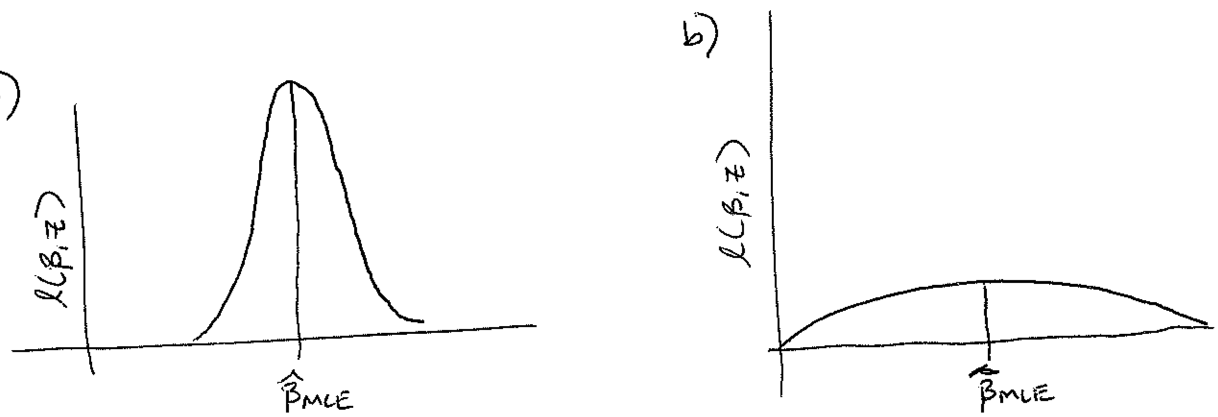
β: univariate

z: one datum

Which $\widehat{\beta}_{MLE}$ is more reliable?

a) b)



more peaked flatter

$\Rightarrow$ curvature aka second derivative info aka the Hessian of the log-likelihood tells us about how much information we have about $\widehat{\beta}_{MLE}$.

Vocab

Score - derivative of log-likelihood

observed information = $-H$ (negative of Hessian of log-likelihood)

Fisher information : $\mathbb{E}$[obs information] or equivalently, var[score].

parameter

datum

$$f(\beta, z) = \text{likelihood}$$

$$\log f(\beta, z) = \text{log-likelihood}$$

Let's look at $\mathbb{E}$[score] :

$$\mathbb{E}_z\left[\frac{d}{d\beta} \log f(\beta, z)\right] = \int \frac{d}{d\beta} \log f(\beta, z) \cdot f(\beta, z)\, dz$$

$$= \int \frac{1}{\cancel{f(\beta, z)}} \cdot f'(\beta, z) \cdot \cancel{f(\beta, z)}\, dz$$

$$= \int \frac{d}{d\beta} f(\beta, z)\, dz$$

$$= \frac{d}{d\beta} \underbrace{\int f(\beta, z)\, dz}_{} \quad (*)$$

$$= \frac{d}{d\beta} 1 = 0$$

(*) Leibniz rule

$$\int f(\beta, z) \overbrace{\frac{\partial}{\partial \beta} \log f(\beta, z)}^{\text{score}} \, dz = 0$$

Differentiate both sides w.r.t. β:

$$\int \frac{\partial f(\beta, z)}{\partial \beta} \left[\frac{\partial}{\partial \beta} \log f(\beta, z) \right] + f(\beta, z) \left[\frac{\partial^2}{\partial \beta^2} \log f(\beta, z) \right] dz = 0$$

Notice: $\frac{\partial f(\beta, z)}{\partial \beta} = f(\beta, z) \cdot \frac{\partial}{\partial \beta} \log f(\beta, z)$

Plugging in: we have

$$\underbrace{\int f(\beta, z) \left[\frac{\partial}{\partial \beta} \log f(\beta, z) \right]^2 dz}_{} = \underbrace{- \int f(\beta, z) \left[\overbrace{\frac{\partial^2}{\partial \beta^2} \log f(\beta, z)}^{H} \right] dz}_{-\mathbb{E}[H]}$$

$$\mathbb{E}\left[\frac{\partial}{\partial \beta} \log f(\beta, z)^2 \right]$$

$$= \mathbb{E}[-H]$$

$$= \mathbb{E}[\text{obs info}]$$

$$\mathbb{E}[\text{score}^2]$$

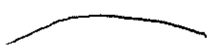
$$= \text{var}[\text{score}]$$

because $\mathbb{E}[\text{score}] = 0$

& recall $\text{var}[\text{score}] = \mathbb{E}[\text{score}^2] - (\mathbb{E}[\text{score}])^2$

Interpretation

(1) var(score)



- low variance score $\Rightarrow$ likelihood barely changes w/ parameter.

$\Rightarrow$ data not informative

- high variance $\Rightarrow$ data strongly identify β

(2) $\mathbb{E}$[obs info]

= $\mathbb{E}$[-Hessian] is info about curvature around true parameter

Flat = low info

Highly curved $\Rightarrow$ more peaked = high info.

i.e. there exists a rapid drop in likelihood for small $\Delta\beta$.