ABHINAV KUMAR BHASKAR GUPTA

180050003 180050022

1\. (a) The first selected will obviously be of a different colour, so $=1$.

Which can't be selected

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

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

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

Probability that a new colour is picked.

Probability that old colour is picked "$k-1$" times

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

(c). Let Z be g.r.v.

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

$$= p\sum_{i=1}^{\infty} i \cdot (1-p)^{i-1}$$

$$= p\left[\frac{-\partial}{\partial p}\Big[\underbrace{\sum_{i=1}^{\infty}(1-p)^{i}}_{\frac{1-p}{1-(1-p)} = \frac{1-p}{p} = \frac{1}{p} - 1}\Big]\right]$$

$$= p\left(\frac{1}{p^2}\right) = 1/p.$$

Scanned by CamScanner

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

$$E(Z^2) = p\left[\sum_{i=1}^{\infty}(1-p)^{i-1} \cdot i \cdot i\right]$$

$$= -\frac{\partial}{\partial p}\left[\underbrace{\left[\frac{-\partial}{\partial p}\sum_{i=1}^{\infty}(1-p)^{i}\right]}_{1/p^2}(1-p)\right](p)$$

$$= -\frac{\partial}{\partial p}\left[1/p^2 - \frac{1}{p}\right]p$$

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

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

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

$$= \sum_{i=1}^{n}\frac{n}{n+1-i} = n\left[1 + \frac{1}{2} + \dots + \frac{1}{n}\right].$$

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

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

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

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

$$= n^2\left[\frac{1}{1^2} + \frac{1}{2} + \dots + \frac{1}{n^2}\right] - n\left[\frac{1}{1} + \dots + \frac{1}{n}\right]$$

Scanned by CamScanner

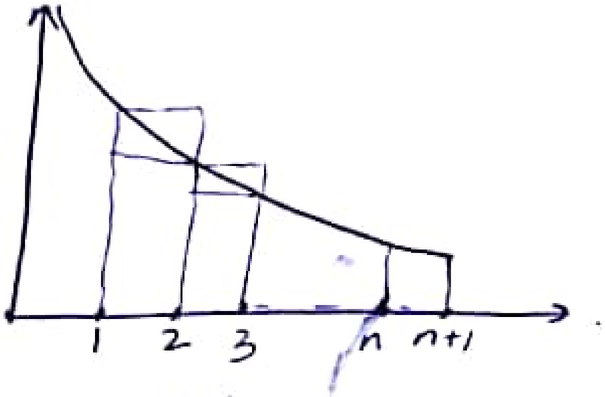
(f) $E[X^n] = n\left[1 + \frac{1}{2} + \dots + \frac{1}{n}\right]$.

as $n \to \infty$ $\quad 1 + \frac{1}{2} + \dots + \frac{1}{n}$ tends to $\int_1^n \frac{dx}{x}$

$\therefore \approx n \ln n$.

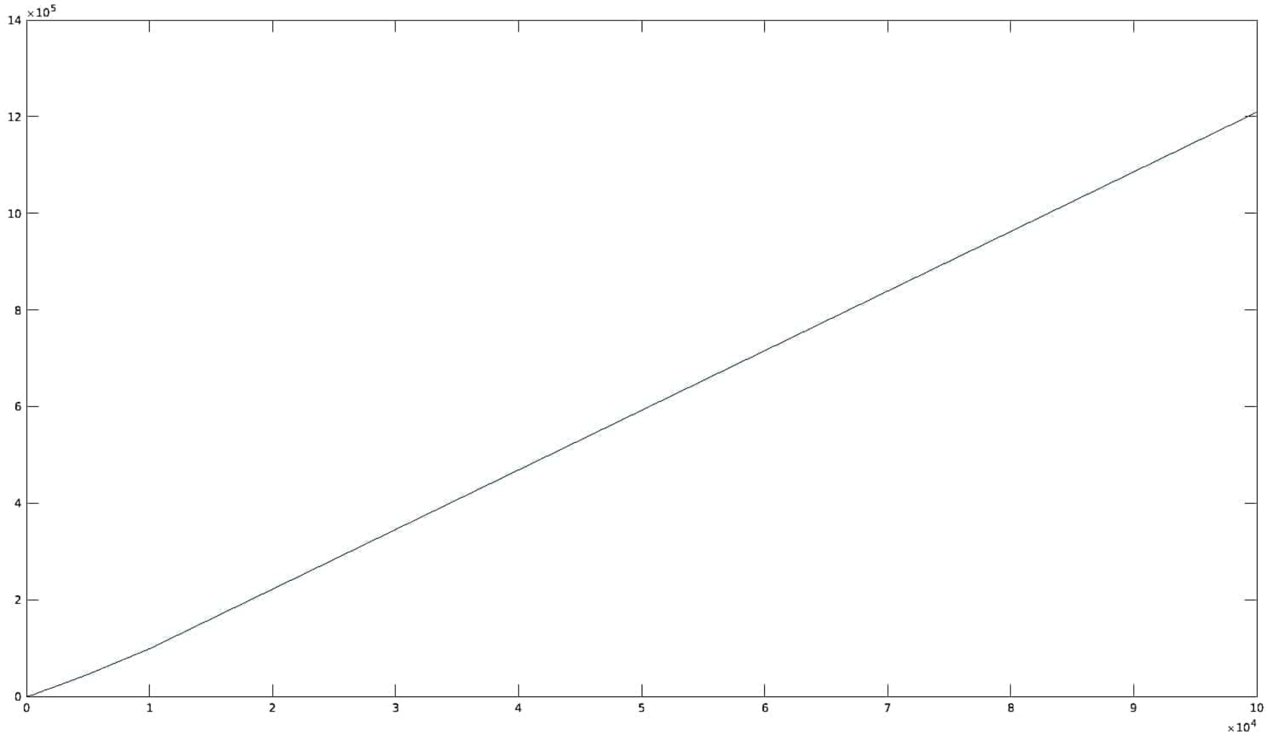
$\Theta(n) = n \ln n$.

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



$$S_n > \int_1^{n+1} \frac{dx}{x} = \ln(n+1) > \ln n.$$

$$S_n < 1 + \int_1^n \frac{1}{x}\,dx = 1 + \ln(n) < 2 \ln n$$


×10^5
14
12
10
8
6
4
2
0
0
1
2
3
4
5
6
7
8
9
10
×10^4

2\. (a). Let $\{V_i\}_{i=1}^{n}$ be the values from some distribution of which X is a random variable such that $X = F^{-1}(U)$ where U is a uniform random variable.

C. D. F. of X.

$$P(X \leq Z) = P(F^{-1}(U) \leq Z) = P(U \leq F(Z)) = F(Z).$$

We use the fact that U is a random uniform variable.

Since X follows the same distribution as that of F. X is a random variable of distribution of F by uniqueness of CDF.

$X = F^{-1}(U)$ follows F. distribution

(b) . $D = \max_x |F_e(x) - F(x)| \geq d$.

$$P(D \geq d) = 1 - P(D \leq d)$$

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

$$= P[(F_x(x) - F(x)) \leq d] \quad \forall x \in (-\infty, \infty)$$

$$P\left[\frac{\sum_{i=1}^{n} 1(Y_i \leq x)}{n} - F(x) \leq d\right] \quad \forall x \in \mathbb{R}$$

$Y = F^{-1}(U)$ follows the F distribution

$$\Rightarrow P\left[\frac{\sum_{i=1}^{n} 1(F^{-1}(u_i) \le x)}{n} - F(x) \le d\right] \forall x \in R$$

$$= P\left[\frac{\sum_{i=1}^{n} 1(u_i \le y_i)}{n} - y \le d\right] \forall y = F(x) \text{ in } [0,1]$$

D qualitatively denotes the maximum deviation of an empirical distribution from the actual distribution. From this result we can conclude that the extent of convergence of the empirical distribution to the actual one is the same as that of the standard uniform distribution. Further we can say that the same argument can be extended to any two arbitrary distributions provided their CDFs are continuous.

Q.4 → and hence M.S.E., so it's not desirable.

Q.3 $Z_i = ax_i + by_i + c + u_i$, $u_i \in N(0, \sigma^2)$

$\Rightarrow Z_i \in N(ax_i + by_i + c, \sigma^2)$ Since Z_i is gaussian$(0,\sigma^2)$ with shifted mean.

$$L = \log[P(Z_1 = z_1, Z_2 = z_2, \ldots ; a,b,c)]$$

since noises are independent

$$L = \log\left[P(Z_1 = z_1) \ldots P(Z_n = z_n)\right]$$

$$= \log \frac{e^{-\sum_{i=1}^{n} \frac{(z_i - (ax_i + by_i + c))^2}{2\sigma^2}}}{(\sigma\sqrt{2\pi})^n}$$

$$L = -\sum_{i=1}^{n} \frac{(z_i - (ax_i + by_i + c))^2}{2\sigma^2} - n\log\sigma - \frac{1}{2}\log(2\pi)$$

is to be maximised

$$\Rightarrow \frac{\partial L}{\partial a} = 0, \quad \frac{\partial L}{\partial b} = 0, \quad \frac{\partial L}{\partial c} = 0$$

$$\Rightarrow -\sum_{i=1}^{n} \frac{2(z_i - (ax_i + by_i + c))(-x_i)}{2\sigma^2} = 0 \quad -(1)$$

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

$$-\sum_{i=1}^{n} \frac{2(z_i - (ax_i + by_i + c))(-1)}{2\sigma^2} = 0 \quad -(3)$$

$$\sum z_i x_i = a\sum x_i^2 + b\sum x_i y_i + c\sum x_i$$

$$\sum z_i y_i = a\sum x_i y_i + b\sum y_i^2 + c\sum y_i$$

$$\sum z_i = a\sum x_i + b\sum y_i + c\sum 1$$

Date:

Matrix form

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

Vector form

Let $\bar{z} = z_1 \hat{i} + z_2 \hat{j} + \cdots z_n(\ )$

$\bar{x} = x_1 \hat{i} + x_2 \hat{j} + \cdots$

$\bar{y} = y_1 \hat{i} + y_2 \hat{j} + \cdots$

$\bar{u} = \hat{i} + \hat{j} + \cdots$

$\bar{z} \cdot \bar{x} = (a, b, c) \cdot (\bar{x} \cdot \bar{x}, \ \bar{x} \cdot \bar{y}, \ \bar{x} \cdot \bar{u})$

$\bar{z} \cdot \bar{y} = (a, b, c) \cdot (\bar{x} \cdot \bar{y}, \ \bar{y} \cdot \bar{y}, \ \bar{y} \cdot \bar{u})$

$\bar{z} = (a, b, c) \cdot (\bar{x} \cdot \bar{u}, \ \bar{y} \cdot \bar{u}, \ \bar{u} \cdot \bar{u})$

The predicted equation of plane is

$$z = 10.0022x + 19.998y + 29.9516$$

The predicted noise variance is 23.057.

- Since the Expectation operator has linearity property, we can apply Expectation operator in $(*)$ row wise in both L.H.S. & R.H.S.

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

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

$$z_i = \underset{\text{fixed}}{a x_i + b y_i} + c + u_i$$

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

$$E(z_i) = a x_i + b y_i + c$$

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

~~On the~~

$\Rightarrow$ This 3×3 matrix comes out to be invertible

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

$E(\hat{a}) = a$ $\quad E(\hat{c}) = c.$

$E(\hat{b}) = b$

Q.4

b) $J.L. = P(x_1, x_2, \ldots x_{n_2}; \sigma)$ $[x_1, x_2, \ldots x_{n_2} \in V]$

Since all samples in V are independent,

$$J.L. = P(x_1;\sigma) \times P(x_2;\sigma) \times \ldots P(x_{n_2};\sigma)$$

$|V| = n_2$

$$= \prod_{i=1}^{n_2} \left(\sum_{j=1}^{n_1} \frac{e^{-(x_i - y_j)^2/2\sigma^2}}{n_1 \sigma \sqrt{2\pi}} \right)$$

$y_j \in T \quad \forall\, j \in [1, n_1]$

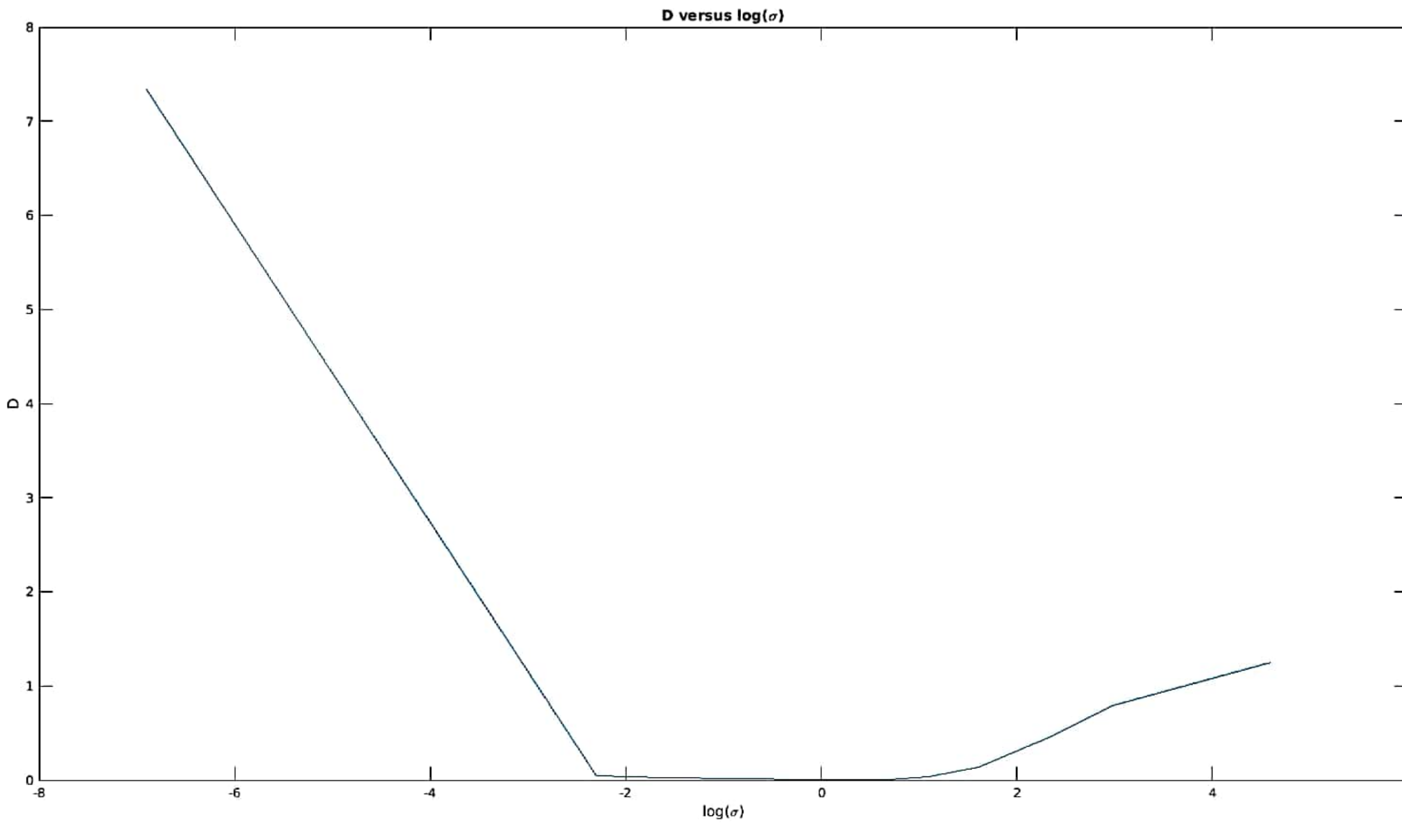
$x_i \in V \quad \forall\, i \in [1, n_2]$

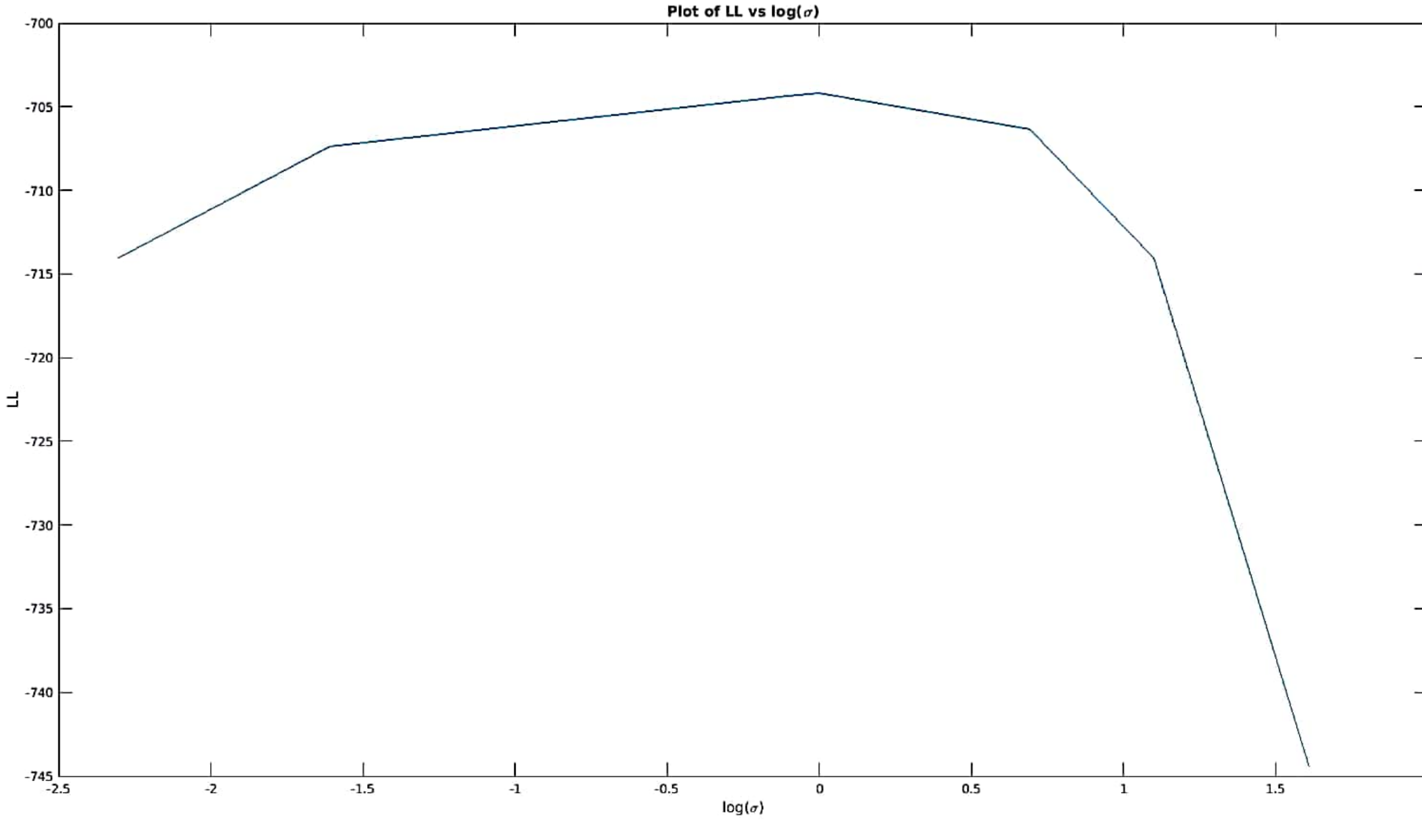
c) Best LL value occurs at sigma = 1
Best LL value is -708

d) Best D value occurs at sigma = 1
Best D value is .0028 (minimum)

e) If T=V, the cross validation procedure will return σ for which Joint likelihood is maximum. But for T=V case, J.L. comes out to be a monotonically decreasing function of σ & hence best value theoretically is 0 but this gives the best J.L. for the set T only. This means that if another sample from the same distribution, which is considerably different than training data set, the results are likely to be inaccurate. This means that the estimate has high variance

Scanned by CamScanner


D versus log(σ)
8
7
6
5
4
3
2
1
0
D
-8
-6
-4
-2
0
2
4
log(σ)


Plot of LL vs log(σ)
LL
log(σ)
-700
-705
-710
-715
-720
-725
-730
-735
-740
-745
-2.5
-2
-1.5
-1
-0.5
0
0.5
1
1.5