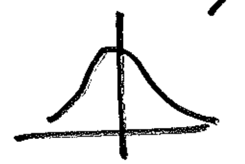
next time: simulation-based computation

read! G. et al. ch. 4, 5, 10

hierarchical models soon

$k = 1$

$\ell(\theta \mid y\,\mathcal{B})$, $\theta \in \mathbb{R}$



$k > 1$ $\quad \underset{\sim}{\theta} = (\theta_1, \dots, \theta_k)$, $\in \mathbb{R}^n$

$\ell\ell(\underset{\sim}{\theta} \mid \underset{\sim}{y}\,\mathcal{B})$, $\in \mathbb{R}^k$

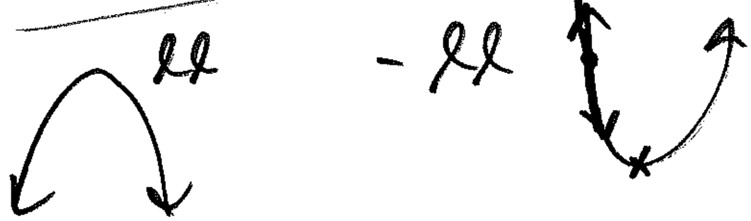
to find MLEs:

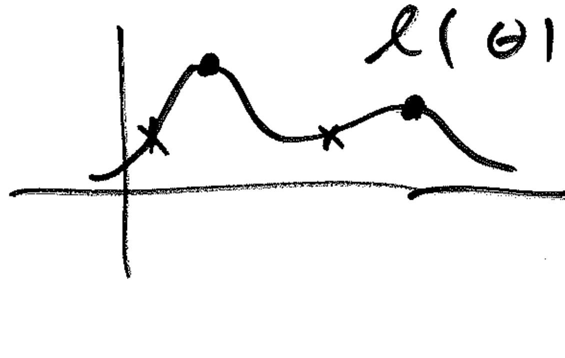
$$\left\{ \begin{array}{l} \frac{\partial}{\partial \theta_1} \ell\ell(\underset{\sim}{\theta} \mid y\,\mathcal{B}) = 0 \\ \frac{\partial}{\partial \theta_2} \ell\ell(\underset{\sim}{\theta} \mid y\,\mathcal{B}) = 0 \\ \vdots \\ \frac{\partial}{\partial \theta_k} \ell\ell(\underset{\sim}{\theta} \mid y\,\mathcal{B}) = 0 \end{array} \right.$$

(the likelihood equations)

$$\nabla \ell\ell(\underset{\sim}{\theta} \mid y\,\mathcal{B})$$

$$\left(\frac{\partial}{\partial \theta_1} \ell\ell, \dots, \frac{\partial}{\partial \theta_k} \ell\ell \right) = \underset{\sim}{0}$$





$k=1$ $$\hat{I} = \left[-\frac{\partial^2}{\partial\theta} \ell\ell(\theta \mid x, y) \right]_{\theta = \hat{\theta}_{MLE}}$$

$$\widehat{SE}_{RS}(\hat{\theta}_{MLE}) = \sqrt{\frac{1}{\hat{I}}} = \sqrt{\hat{I}^{-1}} = \sqrt{\hat{V}_{RS}(\hat{\theta}_{MLE})}$$

$$V_{RS}(\hat{\theta}) = \hat{I}^{-1}$$

$$\begin{array}{c|cccc} & \theta_1 & \theta_2 & \cdots & \theta_k \\ \hline \theta_1 & \frac{\partial^2}{\partial\theta_1^2} & \frac{\partial^2}{\partial\theta_1\partial\theta_2} & - & - \\ \theta_2 & \frac{\partial^2}{\partial\theta_2\partial\theta_1} & \ddots & & \\ \vdots & \vdots & & \ddots & \\ \theta_k & \vdots & & & \ddots \end{array}$$

symmetric

$\ell\ell$ = Hessian of $\ell\ell$

(Otto Hesse)

$$\hat{I}_{k \times k} = \left[-\text{Hessian of } \ell\ell \right]_{\theta = \hat{\theta}_{MLE}}$$

symmetric

$$\begin{array}{c|cccc} & \hat{\theta}_1 & \hat{\theta}_2 & \cdots & \hat{\theta}_k \\ \hline \hat{\theta}_1 & \hat{C}(\hat{\theta}_1, \hat{\theta}_1) = V(\hat{\theta}_1) & & & \\ \hat{\theta}_2 & & & \hat{C}(\hat{\theta}_i, \hat{\theta}_j) & \\ \vdots & & & & \\ \hat{\theta}_k & & & & \end{array}$$

$\hat{\Sigma}_{\hat{\theta}_{MLE}}$

covariance matrix of $\hat{\theta}_{MLE}$

covariance $C(\hat{\theta}_2, \hat{\theta}_1)$

covariance

$$C(X, Y) = E(XY) - E(X)E(Y)$$

$$\underset{\sim MLE}{\Sigma_{\hat{\theta}}} = \hat{I}^{-1}$$ ← inverse of Fisher info matrix

---

$$\hat{SE}_{RS}(\hat{\theta}_j) = \sqrt{\text{diag entry } j \text{ of } \Sigma}$$

$k = 1$

$$\hat{\theta}_{MLE} \underset{RS}{\dot{\sim}} N\left(\theta, \frac{1}{\hat{I}}\right)$$

$k \geq 1$

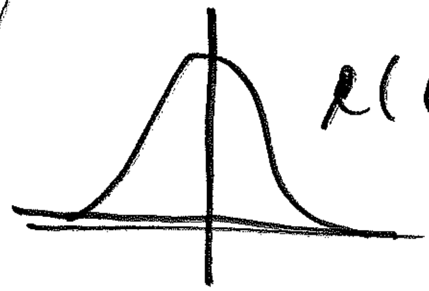
$$\underset{k \times 1}{\hat{\underset{\sim}{\theta}}_{MLE}} \underset{RS}{\dot{\sim}} N_k\left(\underset{k \times 1}{\underset{\sim}{\theta}}, \underset{k \times k}{\hat{I}^{-1}}\right)$$

95% CI for θ_j:

$$(\hat{\theta}_j)_{MLE} \pm 1.96\, \hat{SE}(\hat{\theta}_j)$$

$$\text{bias}\left(\hat{\underset{\sim}{\theta}}_{MLE}\right) = \underset{\sim}{O}\left(\frac{1}{n}\right)$$

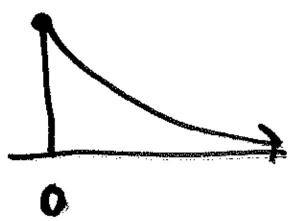
$(k=1)$



$\ell(\theta \mid y\ \mathcal{B})$

$\hat{\theta}_{MLE}$ = mode of posterior dist.

= posterior mean of θ



$\ell(\sigma \mid y\ \mathcal{B})$

if ℓ is strongly non-normal, better from freq. point of view to integrate over ℓ not maximize it

---

CS4: NB10 data

$(\mu\ \sigma\ \nu \mid ?\ \mathcal{B}) \sim p(\mu\ \sigma\ \nu \mid ?\ \mathcal{B})$

$(Y_i \mid \mu\ \sigma\ \nu\ \mathbb{I}\ \mathcal{B}) \stackrel{IID}{\sim} t_\nu(\mu, \sigma^2)$

$(i = 1, \dots, n)$

μ: location — true weight of NB10

$\sigma \leftarrow$ scale (spread)

ν shape (tail-weight)

$\begin{bmatrix} 385 \\ 407 \\ \vdots \end{bmatrix}$ $n = 100$

GIR?

$\underset{\sim}{\theta} = (\mu, \sigma, \nu)$

$\sigma > 0$, $\nu > 0$

there is no conjugate prior for $\tilde{\theta}$

need new way to compute

$\leftarrow$ t sampling dist

$p(\tilde{\theta} \mid y\,\mathcal{T}\,\mathcal{B})$

$\downarrow$

$p(\mu \mid y\,\mathcal{T}\,\mathcal{B}) = ?$

$\ell(\tilde{\theta} \mid y\,\mathcal{T}\,\mathcal{B})$

$p(\tilde{\theta} \mid \mathcal{B})$

$$\boxed{p(\tilde{\theta} \mid y\,\mathcal{T}\,\mathcal{B})} = \boxed{c}\; \underline{p(\theta \mid \mathcal{B})} \cdot \underline{\ell(\tilde{\theta} \mid y\,\mathcal{T}\,\mathcal{B})}$$

$$p(\mu \mid y\,\mathcal{T}\,\mathcal{B}) = \iint p(\mu\,\sigma\,\nu \mid y\,\mathcal{T}\,\mathcal{B})\, d\sigma\, d\nu$$

THT2, ~~2(A)~~

$\tilde{\theta} = (\theta_1, \theta_2, \theta_3)$

Bysh Dnk.

$$\begin{array}{c|c|cc}
 & \theta_1 & \theta_2 & \theta_3 \\
1 & \theta_{11}^{*} & \theta_{12}^{*} & \theta_{13}^{*} \\
2 & \theta_{21}^{*} & \cdots & \\
 & \vdots & & \\
M & & & \\
 & \theta_1^{*} & &
\end{array}$$

$\leftarrow$ monte carlo data set