CS215 HomeWork Assignment - 3

Team members :-

- Kavya Gupta (22B1053)
- Aryan Katiyar (22B1205)
- Arvind Kumar Yadav (22B1025)

Honor Code : No scope of plagiarism and external help from other teams sought, all work distributed and worked on properly.

Solutions to all questions present below.
(Q1 to Q5 in order)

Note : In the case of Q3(c) and Q4, the very first section tells about the location of the code file and output pictures.
For Q1 Part(f), the picture of the plot has been saved as Q1(f).png and also has been included in the report.

22B1053 - 22B1205 - 221025

PAGE NO

DATE / /

Q1

(a) X_1: denotes number of additional books such we move from having picked 0 distinct to 1 distinct colors.

hence X_1: denotes picking any book

$\therefore X_1 = 1$.

When books with $i-1$ distinct colors have been collected, then number option for next book to be of different color

$= n+1-i$

$\therefore P$ (picking different color book after $i-1$ already) $= \frac{n+1-i}{n}$

(b) We calculate PMF of X_i i.e. $P(X_i = r)$. $X_i = r$, means that after we had $i-1$ different color books, it takes r pickups to have i different color i.e. at r^{th} pick we get different color, and at remaining $r-1$ pickups we get any color picked already in those $i-1$ color ($\because$ $i-1$ option for each option).

$$\therefore P(X_i = r) = \frac{\text{considered options}}{\text{Total options}} = \frac{(i-1)^{r-1}(n+1-i)}{n^r}$$

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

$$\therefore \text{parameter } p = \frac{n+1-i}{n}$$

(c) Let Z be Geometric R.V i.e. $P(Z=k) = p(1-p)^{k-1}$, $k = 1, 2 \ldots$

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

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

Subtracting ② from ①,

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

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

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

$$E(Z^2) = \sum_{k=1}^{\infty} p(1-p)^{k-1}k^2 = p + \sum_{k=2}^{\infty} p(1-p)^{k-1}k^2 \quad - (4)$$

$$(1-p)E(Z^2) = \sum_{k=2}^{\infty} p(1-p)^{k-1}(k-1)^2 \quad - (3)$$

Subtracting (3) from (4)

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

$$E(Z^2) = 1 + \sum_{k=2}^{\infty} \cancel{p}(1-p)^{k-1}(2k-1) = 1 + 3\cancel{p}(1-p) + \sum_{k=3}^{\infty} \cancel{p}(1-p)^{k-1}(2k-1) \quad - (5)$$

$$(1-p)E(Z^2) = 1-p + \sum_{k=3}^{\infty} \cancel{p}(1-p)^{k-1}(2k-3) \quad - (6)$$

Subtracting (6) from (5)

$$pE(Z^2) = p + 3\cancel{p}(1-p) + \sum_{k=3}^{\infty} 2\cancel{p}(1-p)^{k-1}$$

$$E(Z^2) = \frac{1+3(1-p)}{p} + \sum_{k=3}^{\infty} \frac{2(1-p)^{k-1}}{p} = \frac{1+3(1-p)}{p} + \frac{2(1-p)^2}{p^2}$$

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

$$= \frac{2}{p^2} + 2 + 1 - 3 + \frac{3}{p} - \frac{4}{p}$$

$$= \frac{2}{p^2} - \frac{1}{p}$$

$$\therefore \operatorname{Var}(Z) = E(Z^2) - E^2(Z)$$

$$\operatorname{Var}(Z) = \frac{1}{p^2} - \frac{1}{p}.$$

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

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

$$= \sum_{i=1}^{n} \frac{n}{n+1-i} = n\left(1 + \tfrac{1}{2} + \tfrac{1}{3} + \dots + \tfrac{1}{n}\right)$$

PAGE No.
DATE / /

(e) $\because$ X_i deals with things after X_{i-1} and before X_{i+1} had happened $\therefore$ There X_{i-1}, X_i, X_{i+1} are indepent, and by induction over i, we claim ~~X_i, X_j~~ X_i's are independent R.V.

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

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

$$= n^2\left(1+\frac{1}{2^2}+\frac{1}{3^2}\cdots\frac{1}{n^2}\right) - n\left(1+\tfrac{1}{2}+\tfrac{1}{3}\cdots+\tfrac{1}{n}\right)$$

$$Var(X^{(n)}) \leq n^2\left(1+\tfrac{1}{2^2}+\tfrac{1}{3^2}\cdots\right) - n\left(1+\tfrac{1}{2}\cdots+\tfrac{1}{n}\right)$$

$$= \frac{n^2\pi^2}{6} - n\left(1+\tfrac{1}{2}\cdots+\tfrac{1}{n}\right)$$

$$Var(X^{(n)}) \leq \frac{n^2\pi^2}{6}$$

(f)
$$1+\frac{1}{2}+\frac{1}{3}\cdots\frac{1}{n} \leq 1+\frac{1}{2}+\frac{1}{3}\cdots\frac{1}{2^m} \leq 1+1+\left(\frac{1}{2}+\frac{1}{2}\right)+\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right)\cdots\left(\frac{1}{2^{m-1}}+\frac{1}{2^{m-1}}\cdots\frac{1}{2^{m-1}}\right)$$

$$\leq 1+1+1+1\cdots+1$$
$$= m$$
$$\leq m+1$$

where $2^m > n \geq 2^{m-1}$ or $m > \log_2 n \geq m-1$

$$\therefore \quad 1+\frac{1}{2}+\frac{1}{3}\cdots\frac{1}{n} \leq \log_2 n+2$$

using same m,

$$1+\frac{1}{2}\cdots\frac{1}{n} \geq 1+\frac{1}{2}+\frac{1}{3}\cdots\frac{1}{2^{m-1}} \geq \left(\frac{1}{2}+\frac{1}{2}\right)+\left(\frac{1}{4}+\frac{1}{4}+\frac{1}{4}+\frac{1}{4}\right)\cdots+\left(\frac{1}{2^{m-1}}+\frac{1}{2^{m-1}}\cdots\frac{1}{2^{m-1}}\right)$$

$$= 1+1+\cdots 1$$
$$= m-1 \geq \log_2 n - 1$$

$$\therefore \log_2 n - 1 < 1+\tfrac{1}{2}\cdots\frac{1}{n} \leq \log_2 n + 2$$

PAGE No.
DATE / /

$\therefore$ for $n > 2$ we have

$$2\log_2 n > \log_2 n + 2$$

or

$$\log_2 n - 1 > \frac{\log_2 n}{2}$$

$\therefore$

$$\frac{\log_2 n}{2} < 1 + \frac{1}{2} + \frac{1}{3} \cdots \frac{1}{n} < 2\log_2 n \quad \text{for } n > 2$$

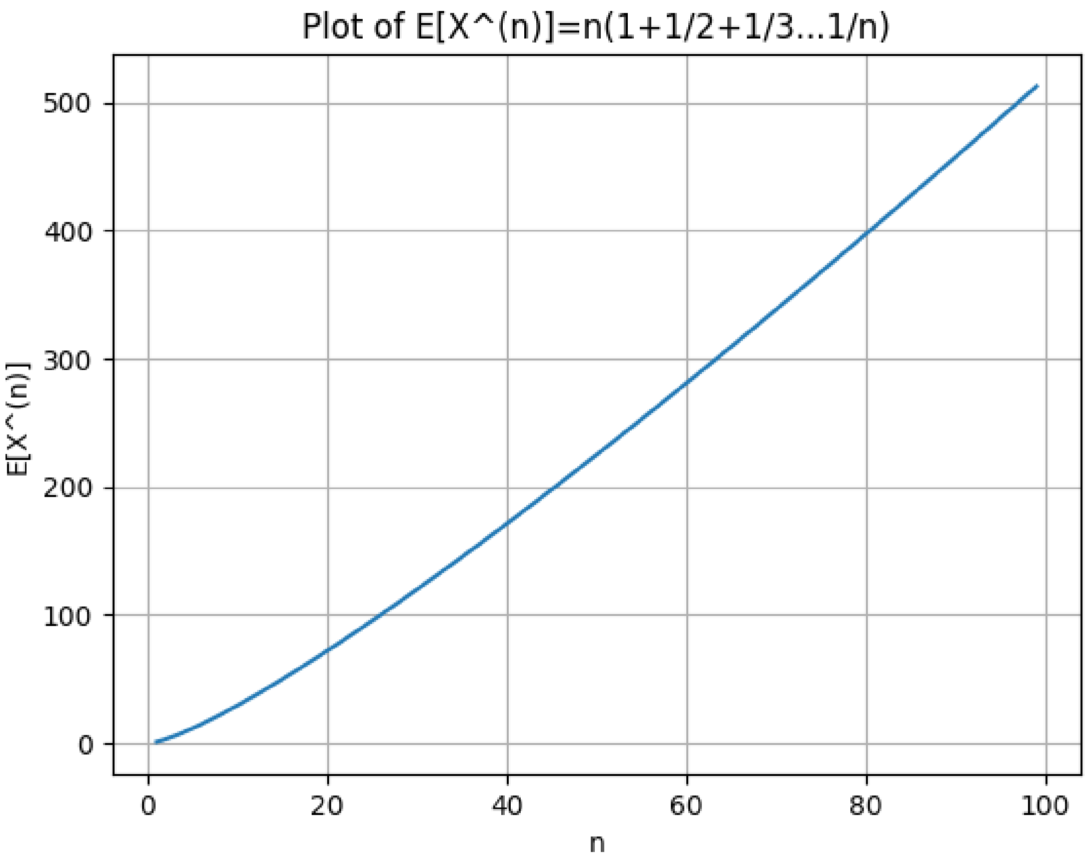
$\therefore$ as $n > 0$

$$\frac{n\log_2 n}{2} < \left(1 + \frac{1}{2} \cdots \frac{1}{n}\right)n < 2(\log_2 n) \times n \quad \text{for } n > 2$$

$\therefore$ $E(x^{(n)}) \in \Theta(n\log_2 n)$.

$\therefore$ $f(n) = n\log_2 n$

Plot for Q1 Part (f) (Homework Assignment-3)



Question (2)

(a) We have $v_i = F^{-1}(u_i)$ for some F^{-1}.

$V = F^{-1}(U)$, V is the random variable generating values $\{v_i\}_{i=1}^{n}$

F^{-1} exists hence F is bijective function — one-to-one onto.

Property of Uniform $(0,1)$ U, for $0 \leq p \leq 1$, $p = P(U \leq p)$

Hence consider $F(y) = P(U \leq F(y))$, $F(y) \in [0,1]$

as $F(y)$ represents probability

Since F^{-1} is bijective,

$$U \leq F(y) \iff F^{-1}(U) \leq F^{-1}(F(y))$$
$$\iff F^{-1}(U) \leq y \iff V \leq y$$

hence $F(y) = P(V \leq y)$
$$= F_V(y)$$

$\Rightarrow F_V = F$ or distribution of V is same as that of F.

(b) Before Proceedingly, let's note that
$0 \le F \le 1$ (covers all values between them)
as $F(y) = P((Y) \le y)$
and F is an increasing function (can be constant at parts also)
as $F(y) = \int_{-\infty}^{y} f_Y(y)\,dy$ and $\underline{f_Y(y) \geq 0}$.

Now for an increasing function f
$$a \le b \Leftrightarrow f(a) \le f(b) \text{ \& hence } \mathbb{1}(a \le b) = \mathbb{1}(f(a) \le f(b))$$
as $a \le b$ & $f(a) \le f(b)$ will be both true or both false at the same time.
and hence $P(Y \le y) = P(F(Y) \le F(y)) = F(y)$
Now $F(y) = z$, $z \in [0,1]$ then
$P(F(Y) \le z) = z \;\forall\; z \in [0,1]$.
This is same for a uniform$(0,1)$ r.v.
$P(U \le z) = z \;\forall\; z \in [0,1]$
hence $F(Y)$ and ~~the~~ Unif$(0,1)$ have same CDFs (distribution) and are equivalent.

We will use this fact later in steps.

So now we have

$$P(D \geq d) = P\left(\max_x \left| \frac{\sum_{i=1}^{n} 1(Y_i \leq x)}{n} - F(x) \right| \geq d\right)$$

as discussed earlier,

$$1(Y_i \leq x) = 1(F(Y_i) \leq F(x))$$

so

$$= P\left(\max_x \left| \frac{\sum_{i=1}^{n} 1(F(Y_i) \leq F(x))}{n} - F(x) \right| \geq d\right)$$

Substitute $F(x) = y$, as $0 \leq F(x) \leq 1, 0 \leq y \leq 1$

so

$$= P\left(\max_{0 \leq y \leq 1} \left| \frac{\sum_{i=1}^{n} 1(F(Y_i) \leq y)}{n} - y \right| \geq d\right)$$

From my claim earlier, $F(Y_i)$ will be acting as a new uniform random variable $Unif(0,1)$ hence $F(Y_i) = Unif(0,1) = U_i$ (say)

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

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

So the previous step is justified only because if I take n-independent uniform variables samples from $[0,1]$ & those from $F(Y_i)$s, they will have same probability for same event (say event being $\geq d$), as $F(Y_i)$s are acting as a new uniform variable, completely randomised now & have same CDFs as a separate uniform random variable and hence the chances of both quantities E & D $\geq d$ is same.

Taking $F(Y_i)$ sample is like taking first lot of n-independent uniform samples and taking U_i sample is like taking another lot of n-independent samples, & prob. for any event for these 2 lots will be same (due to randomness, can't differentiate which is which).

Hence $P(D \geq d) = P(E \geq d)$ proved.

Importance of this equation →

$P(D \geq d)$ is same for whatever distribution F is chosen. It means it is independent of that F distribution.

This can be useful in checking if the data given for a distribution F indeed is of F or not? If the data belonged to F, then $P(D \geq d)$ must not vary too much from $P(E \geq d)$ computed for a uniform(0,1) variable.

If it varies too much, then we can know that the data does not fit F very well.

Hence it forms a Goodness-to-Fit test.

Q3 (a) $z = ax + by + c$.

$$Z_i = ax_i + by_i + c + \varepsilon_i \quad \forall\, i \in \{1, 2, \dots n\}$$

where x_i, y_i are known and also $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$.

All ε_i values are independent. as they are drawn from iid random variable.

$$z_i \sim \mathcal{N}(ax_i + by_i + c, \sigma^2)$$

$\therefore$ ~~$p(z_i; x_i, y_i, a, b, c) = \frac{e^{-\frac{(y_i - mx_i - c)^2}{2\sigma}}}{\sigma \cdot \sqrt{2\pi}}$~~

$$p(z_i; x_i, y_i, a, b, c) = \frac{e^{-\left(\frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2}\right)}}{\sigma\sqrt{2\pi}}.$$

$$p(\{z_i\}; \{x_i\}, \{y_i\}, a, b, c) = \prod_{i=1}^{n} \frac{e^{\left(-\frac{(z_i - ax_i - by_i - c)^2}{2\sigma^2}\right)}}{\sigma\sqrt{2\pi}}.$$

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

$$\frac{1}{p} \cdot \frac{\partial p}{\partial a} = \sum_{i=1}^{n} \frac{+(z_i - ax_i - by_i - c)\, x_i}{\sigma^2} = 0.$$

$$\Rightarrow \boxed{\sum_{i=1}^{n} x_i z_i = a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} y_i x_i + c\sum_{i=1}^{n} x_i} \text{ — } 1$$

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

$$\Rightarrow \boxed{\sum_{i=1}^{n} z_i y_i = a\sum_{i=1}^{n} x_i y_i + b\sum_{i=1}^{n} y_i^2 + c\sum_{i=1}^{n} y_i} \text{ — } 2$$

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

$$\Rightarrow \boxed{\sum_{i=1}^{n} z_i = a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} y_i + nc} \text{ — } 3$$

Matrix Form

$$\begin{bmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} y_i^2 & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} 1 \end{bmatrix}_{3\times 3} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n} x_i z_i \\ \sum_{i=1}^{n} y_i z_i \\ \sum_{i=1}^{n} z_i \end{bmatrix}$$

Vector Form

I can represent every Row of the 3x3 matrix by a 3 dimentional vector. a_{ij} are elements of this 3x3 matrix.

$$\cancel{(R_1 + R_2 + R_3) \cdot (A) = Z}$$

$$R_i = (a_{i1}, a_{i2}, a_{i3}) \qquad Z_1 = \sum_{i=1}^{n} x_i z_i$$

$$A = (a, b, c) \qquad Z_2 = \sum_{i=1}^{n} y_i z_i$$

$$\cancel{Z = \left(\sum x_i z_i, \sum y_i z_i, \sum z_i\right)} \qquad Z_3 = \sum_{i=1}^{n} z_i$$

$$R_i \cdot A = Z_i \quad \forall\, i \in \{1, 2, 3\}$$

(b) $z = ax + by + c$

Note $\sum_{i=1}^{n}$ will be represented as $\sum$

$$z_i = a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6 + \varepsilon_i \quad \forall i \in \{1,2,3,\dots N\}$$

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

All ε_i are independent.

$$z_i \sim N(a_1 x_i^2 + a_2 y_i^2 + a_3 x_i y_i + a_4 x_i + a_5 y_i + a_6,\ \sigma^2)$$

$$\therefore p(z_i; x_i, y_i, a_1, a_2, a_3, a_4, a_5, a_6) = \frac{e^{-\frac{(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

$$p(\{z_i\}; \{x_i\}, \{y_i\}, a_1, a_2, a_3, a_4, a_5, a_6) = \prod_{i=1}^{n} \frac{e^{-\frac{(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

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

$$\frac{1}{p}\frac{\partial p}{\partial a_1} = \sum_{i=1}^{n} \frac{(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)\, x_i^2}{\sigma^2} = 0$$

① $\boxed{\sum z_i x_i^2 = a_1 \sum x_i^4 + a_2 \sum y_i^2 x_i^2 + a_3 \sum x_i^3 y_i + a_4 \sum x_i^3 + a_5 \sum x_i^2 y_i + a_6 \sum x_i^2}$

$$\frac{1}{p}\frac{\partial p}{\partial a_2} = \sum \frac{(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)(y_i^2)}{\sigma^2} = 0$$

② $\boxed{\sum z_i y_i^2 = a_1 \sum x_i^2 y_i^2 + a_2 \sum y_i^4 + a_3 \sum x_i y_i^3 + a_4 \sum x_i y_i^2 + a_5 \sum y_i^3 + a_6 \sum y_i^2}$

$$\frac{1}{p}\frac{\partial p}{\partial a_3} = \sum \frac{(z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)\, x_i y_i}{\sigma^2} = 0$$

③ $\boxed{\sum x_i y_i z_i = a_1 \sum x_i^3 y_i + a_2 \sum x_i y_i^3 + a_3 \sum x_i^2 y_i^2 + a_4 \sum x_i^2 y_i + a_5 \sum x_i y_i^2 + a_6 \sum x_i y_i}$

$$\frac{1}{p}\frac{\partial p}{\partial a_4} = \frac{\sum (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)\, x_i}{\sigma^2} = 0$$

④ $\boxed{\sum x_i z_i = a_1 \sum x_i^3 + a_2 \sum x_i y_i^2 + a_3 \sum x_i^2 y_i + a_4 \sum x_i^2 + a_5 \sum x_i y_i + a_6 \sum x_i}$

$$\frac{1}{p}\frac{\partial p}{\partial a_5} = \frac{\sum (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)\, y_i}{\sigma^2} = 0$$

⑤ $\boxed{\sum y_i z_i = a_1 \sum x_i^2 y_i + a_2 \sum y_i^3 + a_3 \sum x_i y_i^2 + a_4 \sum x_i y_i + a_5 \sum y_i^2 + a_6 \sum y_i}$

$$\frac{1}{p}\frac{\partial p}{\partial a_6} = \frac{\sum (z_i - a_1 x_i^2 - a_2 y_i^2 - a_3 x_i y_i - a_4 x_i - a_5 y_i - a_6)}{\sigma^2} = 0$$

⑥ $\boxed{\sum z_i = a_1 \sum x_i^2 + a_2 \sum y_i^2 + a_3 \sum x_i y_i + a_4 \sum x_i + a_5 \sum y_i + a_6 \sum 1.}$

Matrix Form

$$
\begin{bmatrix}
\sum x_i^4 & \sum x_i^2 y_i^2 & \sum x_i^3 y_i & \sum x_i^3 & \sum x_i^2 y_i & \sum x_i^2 \\
\sum y_i^2 x_i^2 & \sum y_i^4 & \sum x_i y_i^3 & \sum x_i y_i^2 & \sum y_i^3 & \sum y_i^2 \\
\sum x_i^3 y_i & \sum x_i y_i^2 & \sum x_i^2 y_i^2 & \sum x_i^2 y_i & \sum x_i y_i^2 & \sum x_i y_i \\
\sum x_i^3 & \sum x_i y_i^2 & \sum x_i^2 y_i & \sum x_i^2 & \sum x_i y_i & \sum x_i \\
\sum x_i^2 y_i & \sum y_i^3 & \sum x_i y_i^2 & \sum x_i y_i & \sum y_i^2 & \sum y_i \\
\sum x_i^2 & \sum y_i^2 & \sum x_i y_i & \sum x_i & \sum y_i & \sum 1
\end{bmatrix}_{6\times 6}
\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \end{bmatrix}
=
\begin{bmatrix} \sum x_i^2 z_i \\ \sum y_i^2 z_i \\ \sum x_i y_i z_i \\ \sum x_i z_i \\ \sum y_i z_i \\ \sum z_i \end{bmatrix}
$$

Vector Form

$R_i = \{a_{i1}, a_{i2}, a_{i3}, a_{i4}, a_{i5}, a_{i6}\}$ $\quad a_{ij}$ are elements of 6×6 matrix.

$A = \{a_1, a_2, a_3, a_4, a_5, a_6\}$

$Z_1 = \sum x_i^2 z_i \; ; Z_2 = \sum y_i^2 z_i \; ; Z_3 = \sum x_i y_i z_i \; ; Z_4 = \sum x_i z_i$

$Z_5 = \sum y_i z_i \quad Z_6 = \sum z_i$

$\therefore \; \forall \; i \in \{1, 2, 3, 4, 5, 6\}$

$R_i \cdot A = Z_i$

HW3 Q3(c)

Kavya Gupta

September 2023

Instructions to Run the Code

The MATLAB Code for this question is saved in the file `A3Q3.m` present in the main zip directory.
Upon running the code, you shall see the required command line outputs and a figure :-

- `A3Q3_plot.png` Contains the plot of the estimated plane with the data points given from `XYZ.txt`.

Plot

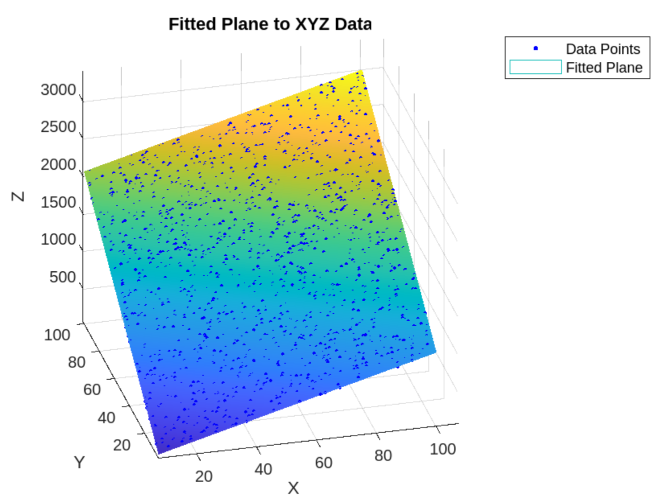


Figure 1: Plot of Estimated Plane

Predicted Equation of the Plane :-

$$z = 10.0022x + 19.9980y + 29.9516$$

Predicted Noise Variance :- 23.0685

HW3 Q4

Kavya Gupta

September 2023

Instructions to Run the Code

The MATLAB Code for **part (c) and (d)** is contained in the file `A3Q4.m` present in the main zip directory. Running that file will give 4 plots **in the same order** :-

- `Q4_Pic1.png` : Contains Plot of Log Likelihood (LL) vs log(sigma)
- `Q4_Pic2.png` : Contains Plot of D Value (D) vs log(sigma)
- `Q4_Pic3.png` : Contains graph for plot of density with best sigma for **maximum LL** and true density.
- `Q4_Pic4.png` : Contains graph for plot of density with best sigma for **minimum D** and true density.

The code will also successfully output the required print statements in the command line output, please see that.
The MATLAB Code for **part (e)** is contained in the file `A3Q4_2.m` present in the main zip directory. Running that file will give 4 plots **in the same order** :-

- `Q4_Pic5.png` : Contains Plot of Log Likelihood (LL) vs log(sigma) (T = V)
- `Q4_Pic6.png` : Contains Plot of D Value (D) vs log(sigma) (T = V)
- `Q4_Pic7.png` : Contains graph for plot of density with best sigma for **maximum LL** and true density (T = V)
- `Q4_Pic8.png` : Contains graph for plot of density with best sigma for **minimum D** and true density (T = V)

The code will also successfully output the required print statements in the command line output, please see that.
Note that all these eight plots are already included in this pdf and are also present in folder `Q4`. This folder is in main zip directory.

Part (a)

I have used `randperm` function to get permutation of the indices from 1 to n and took first 750 indices for Sample space `T` and rest for Validation space `V`. Also I have set the seed to 0 using `rng(0);`, so that I can reproduce same result every time.

Part (b)

We are given the Likelihood of one point x : $\hat{p_n}(x, \sigma)$. So, let's say that i^{th} entry of V as v_i and j^{th} entry of T as t_j, hence likelihood for 1 point of V :-

$$\hat{p_n}(v_i, \sigma) = \frac{\sum_{j=1}^{|T|} e^{-\frac{(v_i - t_j)^2}{2\sigma^2}}}{|T|\sigma\sqrt{2\pi}}$$

Here $|T|$ represents cardinality of set T and hence its size.
So the Joint Likelihood for $\{v_i\}_{i=1}^{|V|}$ will be :-

$$\hat{p_n}(\{v_i\}_{i=1}^{|V|}, \sigma) = \prod_{i=1}^{|V|} \hat{p_n}(v_i, \sigma) = \prod_{i=1}^{|V|} \frac{\sum_{j=1}^{|T|} e^{-\frac{(v_i - t_j)^2}{2\sigma^2}}}{|T|\sigma\sqrt{2\pi}}$$

We were able to split $\hat{p_n}(\{v_i\}_{i=1}^{|V|}, \sigma)$ into products as $\{v_i\}_{i=1}^{|V|}$ sampling is considered **independent**.

Part (c)

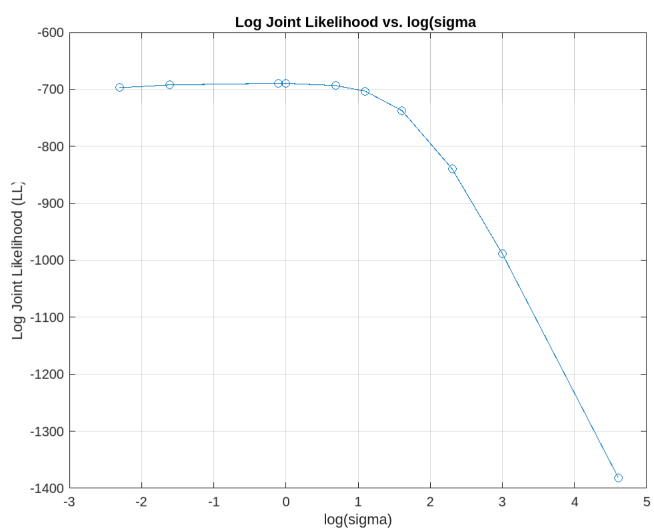


Figure 1: Plot of Log Likelihood (LL) vs log(sigma)

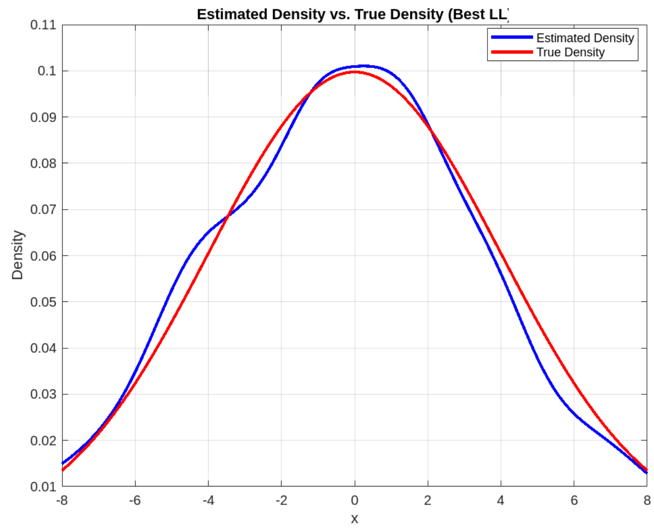


Figure 2: Graph for plot of density with best sigma for maximum LL = **0.9000** and true density

The value of σ for best LL : **0.9000**
Corresponding LL Value : **-689.3216**

Part (d)

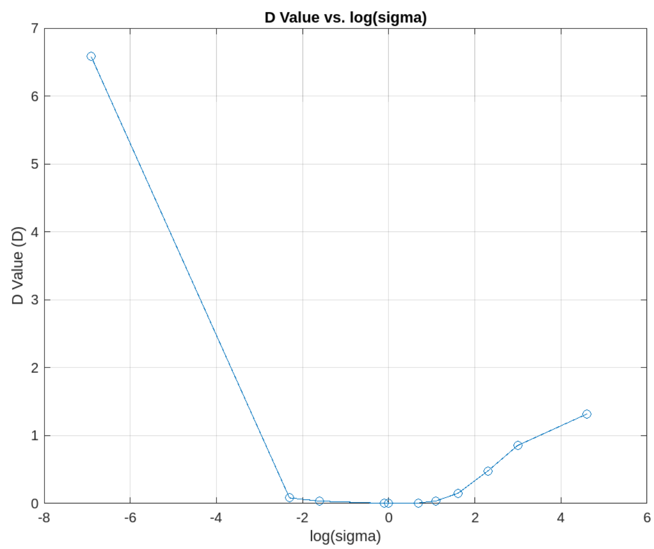


Figure 3: Plot of D value (D) vs log(sigma)

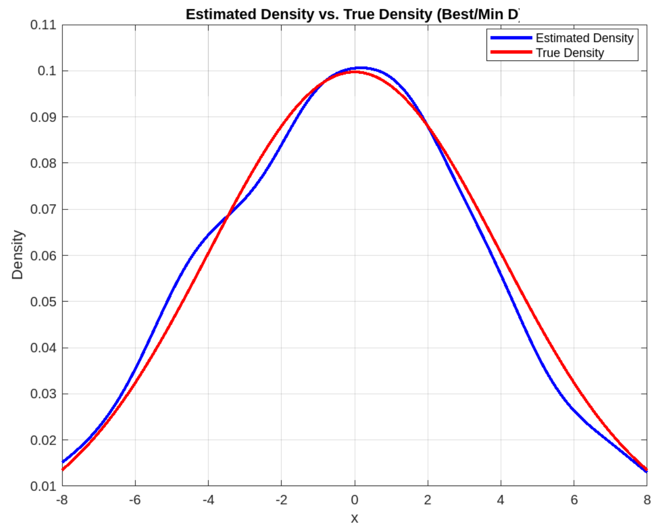


Figure 4: Graph for plot of density with best sigma for minimum D = **1.0000** and true density

The value of σ for best/min D : **1.0000**
Corresponding D Value : **0.0029**
D Value for Sigma with Max LL (1.0000) : **0.0034**

Worth noticing that the estimate graphs found by best LL and best D methods are both **quite close to each other**.

Part (e)

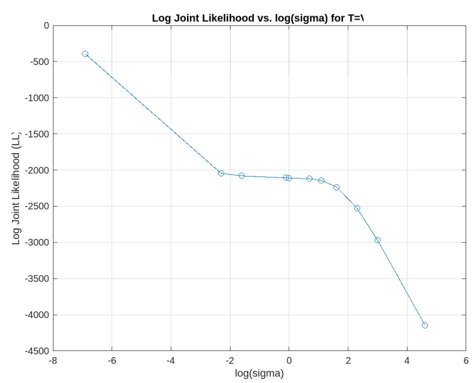


Figure 5: Plot of Log Likelihood (LL) vs log(sigma) (T=V)

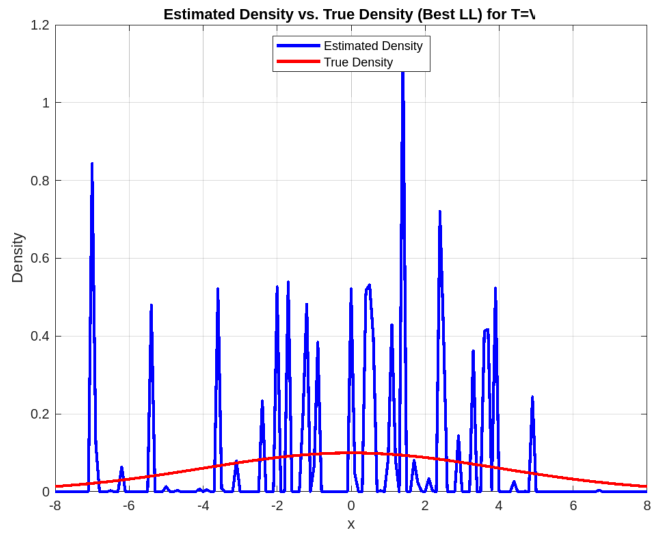


Figure 6: Graph for plot of density with best sigma for maximum LL = **0.001** and true density (T=V)

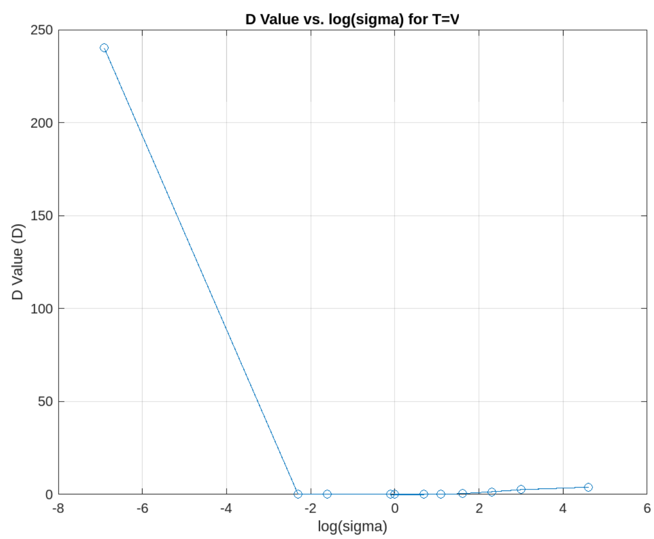


Figure 7: Plot of D value (D) vs log(sigma) (T=V)

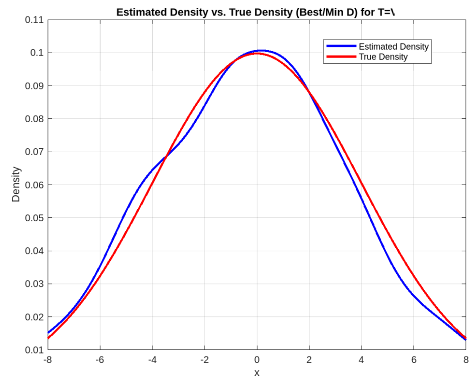


Figure 8: Graph for plot of density with best sigma for minimum D = **1.0000** and true density (T=V)

Best Sigma for Max. LL for T=V: **0.0010**
Best Sigma for Min. D for T=V: **1.0000**

So if the sets T and V became equal to each other, in other words, training sample and validation set became same; then what I observed was from the code that σ for best LL is the **smallest one** = 0.001 from the given set.
Meaning for smaller σ, likelihood is maximum.
Also the graph of LL vs log(sigma) seems to **greatly increase as log(sigma) tends to $-\infty$.**

Reason

When T and V become equal, the density is calculated at the same points from where PDF was estimated. So density of a term t_i will become :-

$$\hat{p_n}(t_i,\sigma) = \frac{\sum_{j=1}^{|T|} e^{-\frac{(t_i-t_j)^2}{2\sigma^2}}}{|T|\sigma\sqrt{2\pi}} = \frac{1+\sum_{j\neq i}\cdots}{|T|\sigma\sqrt{2\pi}} = \frac{1}{|T|\sigma\sqrt{2\pi}} + \ldots$$

The above term becomes independent from the data and rather depends on σ.
This term $\frac{1}{|T|\sigma\sqrt{2\pi}}$ tends to ∞ as σ tends to 0, hence joint likelihood which is product of individual likelihoods also attains max near 0.
Hence max LL will be found at $\sigma \to 0^+$. **Hence we see spikes in data...**
Such a term exists as one term t_i exactly cancels out t_j for $i = j$.

Conclusion

When T and V match then our cross validation procedure tends to fail. **Even for max LL, we got the worst possible matching graph.**

22B1053-22B1205-22B1025

Q5 let Y be R.V st $(X \in [a,b]$ is ~~[illegible]~~ R.V$)$

$Y = e^{s(X - E[X])}$ where $s > 0$

now using markov's inequality on Y, (as Y is positive) (as $e^{st} > 0 \ \forall t \in \mathbb{R}$)

$$P(Y = e^{s(X-E(X))} \geq e^{st}) \leq \frac{E(Y)}{e^{st}}$$

now by IR $E(Y) \leq e^{s^2(b-a)^2/8}$

$$\therefore P(e^{s(X-E(X))} \geq e^{st}) \leq \frac{e^{s^2(b-a)^2/8}}{e^{st}}$$

now as $e^{s(X-E(X))} \geq e^{st} \Leftrightarrow X - E(X) \geq t$ for $t > 0$

$$\therefore P(X - E(X) \geq t) \leq e^{\frac{s^2(b-a)^2}{8} - st} = e^{\lambda} \quad \text{for } t > 0$$

for $t > 0, s > 0$ we minize R.H.S ~~which is obtained at~~

$$\lambda = \frac{s^2(b-a)^2}{8} - st = \left(\frac{s(b-a)}{2\sqrt{2}} - \frac{t\sqrt{2}}{b-a}\right)^2 + \frac{2t^2}{(b-a)^2}$$

$$\therefore \lambda \geq \frac{2t^2}{(b-a)^2}$$

$$\therefore P(X - E(X) \geq t) \leq e^{2t^2/(b-a)^2} \quad \text{for } t > 0$$

for $t < 0$,

let $Y = e^{s(E(X) - X)}$ where $s > 0$, ~~$t<0$~~, $Y > 0$.

applying IR on $E(X) - X$ as R.V (note: if $X \in [a,b]$ then $E(X) - X \in [\mu - b, \mu - a]$ $\mu = E(X)$)

$$E(Y) \leq e^{s^2(b-a)^2/8}$$

now for $t < 0$ we use markov's equality as follows.

$$P(Y \geq e^{-st}) \leq \frac{e^{s^2(b-a)^2/8}}{e^{-st}}$$

now $e^{s(E(X)-X)} \geq e^{-st} \Leftrightarrow E(X) - X \geq -t$ for $t < 0$

$\Leftrightarrow X - E(X) \leq t$

$$1 - P(X - E(X) > t) = P(X - E(X) \leq t) \leq e^{\frac{s^2(b-a)^2}{8} + st}$$

$$\therefore P(X - E(X)) \geq 1 - e^{\frac{s^2(b-a)^2}{8} + st} = 1 - e^{\lambda}$$

we maximize R.H.S

now, $\lambda = \frac{s^2(b-a)^2}{8} + st = \left(s\frac{(b-a)}{2\sqrt{2}} + \frac{t\sqrt{2}}{b-a}\right)^2 - \frac{2t^2}{(b-a)^2} \geq \frac{-2t^2}{(b-a)^2}$

$\therefore$ maximum R.H.S $= 1 - e^{-2t^2/(b-a)^2}$

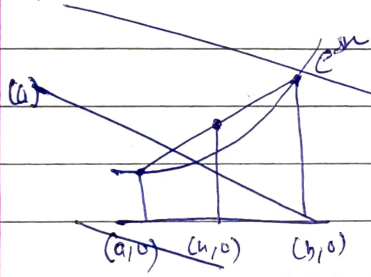
$\therefore \quad P(X - E(X) > t) \geq 1 - e^{-2t^2/(b-a)^2}$ for $t < 0$

$\therefore$ $P(X - E(X) > t) \in [1, 6]$

$\therefore$

$P(X - E(X) \geq t) \leq e^{2t^2/(b-a)^2}$ for $t > 0$

$1 - e^{-2t^2/(b-a)^2} \leq P(X - E(X) > t) \leq 1$ for $t < 0$



by Jensen's inequality,

$e^{su} \leq \frac{(e^{sb} - e^{sa})u}{b-a} - \frac{(ae^{sb} - e^{sa}b)}{b-a} = \frac{(b-u)e^{sa}}{b-a} + \frac{(u-a)e^{sb}}{b-a}$

continue first part ☺

now we need $P(S_n - E(S_n) > t)$

now, we note that

if $x_1 \in [a_1, b_1]$, $x_2 \in [a_2, b_2]$

$x_1 + x_2 \in [a_1 + a_2, b_1 + b_2]$

$\therefore$ by inductive argument

$S_n \in \left[\sum_{i=1}^n a_i, \sum_{i=1}^n b_i\right]$

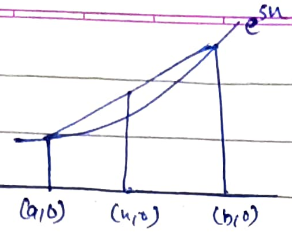
$\therefore$ (also assume x_i's to be continuous)

$\therefore$ from above arguments,

$P(S_n - E(S_n) > t) \leq e^{2t^2/(\sum_{i=1}^n b_i - \sum_{i=1}^n a_i)^2}$ for $t > 0$

$1 - e^{-2t^2/(\sum_{i=1}^n b_i - \sum_{i=1}^n a_i)^2} \leq P(S_n - E(S_n) > t) \leq 1$ for $t < 0$

(a)



by Jensen's inequality

$$e^{sx} \le \frac{(e^{sb}-e^{sa})x}{b-a} - \frac{(ae^{sb}-be^{sa})}{b-a} = \frac{(b-x)e^{sa}}{b-a} + \frac{(x-a)e^{sb}}{b-a}$$

(b)

$$E(e^{sx}) \le E\left[\frac{(e^{sb}-e^{sb})x}{b-a}\right] + E\left[\frac{be^{sa}-ae^{sb}}{b-a}\right] = \frac{be^{sa}-ae^{sb}}{b-a} + \frac{e^{sb}-e^{sa}}{b-a}E[X]$$

$$E(e^{sx}) \le \frac{be^{sa}-ae^{sb}}{b-a} = l$$

~~now its enough to show,~~

$$\cancel{L(s(b-a))} = \cancel{\frac{be^{sa}-ae^{sb}}{b-a}}$$

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

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

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

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

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

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

$$\therefore E(e^{sx}) \le e^{L(s(b-a))}$$

$\cancel{L(0)=0}$

(c) Note that we need only $h>0$ as $s(b-a)>0$.

now, clearly $L(h)$ is continuous and differentiable.

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

$L'(h)$ is differentiable for $h>0$ (as $a<0$), $L'(0)=0$ — (1)

$$\therefore L''(h) = \frac{-a^2e^{2h}}{(b-ae^h)^2} - \frac{ae^h}{(b-ae^h)} = \frac{-ab\,e^h}{(b-ae^h)^2} = \frac{-ab}{(be^{-h/2}-ae^{h/2})^2}$$

for $x,y>0$ $\quad x+y \geq 2\sqrt{xy}$

$$\therefore be^{-h/2}-ae^{h/2} \geq 2\sqrt{-ab}$$

$$\therefore L''(h) \leq \frac{-ab}{4\,-ab} = 1/4$$

(d) now, we have if $f'(x) \leq g'(x)$ ~~or~~ $f'(x)-g'(x) \leq 0 \quad \forall x>0$

and $f(x)-g(x) \leq 0$

$$\Rightarrow f(x)-g(x) \leq 0 \quad \forall x>0 \quad — (2)$$

$\therefore$ as

$$L''(h) - 1/4 \leq 0$$

$$L'(0)=0 \quad \therefore \quad L'(0) - \frac{0}{4} \leq 0 \quad \left(\text{here } g(x)=\frac{h}{4}\right)$$

$$\therefore L'(h) - \frac{h}{4} \leq 0 \quad \forall h>0$$

now as $L(0)=0$,

$$\therefore L(0) - \frac{0}{8} \leq 0 \quad \left(\text{here } g(h) = \frac{h^2}{8}\right)$$

$$\therefore L(h) - \frac{h^2}{8} \leq 0 \quad \forall h>0$$

$$\therefore L(h) \leq \frac{h^2}{8} \quad \forall h>0$$

$$\therefore P[e^{sx}] \leq e^{L(s(b-a))} \leq e^{\frac{s^2(b-a)^2}{8}}$$

prove for (2),

let $h(x) = f(x)-g(x)$ then given $h'(x) \leq 0$, $h(0) \leq 0$

Then by Taylor's theorem $\exists\, c \in [0,x)$ s.t

$$h(x) = h(0) + h'(c)(x-0) = h(0) + x h'(c)$$

as $h(0) \leq 0$ and $xh'(x) \leq 0$ for $x>0$ $\therefore h(x) \leq 0 \;\forall x>0$.