LL(i) = 1;
    for j = 1:N
        LL(i) = LL(i)+log(p_n_updated_x(V(j), T, sigma(i)));
    end
end

```

```
[maxima, index] = max(LL);
sigma_best = sigma(index)

function p_n = p_n_x(x, V, sigma_best)
    arr = exp(-(x - V).*(x - V)/(2*sigma_best*sigma_best));
    p_n = sum(arr(:))/(size(V,2)*sigma_best*sqrt(2*pi));
end

function p_n = p_n_updated_x(x, V, sigma_best)
    arr = exp(-(x - V).*(x - V)/(2*sigma_best*sigma_best));
    p_n = (sum(arr(:)) - 1)/(size(V,2)*sigma_best*sqrt(2*pi));
end
```

Question 5

Consider n independent random variables $X_1, X_2, \dots, X_n$ such that for every i, X_i always lies in $[a_i, b_i]$ where $a_i < b_i$.

Let $S_n = \sum_{i=1}^{n} X_i$

$$P(S_n - E(S_n) > t) = P\left(e^{s[S_n - E(S_n)]} > e^{st}\right)$$

[Since e^x is a strictly increasing function, i.e., $x > y \Leftrightarrow e^x > e^y$]

$$\leqslant \frac{E\left[e^{s(S_n - E(S_n))}\right]}{e^{st}}$$

[By Markov's inequality, $P(X > a) \leqslant \frac{E(X)}{a}$, if $a > 0$ and X is a non-negative random variable.

Here, $e^x > 0 \ \forall x \in \mathbb{R}$. So, $X = e^{s[S_n - E(S_n)]}$ is a non-negative random variable, and $a = e^{st} > 0$]

$$\Rightarrow \boxed{P(S_n - E(S_n) > t) \leqslant e^{\frac{s^2(b^2 - a^2)}{8} - st} \quad \forall s > 0}$$

[where $b = \sum_{i=1}^{n} b_i$, $a = \sum_{i=1}^{n} a_i$]

[Note that $a_i \leqslant X_i(\omega_i) \leqslant b_i$ for $i = 1, 2, \dots, n$

$\Rightarrow a = \sum_{i=1}^{n} a_i \leqslant S_n(\vec{\omega}) \leqslant \sum_{i=1}^{n} b_i = b$

$\Rightarrow$ S_n always lie in $[a, b]$. So, we can apply the Intermediate Result (IR) on the random variable S_n.]

For tighter bound on S_n, we have to minimise $f(s) = \frac{s^2(b^2 - a^2)}{8} - st$ (since e^x is an increasing function.)

$$f(s) = \frac{(b^2 - a^2)}{8}\left[s^2 - \frac{8st}{(b^2 - a^2)} + \left(\frac{4t}{b^2 - a^2}\right)^2 - \left(\frac{4t}{b^2 - a^2}\right)^2\right]$$

$$= \frac{(b^2 - a^2)}{8}\left(s - \frac{4t}{b^2 - a^2}\right)^2 - \frac{2t^2}{b^2 - a^2} \geqslant \frac{-2t^2}{b^2 - a^2}$$

$f(s)$ is minimum at $s = s_0 = \frac{4t}{b^2 - a^2}$, and $f(s_0) = \frac{-2t^2}{b^2 - a^2}$ //

$\therefore$ The minimum value of $e^{f(s)}$ is $e^{-2t^2/(b^2-a^2)}$. So,

$$\boxed{P(S_n - E(S_n) > t) \leqslant e^{-2t^2/(b^2-a^2)} \quad \text{where } a = \sum_{i=1}^{n} a_i,\ b = \sum_{i=1}^{n} b_i}$$

Now, we will prove IR.

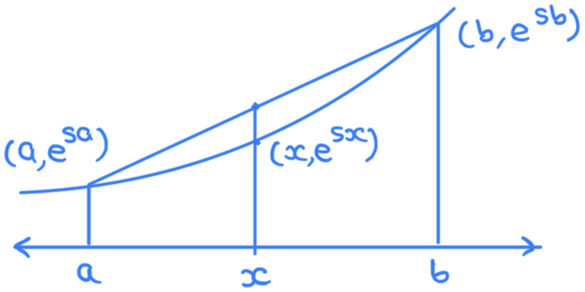
(a) Without loss of generality, take $E(X) = 0$, and hence $a \leqslant 0 \leqslant b$.

Since e^{sx} is a convex function,

$$e^{s(\lambda a + (1-\lambda)b)} \leqslant \lambda e^{sa} + (1-\lambda)e^{sb} \qquad \forall\ \lambda \in [0,1]$$

Taking $x = \lambda a + (1-\lambda)b$, we get $\lambda = \frac{b-x}{b-a}$. Substituting λ in terms of x, we get

$$e^{sx} \leqslant \left(\frac{b-x}{b-a}\right)e^{sa} + \left(\frac{x-a}{b-a}\right)e^{sb}$$



(b) Taking expectation of both sides,

$$E(e^{sX}) \leqslant \frac{be^{sa} - ae^{sb}}{b-a} + \frac{e^{sb} - e^{sa}}{b-a}E(X) = \frac{be^{sa} - ae^{sb}}{b-a}$$

$(\because E(X) = 0)$

$$\frac{be^{sa} - ae^{sb}}{b-a} = e^{sa}\left(\frac{b - ae^{s(b-a)}}{b-a}\right)$$

$$= e^{sa}\left(1 + \frac{a}{b-a}\left(1 - e^{s(b-a)}\right)\right)$$

$$= e^{\left[\frac{s(b-a)a}{(b-a)} + \log\left(1 + \frac{a}{b-a}\left(1 - e^{s(b-a)}\right)\right)\right]}$$

$$= e^{L(s(b-a))}$$

$\left[\text{where } L(h) = \frac{ha}{b-a} + \log\left(1 + \frac{a(1-e^h)}{b-a}\right)\right]$

$$\therefore\ E(e^{sX}) \leqslant e^{L(s(b-a))}$$

(c) $L(h) = \frac{ha}{b-a} + \log\left(1 + \frac{a(1-e^h)}{b-a}\right)$

$$L'(h) = \frac{a}{b-a} + \frac{1}{1+\frac{a(1-e^h)}{b-a}} \times \frac{-ae^h}{b-a} = \frac{a}{b-a} - \underline{\underline{\frac{ae^h}{b-ae^h}}}$$

$$L''(h) = \frac{ae^h}{ae^h-b} - \frac{ae^h \times ae^h}{(ae^h-b)^2} = \frac{(-ae^h)b}{(b-ae^h)^2} \leq \underline{\underline{\frac{1}{4}}} \quad \forall h \in \mathbb{R}$$

[Taking $x = b > 0$ and $y = -ae^h > 0$,

$\frac{(x+y)}{2} \geq \sqrt{xy} \Rightarrow \frac{\sqrt{xy}}{(x+y)} \leq \frac{1}{2} \Rightarrow \frac{xy}{(x+y)^2} \leq \frac{1}{4}$ (since $x, y > 0$)

$\Rightarrow \underline{\underline{\frac{(-ae^h)b}{(b-ae^h)^2} \leq \frac{1}{4}}}$]

(d) $L''(h) \leq \frac{1}{4} \quad \forall h \in \mathbb{R}$

$\Rightarrow \int_0^h L''(x)\,dx \leq \int_0^h \frac{1}{4}\,dx \quad \forall h \geq 0 \Rightarrow L'(h) - L'(0) \leq \frac{h}{4} \quad \forall h \geq 0$

$\Rightarrow \underline{\underline{L'(h) \leq \frac{h}{4}}} \quad \forall h \geq 0$ [$\because L'(0) = \frac{a}{b-a} - \frac{a}{b-a} = \underline{\underline{0}}$]

$\Rightarrow \int_0^h L'(x)\,dx \leq \int_0^h \frac{x}{4}\,dx \quad \forall h \geq 0 \Rightarrow L(h) - L(0) \leq \frac{h^2}{8} \quad \forall h \geq 0$

$\Rightarrow \underline{\underline{L(h) \leq \frac{h^2}{8}}} \quad \forall h \geq 0$ [$\because L(0) = 0 + \log(1) = \underline{\underline{0}}$]

$$\therefore \boxed{E\left[e^{s(X-E(X))}\right] \leq e^{L(s(b-a))} \leq e^{s^2(b-a)^2/8} \quad \forall s > 0}$$

Hence, IR is proved.

Attached MATLAB code files

1. q1.m
2. q3.m
3. q4.m