12/19/14.

FINAL EXAM- STAT-511.

AMAL AGARWAL.

SHORT Q & A-

1) $y_i = k_i + \frac{1}{2a} x_i^2$, $k_i \sim N(b, c x_i^2)$

Since $\frac{1}{2a} x_i^2$ is constant at each i,

$y_i \sim N\left(\left(b + \frac{1}{2a} x_i^2\right), c x_i^2\right)$

$f_{Y_i}(y_i) = \frac{1}{\sqrt{2\pi (c x_i^2)}} \exp\left[\frac{-1}{2c x_i^2} \left(y_i - b - \frac{x_i^2}{2a}\right)^2\right]$

Since k_i's are independent $\forall i$, Y_i's are independent.

The joint pdf of Y_i's is given as

$f_{\underset{\sim}{Y}}(\underset{\sim}{y}) = \prod_{i=1}^{n} f_{Y_i}(y_i)$ (by the independence of Y_i's)

$= \prod_{i=1}^{n} \left\{\frac{1}{\sqrt{2\pi (c x_i^2)}} \exp\left[\frac{-1}{2c x_i^2} \left(y_i - b - \frac{x_i^2}{2a}\right)^2\right]\right\}$

$= \left(\frac{1}{\sqrt{2\pi c}}\right)^n \prod_{i=1}^{n}\left\{\left(\frac{1}{x_i}\right)\right\} \exp\left\{\left(\frac{-1}{2c}\right) \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \frac{b}{x_i} - \frac{x_i}{2a}\right)^2\right\}$

The joint pdf of $\underset{\sim}{Y}$ is equal to the likelihood function.

Thus.

$L(a, b, c \mid \underset{\sim}{y}) = f_{\underset{\sim}{Y}}(\underset{\sim}{y})$

Scanned by CamScanner

$$\sum_{i=1}^{n} y_i - nb = \frac{1}{2a} \sum_{i=1}^{n} x_i^2$$

$$\Rightarrow \frac{1}{2a} = \frac{\sum_{i=1}^{n} y_i - nb}{\sum_{i=1}^{n} x_i^2} \longrightarrow (3)$$

Substituting eq. (3) in eq. (2),

$$\sum_{i=1}^{n}\left(\frac{y_i}{x_i^2}\right) - b \sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right) = n\left[\frac{\sum_{i=1}^{n} y_i - nb}{\sum_{i=1}^{n} x_i^2}\right]$$

$$\Rightarrow \frac{n^2 b}{\sum_{i=1}^{n} x_i^2} - b \sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right) = \frac{n \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2} - \sum_{i=1}^{n}\left(\frac{y_i}{x_i^2}\right)$$

$$\Rightarrow b\left[n^2 - \sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right)\left(\sum_{i=1}^{n} x_i^2\right)\right] = n \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\left(\frac{y_i}{x_i^2}\right) \sum_{i=1}^{n}(x_i^2)$$

$$\Rightarrow \hat{b}_{MLE} = \frac{n \sum_{i=1}^{n} y_i - \sum_{i=1}^{n}\left(\frac{y_i}{x_i^2}\right) \sum_{i=1}^{n} x_i^2}{n^2 - \sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right)\left(\sum_{i=1}^{n} x_i^2\right)} \longrightarrow (4)$$

Substituting this in eq. (3), we get

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

Now $\log[L(a,b,c \mid \underset{\sim}{y})]$ is the log likelihood

To estimate a, b, c we must attempt to maximize this to find the MLE's.

Now $l(a,b,c \mid \underset{\sim}{y}) = \log\left(L(a,b,c \mid \underset{\sim}{y})\right)$

$$= \frac{-n}{2}\log(2\pi) - \frac{n}{2}\log c - \sum_{i=1}^{n}\log x_i$$

$$\frac{-1}{2c}\sum_{i=1}^{n}\left(\frac{y_i}{x_i} - \frac{b}{x_i} - \frac{x_i}{2a}\right)^2$$

Setting the partial derivatives zero for MLE's →

$$\frac{\partial l}{\partial a} = 0$$

$$\Rightarrow \frac{-2}{2c}\sum_{i=1}^{n}\left(\frac{y_i}{x_i} - \frac{b}{x_i} - \frac{x_i}{2a}\right)\left(\frac{+x_i}{2}\right)\left(\frac{+1}{a^2}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(y_i - b - \frac{x_i^2}{2a}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n} y_i - nb - \frac{1}{2a}\sum_{i=1}^{n} x_i^2 = 0 \qquad \rightarrow (1)$$

Now $\frac{\partial l}{\partial b} = 0$

$$\Rightarrow \frac{+2}{2c}\sum_{i=1}^{n}\left(\frac{y_i}{x_i} - \frac{b}{x_i} - \frac{x_i}{2a}\right)\left(\frac{+1}{x_i}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{y_i}{x_i^2} - \frac{b}{x_i^2} - \frac{1}{2a}\right) = 0$$

$$\Rightarrow \sum_{i=1}^{n}\left(\frac{y_i}{x_i^2}\right) - b\cdot\sum_{i=1}^{n}\left(\frac{1}{x_i^2}\right) - \frac{n}{2a} = 0 \qquad \rightarrow (2)$$

Now from eq. (1) we obtain

Scanned by CamScanner

$$\Rightarrow \frac{1}{2a} = \frac{n \sum_{i=1}^{n} \left(\frac{y_i}{x_i^2}\right) - \sum_{i=1}^{n} \left(\frac{1}{x_i^2}\right)\left(\sum_{i=1}^{n} y_i\right)}{n^2 - \sum_{i=1}^{n} \left(\frac{1}{x_i^2}\right) \sum_{i=1}^{n} \left(x_i^2\right)}$$

$$\Rightarrow \hat{a}_{MLE} = \frac{n^2 - \sum_{i=1}^{n} \left(\frac{1}{x_i^2}\right) \sum_{i=1}^{n} \left(x_i^2\right)}{2\left[n \sum_{i=1}^{n} \frac{y_i}{x_i^2} - \sum_{i=1}^{n} \left(\frac{1}{x_i^2}\right)\left(\sum_{i=1}^{n} y_i\right)\right]}$$

Now $\frac{\partial l}{\partial c} = 0.$

$$\Rightarrow \frac{-n}{2c} + \frac{1}{2c^2} \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \frac{b}{x_i} - \frac{x_i}{2a}\right)^2 = 0.$$

$$\Rightarrow \hat{c}_{MLE} = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{y_i}{x_i} - \frac{\hat{b}}{x_i} - \frac{x_i}{2\hat{a}}\right)^2$$

Scanned by CamScanner

2) $\underset{\sim}{y} \sim N(X\underset{\sim}{\beta}, \sigma^2 I)$

Now $\underset{\sim}{\epsilon} \sim N(0, \sigma^2 I)$

Residuals are given as. $\hat{\underset{\sim}{\epsilon}} = (I-H)\underset{\sim}{y}$.

$(I-H)\underset{\sim}{y} \sim N(\underset{\sim}{0}, \sigma^2(I-H))$

Thus marginally

$\hat{\epsilon}_i \sim N(0, \sigma^2(1-h_{ii}))$

Now for testing one outlier we considered a leave one-out analysis to construct Studentized residuals which followed a t-distribution.

Working along the same lines, consider a leave 3 (consecutive) observations out analysis. Thus we will have $(n-3)$ datasets each of $(n-3)$ dimensions. Consider the dataset where $(i-1)^{th}$, i^{th} and $(i+1)^{th}$ observations are left out i.e.

$$\underset{\sim}{y}_{(i)} = (y_1, y_2, \ldots, y_{i-2}, y_{i+2}, \ldots, y_n)^T_{(n-3)\times 1}.$$

$X_{(i)} \rightarrow$ matrix X without $(i-1)^{th}$, i^{th} and $(i+1)^{th}$ rows

Scanned by CamScanner

$$\underset{\sim}{y}_{(i)} \sim N\left(X_{(i)}\underset{\sim}{\beta},\ \sigma^2 I\right) \longrightarrow (1)$$

$$\hat{\underset{\sim}{\beta}}_{(i)} = \left(X'_{(i)} X_{(i)}\right)^{-1} X'_{(i)} \underset{\sim}{y}_{(i)} \longrightarrow (2)$$

$$\hat{\sigma}^2_{(i)} = \frac{1}{[(n-3)-p]}\left(\underset{\sim}{y}_{(i)} - X_{(i)}\hat{\underset{\sim}{\beta}}_{(i)}\right)'\left(\underset{\sim}{y}_{(i)} - X_{(i)}\hat{\underset{\sim}{\beta}}_{(i)}\right) \longrightarrow (3)$$

Thus the vector of $(i-1)^{th}$, i^{th} and $(i+1)^{th}$ residuals is

$$\begin{bmatrix} \hat{\epsilon}_{(i-1)} \\ \hat{\epsilon}_{(i)} \\ \hat{\epsilon}_{(i+1)} \end{bmatrix} = \begin{bmatrix} y_{i-1} - \underset{\sim}{X}_{i-1}\hat{\underset{\sim}{\beta}}_{(i)} \\ y_i - \underset{\sim}{X}_i \hat{\underset{\sim}{\beta}}_{(i)} \\ y_{i+1} - \underset{\sim}{X}'_{i+1}\hat{\underset{\sim}{\beta}}_{(i)} \end{bmatrix}$$

Now
$$\begin{bmatrix} y_{i-1} \\ y_i \\ y_{i+1} \end{bmatrix} \sim N\left[\begin{pmatrix} \underset{\sim}{X}'_{i-1}\underset{\sim}{\beta} \\ \underset{\sim}{X}'_i\underset{\sim}{\beta} \\ \underset{\sim}{X}'_{i+1}\underset{\sim}{\beta} \end{pmatrix},\ \sigma^2\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right] \longrightarrow (4)$$

Using (1), (2) and (4), we get

$$\underset{\sim}{\epsilon}^{(i)} = \begin{bmatrix} \hat{\epsilon}_{(i-1)} \\ \hat{\epsilon}_{(i)} \\ \hat{\epsilon}_{(i+1)} \end{bmatrix} \sim N\left[\underset{\sim}{0},\ \begin{pmatrix} \sigma^2\left[1+\underset{\sim}{X}'_{i-1}(X'_{(i)}X_{(i)})^{-1}\underset{\sim}{X}_{i-1}\right] & 0 & 0 \\ 0 & \sigma^2\left[1+\underset{\sim}{X}'_i(X'_{(i)}X_{(i)})^{-1}\underset{\sim}{X}_i\right] & 0 \\ 0 & 0 & \sigma^2\left[1+\underset{\sim}{X}'_{i+1}(X'_{(i)}X_{(i)})^{-1}\underset{\sim}{X}_{i+1}\right] \end{pmatrix}\right]$$

$$= N\left(\underset{\sim}{0},\ \sigma^2 A\right)$$

where A is the 3x3 diagnol matrix.

Scanned by CamScanner

Since the variance covariance matrix is $\sigma^2 A$, the precision matrix for $\underset{\sim}{\epsilon}^{(i)}$ is $\left(\frac{1}{\sigma^2} A^{-1}\right)$. Now we know that if $\underset{\sim}{X} \sim N_p(\underset{\sim}{\mu}, \Sigma)$ then $(\underset{\sim}{X}-\underset{\sim}{\mu})' \Sigma^{-1} (\underset{\sim}{X}-\underset{\sim}{\mu}) \sim \chi^2_p$

Applying this here we get

$$\underset{\sim}{\epsilon}^{(i)'} \left(\frac{1}{\sigma^2} A^{-1}\right) \underset{\sim}{\epsilon}^{(i)} \sim \chi^2_3.$$

Now we also know that

$$\frac{\hat{\sigma}^2_{(i)}}{\sigma^2}(n-3-p) \sim \chi^2_{(n-3-p)}$$

Thus

$$\frac{\left(\underset{\sim}{\epsilon}^{(i)'} \left(\frac{1}{\sigma^2} A^{-1}\right) \underset{\sim}{\epsilon}^{(i)}\right) \Big/ 3}{\left(\frac{\hat{\sigma}^2_{(i)} (n-3-p)}{\sigma^2}\right) \Big/ (n-3-p)} \sim F_{3,\, n-3-p}$$

or

$$\frac{\underset{\sim}{\epsilon}^{(i)'} A^{-1} \underset{\sim}{\epsilon}^{(i)}}{3\, \hat{\sigma}^2_{(i)}} \sim F_{3,\, n-3-p}. \longrightarrow (5)$$

Thus our test statistic for testing the null hypothesis that 1st 3 observations come from ~~the~~ the specified linear regression vs. the alternative hypothesis that they are outliers is

$$T = \frac{\underset{\sim}{\epsilon}^{(i)'} A^{-1} \underset{\sim}{\epsilon}^{(i)}}{3 \, \hat{\sigma}_{(i)}^{2}}$$

This use $\hat{\sigma}_{(i)}^{2}$ given by eq. 3.

The distribution of T is $F_{3, n-p-3}$ as derived in 5.

Test: Reject H_0 if $|T| > qf\left(\left(1-\frac{\alpha}{2}\right), (3, n-3-p)\right)$

where $qf\left(\left(1-\frac{\alpha}{2}\right), (3, n-3-p)\right)$ is the $\left(1-\frac{\alpha}{2}\right)$ quantile of F distribution with degrees of freedom 3 and $n-3-p$.

Scanned by CamScanner