STAT 535

DIRECT FUNCTION MAXIMIZATION

Good Ref: Ch. 11 of Lange

Newton's Method

Want to mazimize $g(\theta)$.

Willing and able to evaluate $g()$, $g'()$, $g''()$ at arbitrary points.

$$\theta^{(i+1)} = \theta^{(i)} - \left[g''(\theta^{(i)})\right]^{-1} g'(\theta^{(i)})$$

Always feasible? Always 'uphill'?

More generally...

$$\theta^{(i+1)} = \theta^{(i)} + \left[M(\theta^{(i)})\right]^{-1} g'(\theta^{(i)})$$

If M is $(1/\epsilon) \times$ 'something P.D.', then this will go uphill for small enough ϵ. Proof?

Example: gradient ascent algorithm - set $M = (1/\epsilon)I$. Use backtracking to force uphill moves.

(Fisher) Scoring

Say $g(\theta)$ is a log-likelihood based on n *iid* observations $y_1, \ldots, y_n$.

Newton's method based on

$$M_{ij}(\theta) = -\sum_{k=1}^{n} \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(y_k; \theta).$$

wheras scoring is based on

$$\begin{aligned} M_{ij}(\theta) &= -nE_\theta \left\{ \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(Y; \theta) \right\} \\ &= nE_\theta \left\{ \frac{\partial}{\partial\theta_i} \log f(Y; \theta) \frac{\partial}{\partial\theta_j} \log f(Y; \theta) \right\}. \end{aligned}$$

Former may only be P.D., for θ near mode, latter will be P.D. everywhere. (Asymptotic SE's either way).

Does Scoring Help?

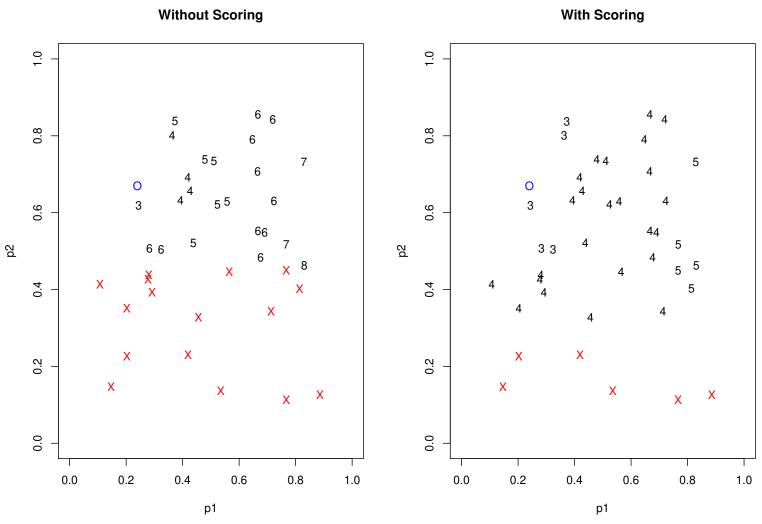
$(X_1, X_2, X_3, X_4) \sim \text{Multinomial}(100; \nu_1, \nu_2, \nu_3, \nu_4)$

$$\begin{aligned}
\log \nu_1 &= \theta_1 + \theta_2 \\
\log \nu_2 &= \theta_1 + \log(1 - \exp(\theta_2)) \\
\log \nu_3 &= \log(1 - \exp(\theta_1)) + \theta_2 \\
\log \nu_4 &= \log(1 - \exp(\theta_1)) + \log(1 - \exp(\theta_1))
\end{aligned}$$

Observed data (14 10 53 23), simulated under $(\theta_1, \theta_2) = (\log .2, \log .7)$

Maximize log-likelihood $g(\theta_1, \theta_2)$, get $(\hat{\theta}_1, \hat{\theta}_2) = (\log .24, \log .67)$

Try different starting values (sampled from distribution), compare Newton's method versus scoring...



Interesting Application

Logistic regression:

$$\text{logit} Pr(Y_i = 1 | X_{i1}, \ldots, X_{ip}) = \beta_1 X_{i1} + \ldots + \beta_p X_p,$$

where $\text{logit}(p) = log(p/(1-p))$ (link or inverse link?).

Claim: can write Newton's method in the form

$$\beta^{(i+1)} = (X'WX)^{-1} X'W\tilde{y},$$

where the diagonal 'weight' matrix W and the 'psuedo-response' vector $\tilde{y}$ depend on $\beta^{(i)}$.

Familiar looking form? Software implications?

More to say...

Generality of IRWLS for GLM's

Quasi-Newton methods