M	T	W	R	F
	8		10	11 ← lecture in 410 not in 450
	$\boxed{15}$ ✓ usual		~~17~~	NR ~~18~~ away
21	22 test		24	?
	29		31	?
	5		7	?

office hr
3-5

email coming

- optimization
- nonlinear least squares (extends linear regression, different than glm)
  Ch. 8.1-8.4
- mixed & random effects Ch. 10
- numerical integration
- Bayesian inference

---

3 (a) given $y \sim$ Cauchy$(\mu, \sigma = 1)$ ← assume.

[(*) rcauchy(n = 10, 7, 5 ?) ← data]

your choice

fit model $\hat{\mu}$ plot log-likelihood

N-R ↑ to solve

repeat (*) a few times

(b) simulation : repeat (a) 100 times

$$n=5 \quad \left\{ \begin{array}{l} y_1 \\ y_2 \\ \vdots \\ y_{100} \end{array} \right.$$

all random r. from same Cauchy

w same n, same μ

$\hat\mu_1 \cdots \hat\mu_{100}$ $\quad \frac{1}{100}\sum_{i=1}^{100}\hat\mu_i \overset{?}{=} \mu$

$$\frac{1}{99}\sum(\hat\mu_i - \bar{\hat\mu})^2 \overset{?}{=} (-E\,\ell''(\mu))^{-1}$$

or $\dfrac{\sum^N \ell''(\hat\mu_i)}{N}$

$E\{-\ell''(\mu)\}$ exp'd F. info.

$\frac{1}{N}\sum \ell''(\hat\mu)$ estimates the observed F info.

$n=10 \qquad 20 \qquad \cdots \qquad n\to\infty \quad \bar{\hat\mu} \neq \mu$

simulation error

$$\left.\begin{array}{l} \operatorname{var}\hat\mu \simeq \{i(\hat\mu)\}^{-1} \\ \phantom{\operatorname{var}\hat\mu} \simeq \{i(\mu)\}^{-1} \\ \phantom{\operatorname{var}\hat\mu} \simeq \{j(\hat\mu)\}^{-1} \\ \phantom{\operatorname{var}\hat\mu} \simeq \{j(\mu)\}^{-1} \end{array}\right] *$$

$N\to\infty$

$$\lim_{\substack{N\to\infty \\ n\to\infty}} \bar{\hat\mu} = \mu$$

$$-\ell''(\mu) = -\frac{\partial^2}{\partial\mu^2}\sum_{i=1}^{n}\log f(y_i;\mu)$$

$$E[\quad] = \cdots = n\,E\,\frac{\partial^2 \log f(y_i;\mu)}{\partial\mu^2}$$

1. - dimensional optimization or <u>root - finding</u>.

$$\max_{\theta} \ell(\theta) \quad \text{or} \quad \ell'(\theta) = 0$$

$\longleftrightarrow$
? same ?

a) simple iteration $\leftarrow$ guaranteed to converge (under weak cond$^{=}$s) ; slow

b) N - Raphson $\leftarrow$ fast ; conv. guaranteed under <u>strong</u> cond$^{=}$s *

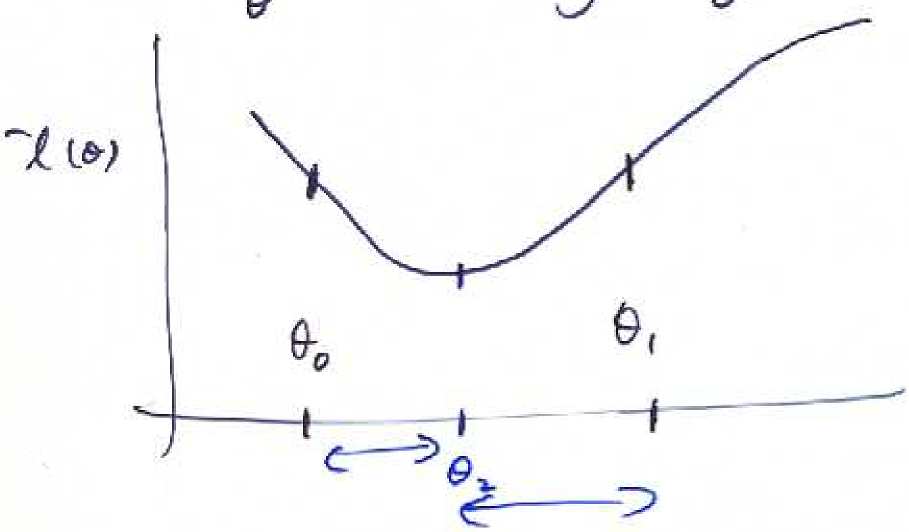
c) bracketing method $\leftarrow$ a bit faster than (a), usually converges

* need <u>really</u> good starting values

(R uniroot)

- these were described as solving $\ell'(\hat{\theta}) = 0$
- but (c) can be used without derivatives to find a min or max (R : optimize),

$\max_{\theta} \ell(\theta)$ by $\min_{\theta} -\ell(\theta)$



check (θ_0, θ_2) brackets min
if not (θ_2, θ_1) does
bisect that etc.

- clever variations on all of these
- in statistics if we really only have $\theta \in \mathbb{R}$, any methods work reasonably well, & quickly
- no standard fix for $\exists$ multiple minima or maxima ; you have to check
- no gold standard for stopping

(*) • $\left| \hat{\theta}^{(t)} - \hat{\theta}^{(t-1)} \right| < \varepsilon$ = "tolerance"

• $\left| \dfrac{\hat{\theta}^{(t)} - \hat{\theta}^{(t-1)}}{\hat{\theta}^{(t-1)}} \right| < \varepsilon$ relative error

(**) • $\left| \ell(\hat{\theta}^{(t)}) - \ell(\hat{\theta}^{(t-1)}) \right| < \varepsilon$ function is converging

(*) usually default
(**) should be checked

NR at convergence, we have $\ell''(\hat{\theta})$ already computed — know a var. of mle

2. multi-dimensional root finding

N-R $\quad \underset{p\times 1}{\hat{\underset{\sim}{\theta}}^{(t+1)}} = \underset{p\times 1}{\hat{\underset{\sim}{\theta}}^{(t)}} - \underset{p\times p}{\left[\ell''(\hat{\underset{\sim}{\theta}}^{(t)})\right]^{-1}} \underset{p\times 1}{\underset{\sim}{\ell}'(\hat{\underset{\sim}{\theta}}^{(t)})}$

- computation & inversion of $\ell''(\hat{\theta}^{(t)})$ can be expensive

- in theory can be computed numerically

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}$$

$$\hat{f}'(x) \simeq \frac{f(x+h)-f(x)}{h} \quad \text{some small } h$$

our f is ℓ'

$\cancel{x}\,x$ is θ

$\cancel{\ell'(\hat{\underset{\sim}{\theta}}^{(t)})}$ $\quad \ell'(\hat{\theta}_1^{(t)}, \hat{\theta}_2^{(t)})$

finite differences

$$\frac{\ell'(\hat{\theta}_1^{(t)}+h_1, \hat{\theta}_2^{(t)})}{h_1} \quad \text{est } \frac{\partial}{\partial\theta_1}$$

$$\frac{\ell'(\hat{\theta}_1^{(t)}, \hat{\theta}_2^{(t)}+h_2)}{h_2} \quad \text{est. } \frac{\partial}{\partial\theta_2}$$

$\exists$ many, many clever improvements

- most are more efficient way to get $\ell''(\hat{\theta}^{(c)})$

- online textbook Robert Gray : Ch. 4 on non linear methods

Bible is "Numerical Recipes" Press et al.

$i = 1, \ldots, n$

model $\quad y_i \sim f(\cdot\, ; \mu_1, \mu_2, \sigma_1, \sigma_2, \pi)$

w prob $\pi \quad y_i \sim N(\mu_1, \sigma_1^2)$

$(1-\pi) \quad \sim N(\mu_2, \sigma_2^2)$

$$f(y_i) = \pi \cdot \frac{1}{\sigma_1} \varphi\left(\frac{y_i - \mu_1}{\sigma_1}\right) + (1-\pi) \frac{1}{\sigma_2} \varphi\left(\frac{y_i - \mu_2}{\sigma_2}\right)$$

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

$$\ell(\mu_1, \mu_2, \sigma_1, \sigma_2, \pi ; \underset{\sim}{y})$$

$$= \sum_{i=1}^{n} \log \left\{ \frac{\pi}{\sigma_1} \varphi\left(\frac{y_i - \mu_1}{\sigma_1}\right) + \frac{(1-\pi)}{\sigma_2} \varphi\left(\frac{y_i - \mu_2}{\sigma_2}\right) \right\}$$

log-lik. $\qquad$ mix.obj $\qquad p \Leftrightarrow \underset{\sim}{\theta} = \theta_1 \cdots \theta_5$

$x \Leftrightarrow y_1 \cdots y_n$