Problem 1

Topic :- Linear Regression.

(a) (Theoretical problem) - 5 points :- Find the equation of the regression line that fits the points (0,4), (1,7), (2,9), (3,12) & (5,18) by using and solve the normal equations.

Ans The points given in the vector form as follows:-

$$x = \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 5 \end{bmatrix}, \quad y = \begin{bmatrix} 4 \\ 7 \\ 9 \\ 12 \\ 18 \end{bmatrix}$$

$y = b_0 x_0 + b_1 x_1$ $\quad (x_0 = 1)$ ... (equation of regression line)

Including the x_0 in x the vector becomes

$$X = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 5 \end{bmatrix}$$

Finding b_0 & b_1 using the normal equation

$$B = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = (X^T X)^{-1} X^T y$$

$$X^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 5 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 5 & 11 \\ 11 & 39 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{74} \times \begin{bmatrix} 39 & -11 \\ -11 & 5 \end{bmatrix}$$

$$(X^T X)^{-1} X^T = \frac{1}{74} \begin{bmatrix} 39 & -11 \\ -11 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 3 & 5 \end{bmatrix}$$

$$= \frac{1}{74} \begin{bmatrix} 39 & 28 & 17 & 6 & -16 \\ -11 & -6 & -1 & 4 & 14 \end{bmatrix}$$

$$(X^T X)^{-1} X^T y = \frac{1}{74} \begin{bmatrix} 39 & 28 & 17 & 6 & -16 \\ -11 & -6 & -1 & 4 & 14 \end{bmatrix} \begin{bmatrix} 4 \\ 7 \\ 9 \\ 12 \\ 18 \end{bmatrix}$$

$$\begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \frac{1}{74} \begin{bmatrix} 289 \\ 205 \end{bmatrix} = \begin{bmatrix} 3.90 \\ 2.77 \end{bmatrix}$$

Therefore, regression equation of line is $\boxed{y = 3.9 + 2.77x}$

Scanned by CamScanner

Problem 3.

Topic :: Maximum likelihood Estimate (MLE)

a) In a lab testing, the number of cycles to fail in five tests are $x_i = 25, 20, 28, 33, 26$. In this case, the distribution parameters are λ and ξ, and follows log-normal distribution: $\ln(x) \sim N(\mu, \sigma^2)$. Estimate the distribution parameter using MLE

Ans. The probability density function of log-normal distribution is,

$$f(x|\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log x - \mu)^2}{2\sigma^2}\right]$$

Assumption: Here $\log x = \log_e x$ (base is e)

$$l = \sum_{i=1}^{N} \log\left(\left[\frac{1}{x_i\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\log x_i - \mu)^2}{2\sigma^2}\right]\right]\right) \cdots \boxed{\text{Likelihood}}$$

$$l = \sum_{i=1}^{N} \left[\cancel{\log 1}^{0} - \log x_i - \log(\sqrt{2\pi\sigma^2}) - \frac{(\log x_i - \mu)^2}{2\sigma^2}\right]$$

To maximize likelihood we differentiate l w.r.t μ & σ and equate to zero.

$$\hat{\mu} = \frac{\partial l}{\partial \mu} = +\frac{2}{2\sigma^2} \sum_{i=1}^{N} (\log x_i - \mu) = 0$$

$$N\mu = \sum_{i=1}^{N} \log x_i$$

$$\mu = \frac{1}{N}\left[\sum_{i=1}^{N} \log x_i\right] \quad \text{———— (1)}$$

$$\hat{\sigma} = \frac{\partial l}{\partial \sigma} = \sum_{i=1}^{N}\left[\frac{-1}{\sigma} - \frac{(\log x_i - \mu)^2}{2} \times (-2\sigma^{-3})\right] = 0$$

$$0 = \sum_{i=1}^{N}\left[\frac{1}{\sigma^3}(\log x_i - \mu)^2 - \frac{1}{\sigma}\right]$$

$$0 = \frac{1}{\sigma^3}\sum_{i=1}^{N}(\log x_i - \mu)^2 - \frac{N}{\sigma}$$

$$\frac{N}{\sigma} \times \sigma^3 = \sum_{i=1}^{N}(\log x_i - \mu)^2$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\log x_i - \mu)^2} \quad \text{———— (2)}$$

Substituting the values $x_1 = 25$, $x_2 = 20$, $x_3 = 28$, $x_4 = 33$, $x_5 = 26$ & taking summation of x for $i = 1$ to 5 we get,

$$\mu = \frac{16.3}{5} = 3.26$$

$$\sigma = 0.163$$

$$\boxed{\begin{aligned} \lambda &= \mu = 3.26 \\ \xi &= \sigma^2 = 0.0265 \end{aligned}}$$

3b For two parameter distribution, the probability density function is :

$$f(x) = \theta\sigma^{\theta} x^{(-\theta-1)}$$. What is MLE for σ & θ?

Ans. $f(x) = \theta\sigma^{\theta} x^{(-\theta-1)}$ .... Given in the problem.

$$\text{Likelihood } (L) = \sum_{i=1}^{N} \log\left(\theta\sigma^{\theta} x^{(-\theta-1)}\right)$$

$$= \sum_{i=1}^{N}\left(\log\theta + \theta\log\sigma - (\theta+1)\log x_i\right)$$

$$= N\log\theta + N\theta\log\sigma - \sum_{i=1}^{N}(\theta+1)\log x_i$$

To find MLE, equating $\frac{\partial l}{\partial \theta} = 0$

$$\hat{\theta} = \frac{\partial l}{\partial \theta} = \frac{N}{\theta} + N\log\sigma - \sum_{i=1}^{N}\log x_i = 0$$

$$\frac{N}{\theta} = \sum_{i=1}^{N}\log x_i - N\log\sigma$$

$$\theta = \frac{N}{\left(\sum_{i=1}^{N}\log x_i - N\log\sigma\right)}$$

$\sigma = x_i$ .. As $x_i \geq \sigma$

3c The probability mass function of Poisson distribution is: $P(x) = e^{-\lambda} \frac{\lambda^x}{x!}$. Find MLE for λ

Ans Likelihood $(l) = \sum_{i=1}^{N} \log\left(\frac{e^{-\lambda} \lambda^{x_i}}{x_i!}\right)$

$$= \sum_{i=1}^{N} \left(\log e^{-\lambda} + \log \lambda^{x_i} - \log x_i!\right)$$

$$= \sum_{i=1}^{N} \left(-\lambda + x_i \log \lambda - \log x_i!\right)$$

For MLE for $\lambda \Rightarrow \frac{\partial l}{\partial \lambda} = 0$

$$\frac{\partial l}{\partial \lambda} = \sum_{i=1}^{N} \left[-1 + \frac{x_i}{\lambda}\right]$$

$$0 = -N + \sum_{i=1}^{N} \frac{x_i}{\lambda}$$

$$\boxed{\lambda = \frac{\sum_{i=1}^{N} x_i}{N}}$$