Comparing models

- Have f_j, $j=1,\dots,J$ with $\{f_j = f_{\theta_j} : \theta_j \in \Theta_j\}$, $\dim(\Theta_j) = p_j$

- $\hat\theta_j$ is the maximizer of $\sum_{i=1}^{n} \log f_{\theta_j}(X_i) = \ell_{f_j}(\theta_j)$

- maximum value $\ell_{f_j}(\hat\theta_j)$

- θ_{j0} is the maximizer of $E_g[\log f_{\theta_j}(X)]$

- Criterion used to compare f_j:

$$Q_j = \underset{\substack{\text{over}\\ X \text{ and}\\ \hat\theta_j = \hat\theta_j(X^n)}}{E}\left[\log f_j(X \mid \hat\theta_j)\right] = E\left[\log f_j(X \mid \hat\theta_j(X^n))\right]$$

- We know

$$\sqrt{n}(\hat\theta_j - \theta_{j0}) \to^d N(0, FI_j(\theta_{j0})^{-1} J_j(\theta_{j0}) FI_j(\theta_{j0})^{-1})$$

We will work toward an estimate of θ_j.

We will start by examining the estimation variance

$\rightarrow$ how $\hat{\theta}_j$ varies for different samples of x^n.

- We have $l_{f_j}(\theta_j) = \sum_{i=1}^{n} \log f_j(x_i|\theta_j)$.

- Expand $l_{f_j}(\theta_j)$ near $\hat{\theta}_j$

$$l_{f_j}(\theta_j) = l_{f_j}(\hat{\theta}_j) + \overbrace{\frac{\partial}{\partial \theta_j} l_{f_j}(\theta_j)\Big|_{\hat{\theta}_j}}^{=0 \text{ by definition}} (\theta_j - \hat{\theta}_j)$$

$$+ \frac{1}{2}(\theta_j - \hat{\theta}_j)' \left(\frac{\partial^2}{\partial \theta_j \partial \theta_j'} l_{f_j}(\theta_j)\Big|_{\theta_j^*} \right) (\theta_j - \hat{\theta}_j)$$

where $\|\theta_j - \theta_j^*\| \leq \|\theta_j - \hat{\theta}_j\|$

$$= l_{f_j}(\hat{\theta}_j) - \frac{n}{2}(\theta_j - \hat{\theta}_j)' \left[** \right] (\theta_j - \hat{\theta}_j)$$

where $** = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial \theta_j \partial \theta_j'} \log f_j(x_i|\theta_j)\Big|_{\theta_j^*}$

$$\longrightarrow E\left[-\frac{\partial^2}{\partial \theta_j \partial \theta_j'} \log f_j(X|\theta_j)\Big|_{\theta_{j0}} \right] = FI_j(\theta_{j0})$$

So we have

$$l_{f_j}(\theta_j) \simeq l_f(\hat{\theta}_j) - \frac{n}{2} \underbrace{(\theta_j - \hat{\theta}_j)' \mathcal{FI}_j(\theta_{j0}) (\theta_j - \hat{\theta}_j)}_{\text{Can we simplify this?}}$$

Look at $\theta_j = \theta_{j0}$

- $$n\,(\hat{\theta}_j - \theta_{j0})' \,\mathcal{FI}_j(\theta_{j0})\, (\hat{\theta}_j - \theta_{j0})$$

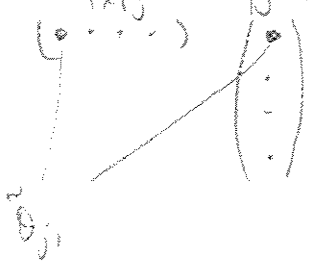
$$= \left(\sqrt{n}\, \mathcal{FI}_j(\theta_{j0})^{1/2} (\hat{\theta}_j - \theta_{j0})\right)' \left(\sqrt{n}\, \mathcal{FI}_j(\theta_{j0})^{1/2} (\hat{\theta}_j - \theta_{j0})\right)$$

$$= \tilde{\theta}_j' \tilde{\theta}_j \qquad \text{where} \quad \tilde{\theta}_j = \sqrt{n}\, \mathcal{FI}_j(\theta_{j0})^{1/2} (\hat{\theta}_j - \theta_{j0})$$

- Think of $\tilde{\theta}_j$ as a 'standardized' version of $\hat{\theta}_j$

dim($\tilde{\theta}_j$)

$$\tilde{\theta}_j' \tilde{\theta}_j = \sum_{p=1}^{p_j} \tilde{\theta}_{jp}^2$$



sum of squares of components of $\tilde{\theta}_j$ vector

Now $\tilde{\theta}_j' \tilde{\theta}_j = \sum_{p=1}^{p_j} \tilde{\theta}_{jp}^2$ = sum of diagonal elements of $\underbrace{\underset{p_j \times 1}{\tilde{\theta}_j} \; \underset{1 \times p_j}{\tilde{\theta}_j'}}_{p_j \times p_j \text{ matrix}}$ (4)

$$\begin{pmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{pmatrix} \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \end{pmatrix} = \begin{pmatrix} \theta_{j1}^2 & \theta_{j1}\theta_{j2} & \cdots \\ & \theta_{j2}^2 & \\ & & \ddots \end{pmatrix}$$

Trick: We are interested in $E[\tilde{\theta}_j' \tilde{\theta}_j]$

but from above $E[\tilde{\theta}_j' \tilde{\theta}_j] = \text{Trace}\, E[\tilde{\theta}_j \tilde{\theta}_j']$

Here

$$\text{Trace}\, E(\tilde{\theta}_j \tilde{\theta}_j') = \text{Trace}\, E\left[n\, FI_j(\theta_{j0}) (\hat{\theta}_j - \theta_{j0}) (\hat{\theta}_j - \theta_{j0})'\right]$$

$$= \text{Trace}\left[n\, FI_j(\theta_{j0})\, \underbrace{E\left[(\hat{\theta}_j - \theta_{j0})(\hat{\theta}_j - \theta_{j0})'\right]}_{\text{Cov}(\hat{\theta}_j) = \frac{FI_j(\theta_{j0})^{-1} J_j(\theta_{j0}) FI_j(\theta_{j0})^{-1}}{n}}\right]$$

J

So

$$E\left(n\,(\hat{\theta}_j-\theta_{j0})'\,FI_j(\theta_{j0})\,(\hat{\theta}_j-\theta_{j0}) \right)$$

$$= \text{Trace}\left[n\,FI_j(\theta_{j0}) \left(\frac{FI_j(\theta_{j0})^{-1}\,J_j(\theta_{j0})\,FI_j(\theta_{j0})^{-1}}{n} \right) \right]$$

$$= \text{Trace}\left[J_j(\theta_{j0})\,FI_j(\theta_{j0})^{-1} \right]$$

Back to our approximation

$$l_{f_j}(\theta_j) \approx l_{f_j}(\hat{\theta}_j) - \frac{n}{2}(\hat{\theta}_j-\theta_j)\,FI_j(\theta_{j0})\,(\hat{\theta}_j-\theta_j)$$

And specifically at θ_{j0}

$$l_{f_j}(\theta_{j0}) \approx l_{f_j}(\hat{\theta}_j) - \frac{n}{2}\underbrace{(\hat{\theta}_j-\theta_{j0})\,FI_j(\theta_{j0})\,(\hat{\theta}_j-\theta_{j0})}_{(\bullet\bullet\bullet)}$$

$$\approx l_f(\hat{\theta}_j) - \frac{1}{2}\text{Trace}\{J_j(\theta_{j0})\,FI_j(\theta_{j0})\}$$

replacing $(\bullet\bullet\bullet)$
by its expectation

So

$$l_{f_j}(\theta_{j0}) = \sum_{i=1}^{n} \log f_j(x_i \mid \theta_{j0})$$

$$\approx \sum_{i=1}^{n} \log f_j(x_i \mid \hat{\theta}_j) - \frac{1}{2} \mathrm{Trace}\{ J_j(\theta_{j0}) \{ I_j(\theta_{j0}) \}^{-1} \}$$

$\downarrow$ take expectation over X^n

$$E\left(\sum_{i=1}^{n} \log f_j(X_i \mid \theta_{j0}) \right) \approx E\left(l_f(\hat{\theta}_j) \right) - \frac{1}{2} \mathrm{Trace}\{ J_j(\theta_{j0}) \{ I_j(\theta_{j0}) \}^{-1} \}$$

$$\parallel$$

$$n E\left[\log f_j(X \mid \theta_{j0}) \right]$$

Remember, we are after

$$Q_j = E\left[\log f_j(X \mid \hat{\theta}_j(X^n)) \right]$$

- We have 'done' the training part (E over X^n)
- Now, do testing.