Ex. 3 p. 146

$X_1, \dots, X_n \overset{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$

you want to estimate the .95 percentile

$\tau : \mathbb{P}\{X_i < \tau\} = .95$

a) find $\hat{\tau}_{MLE}$

ANSWER: $\mathbb{P}\{X_i < \tau\} = \mathbb{P}\{ \underbrace{\frac{X_i - \mu}{\sigma}}_{Z \sim \mathcal{N}(0,1)} < \frac{\tau - \mu}{\sigma} \}$

$\Rightarrow 0.95 = \underbrace{\mathbb{P}\{Z < \frac{\tau - \mu}{\sigma}\}}_{\Phi\left(\frac{\tau-\mu}{\sigma}\right)}$ where Φ is the cdf of a $\mathcal{N}(0,1)$

Recall: $\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du.$

$\Rightarrow \frac{\tau - \mu}{\sigma} = \Phi^{-1}(0.95) \Rightarrow \tau = \mu + \sigma \Phi^{-1}(.95)$

$\hat{\tau}_{ML} = \hat{\mu}_{ML} + \hat{\sigma}_{ML} \Phi^{-1}(.95)$

b) Find a $1-\alpha$ CI (approx.) for τ.

Here $\theta = \{(\mu, \sigma)\}$

Recall: A vector $Z = \begin{pmatrix} Z_1 \\ \vdots \\ Z_n \end{pmatrix}$ of random variables, is called random vector.

i) $\mathbb{E}(Z) := \begin{pmatrix} \mathbb{E}(Z_1) \\ \mathbb{E}(Z_2) \\ \vdots \\ \mathbb{E}(Z_n) \end{pmatrix}$

ii) $\mathbb{V}ar(Z) := \mathbb{E}[\underbrace{(Z - \mathbb{E}(Z))}_{n \times 1}\underbrace{(Z - \mathbb{E}(Z))^T}_{1 \times n}]$

$\Rightarrow \mathbb{V}ar(Z) \in \mathbb{R}^{+\, n \times n}$ and

$[\mathbb{V}ar(Z)]_{ij} = Cov(Z_i, Z_j)$

Our $\tau = \mu + \sigma \Phi^{-1}(.95) =: g(\mu, \sigma)$

So $\hat{\tau}_{ML} = g(\hat{\mu}_{ML}, \hat{\sigma}_{ML})$ (equivariance of ML)

$\Rightarrow \hat{\tau}_{ML} - \tau = g(\hat{\mu}_{ML}, \hat{\sigma}_{ML}) - g(\mu, \sigma)$

$\cong \nabla^T g(\mu, \sigma)[(\hat{\mu}_{ML}, \hat{\sigma}_{ML}) - (\mu, \sigma)]$

Now: $(\hat{\mu}_{ML}, \hat{\sigma}_{ML}) - (\mu, \sigma) \rightsquigarrow \mathcal{N}(0, I_n^{-1}(\mu, \sigma))$

$\hat{\tau}_{ML} - \tau \overset{d}{\rightsquigarrow} \mathcal{N}(0, \nabla^T g(\mu, \sigma) I_n^{-1}(\mu, \sigma) \nabla g(\mu, \sigma))$ *

Last time I showed you that

$$I_n(\mu,\sigma) = \begin{bmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{bmatrix}$$

$$\Rightarrow I_n^{-1}(\mu,\sigma) = \frac{1}{n}\begin{pmatrix} \sigma^2 & 0 \\ 0 & \frac{\sigma^2}{2} \end{pmatrix}$$

From $\ast$ it follows that

$$\left[\hat{\tau}_{ML} \pm z_{\frac{\alpha}{2}} \cdot \sqrt{\nabla g(\mu,\sigma)^T I_n^{-1}(\mu,\sigma) \nabla g(\mu,\sigma)}\right]$$

is an asymptotic $1-\alpha$ CI for τ

$$g(\mu,\sigma) = \mu + \sigma\,\Phi^{-1}(.95)$$

$$\Rightarrow \nabla g(\mu,\sigma) = \begin{bmatrix} 1 \\ \Phi^{-1}(.95) \end{bmatrix}$$

$$\Rightarrow sd(\hat{\tau}_{ML}) = \sqrt{\frac{1}{n}\left[\sigma^2 + \left[\Phi^{-1}(.95)\right]^2 \frac{\sigma^2}{2}\right]}$$

Since σ is unknown you say

$$\hat{sd}(\hat{\tau}_{ML}) = \left[\frac{1}{n}\left(\hat{\sigma}^2_{ML} + \left[\Phi^{-1}(.95)\right]^2 \frac{\hat{\sigma}^2_{ML}}{2}\right)\right]^{1/2}$$

Linear Regression Model

$$Y = X\beta + \varepsilon \qquad (1)$$

you have N individuals

$\Rightarrow Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}$ (response vector)

and p features collected into the design matrix $X \in \mathbb{R}^{N \times p}$:

for the i-th individual, you observe feature $(x_{i1}, x_{i2}, \dots, x_{ip})$, the i-th row of X

From (1):

$$y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$$

$E(\varepsilon_i) \overset{Hp}{=} 0$

$Var(\varepsilon_i) \overset{Hp}{=} \sigma^2$

$\varepsilon_i \perp \varepsilon_j$

$\Rightarrow \beta = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$ unknown parameters

$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{pmatrix}$ is the residual vector

Additional Hps:

i) $X^T X$ is invertible (rank $(X) = p$)

ii) $\varepsilon_i \overset{i.i.d}{\sim} \mathcal{N}(0, \sigma^2) \Rightarrow \varepsilon \sim \mathcal{N}(0, \sigma^2 I_m)$

Main problem is :

- β, σ^2 are unknown and should be estimated

For a given $\tilde{\beta}$, your guess of y will be $\tilde{y} = X\tilde{\beta}$

$\Rightarrow \tilde{\varepsilon} = y - \tilde{y}$

idea : $\min\limits_{\tilde{\beta}} \sum\limits_{i=1}^{N} \tilde{\varepsilon}_i^2 = \min\limits_{\tilde{\beta}} \|\tilde{\varepsilon}\|^2$

$\Rightarrow \min\limits_{\tilde{\beta}} \|y - X\tilde{\beta}\|^2 = \min\limits_{\tilde{\beta}} (y - X\tilde{\beta})^T (y - X\tilde{\beta})$

$= \min\limits_{\tilde{\beta}} [\underbrace{y^T y + \tilde{\beta}^T X^T X \tilde{\beta} - 2\tilde{\beta}^T X^T y}_{g(\tilde{\beta})}]$

$$\nabla_{\tilde\beta}\, g(\tilde\beta) = \cancel{2}\, X^T X \tilde\beta - \cancel{2}\, X^T y = 0_{\mathbb{R}^p}$$

$$\Longleftrightarrow \quad X^T X \tilde\beta = X^T y$$

$$\Longleftrightarrow \quad \tilde\beta = (X^T X)^{-1} X^T y$$

$$\hat\beta_{OLS} = \tilde\beta$$