We want to expand $E[\log f_j(X|\theta_j)]$ near θ_{j0}
(Here all E's are over $X \sim g$)

$$E\left[\log f_j(X|\theta_j)\right] = E\left[\log f_j(X|\theta_{j0})\right]$$

$$+ \underbrace{\frac{\partial}{\partial\theta_j}\left(E \log f_j(X|\theta_j)\right)\Big|_{\theta_{j0}}}_{=0 \text{ by definition of } \theta_{j0}} (\theta_j - \theta_{j0})$$

$$+\frac{1}{2}(\theta_j-\theta_{j0})'\frac{\partial^2}{\partial\theta_j\partial\theta_j'}E\log f_j(X|\theta_j)\Big|_{\theta^*}(\theta_j-\theta_{j0})$$

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

$$-\frac{1}{2}(\theta_j-\theta_{j0})'\underbrace{E\left(\frac{-\partial^2}{\partial\theta_j\partial\theta_j'}\log f_j(X|\theta_j)\Big|_{\theta_{j0}}\right)}_{\mathcal{I}_j(\theta_{j0})}(\theta_j-\theta_{j0})$$

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

$$-\frac{1}{2n}\left[n(\theta_j-\theta_{j0})'\mathcal{I}_j(\theta_{j0})(\theta_j-\theta_{j0})\right]$$

(consider $\theta_j = \hat{\theta}_j$ especially. (thinking of $\hat{\theta}_j$ as number)

Here $\underset{\text{over } X}{E}\left[\log f_j(X|\hat{\theta}_j)\right] \approx E\left[\log f_j(X|\theta_{j0})\right] - \frac{1}{2n}\left[n(\hat{\theta}_j-\theta_{j0})'\mathcal{I}_j(\theta_{j0})(\hat{\theta}_j-\theta_{j0})\right]$

Now take expectation ~~over~~ X^n (i.e. over $\hat{\theta}_j$) ⑧

$$\underset{X^n}{E}\left(E_X\left[\log f_j(X|\hat{\theta}_j(X^n))\right]\right) \approx \underset{X}{E}\left(\log f_j(X|\theta_{j0})\right)$$

$$\underbrace{\qquad}_{\parallel}$$

$$Q_j \qquad -\frac{1}{2n}\text{Trace}\left\{J_j(\theta_{j0})FI_j(\theta_{j0})^{-1}\right\}$$

Our validation criterion

Compare with result from before (page 6)

$$nE\left(\log f_j(X|\theta_{j0})\right) \approx E\left(l_f(\hat{\theta}_j)\right) - \frac{1}{2}\text{Trace}\left\{J_j(\theta_{j0})FI_j(\theta_{j0})^{-1}\right\}$$

$$E\left(\sum_{i=1}^n \log f_{\hat{\theta}_j}(x_i)\right)$$

$$= E\left(\sum_{i=1}^n \log f_j(x_i|\hat{\theta}_j(x^n))\right)$$

Average <u>training</u> loglikelihood at MLE for each ~~training~~ set

$$\Rightarrow nQ_j \approx \underbrace{E\left[\sum_{i=1}^n \log f_j(x_i|\hat{\theta}_j(x^n))\right]}_{} - \text{Trace}\left\{J_j(\theta_{j0})FI_j(\theta_{j0})^{-1}\right\}$$

This average training loglik is <u>estimated by the observed</u>

$$\sum_{i=1}^n \log f_j(x_i|\hat{\theta}_j(x^n))$$

We thus arrive at the criterion

$$n Q_j \hat{=} \sum_{i=1}^{n} \log f_j(x_i; \hat{\theta}_j(x^n)) - \mathrm{Trace}\{ J_j(\theta_{j0}) FI_j(\theta_{j0})^{-1} \}$$

$$= \ell_f(\hat{\theta}_j(x^n)|x^n) - \underbrace{\mathrm{Trace}\{ J_j(\theta_{j0}) FI_j(\theta_{j0})^{-1} \}}$$

this depends on g

and on θ_{j0} - so not

very practical

but if we assume

f_j is not too far from g

then $J_j(\theta_{j0}) \approx FI_j(\theta_{j0})$

$$\Rightarrow n Q_j \hat{=} \ell_f(\hat{\theta}_j(x^n)|x^n) - \mathrm{Trace}\{ I_{p_j \times p_j} \}$$

$$= \ell_f(\hat{\theta}_j(x^n)|x^n) - p_j$$

Usually see this as the $AIC_j = -2n Q_j$

$$AIC_j = \underbrace{-2 \log L_f(\hat{\theta}_j(x^n)|x^n)}_{\text{deviance}} + \underbrace{2 p_j}_{\text{penalty on model complexity}}$$

Strategy → find for each f_j maximizer $\hat{\theta}_j$

→ evaluate maximum loglike $\log L_j(\hat{\theta}_j | x^n)$

→ evaluate $AIC_j = \text{Deviance}_j + 2p_j$

→ Choose f_j with minimum AIC

- Alternative – use cross-validation procedure

* Estimate θ_j by creating a training & test data from data available

* Example Leave-one-out CV or jacknife

$$\hat{Q}_j = \frac{1}{n} \sum_{i=1}^{n} \log f_j(x_i | \underbrace{\hat{\theta}_j(x^{-i})}) = LOOCV_j$$

$\hat{\theta}_j(x^{-i})$ is maximizer of

$$\sum_{i' \neq i} \log f_j(x_{i'} | \theta_j)$$

Can show $AIC_j \approx -2n\, LOOCV_j$

* Can try Leave-v-out procedures as well
* High variance of $\hat{\theta}_j$ estimate a problem for LOOCV