Maximum likelihood

Maximize $l(\theta) = \ln P(x|\theta)$

Set $0 = l'(\theta) \stackrel{\text{def}}{=} s$

s is called (Fisher) score

Newton's method

Let $V = -l''(\theta)$

Have $s_{\mathrm{new}} \approx s - V(\theta_{\mathrm{new}} - \theta)$

So set $\theta_{\mathrm{new}} = \theta + V^{-1}s$

Fisher scoring

Same, but $V = \text{Var}(s|\theta)$

V is called (Fisher) information

For exponential family w/ natural parameter, $V_{\text{Fisher}} = V_{\text{Newton}}$

Newton: $l(\theta) \approx$ quadratic (Taylor expansion)

Fisher: $P(x|\theta) \approx$ normal (Edgeworth expansion)

Comparison

Sometimes easier to compute variance than derivative

Fisher is not always quadratically convergent

But sometimes more robust than Newton away from minimum

An example

Exponential family $P(x|\theta) = \exp(\theta \cdot x - a(x) - b(\theta))$

Nonlinear transform $\theta = f(\mu)$

Can show

- $s = f'(\mu)(x - b'(\theta))$
- $V_{\text{Fisher}} = f'(\mu)^2 b''(\theta)$
- $V_{\text{Newton}} = V_{\text{Fisher}} - f''(\mu)(x - b'(\theta))$

Example cont'd

Let

- P be normal
- $f(\mu) = \frac{1}{1+e^{-\mu}}$
- $x = .8$ (so $f^{-1}(x) \approx 1.386294$)

Fisher	Newton
0	0
1.2	1.2
1.3769319	1.3615785
1.3862683	1.3857626
1.3862947	1.3862946
1.3862944	1.3862946
1.3862944	

Approx convergence interval

Fisher	$[-1.7312, 3.3512]$
Newton	$[-0.3958, 2.0478]$

Observed vs. expected information

V_{Newton} depends on x, so is a random variable

Usually $E(V_{\text{Newton}}|\theta) = V_{\text{Fisher}}$ (*e.g.*, curved exp fam)

So V_{Newton} is called "observed information" and V_{Fisher} is "expected information"

When does $V_{\text{Fisher}} = V_{\text{Newton}}$

In example, for any θ, $V_{\text{Fisher}} = V_{\text{Newton}}$ at MLE

Holds only if f is linear or onto

Linear means $f''(\mu) = 0$, onto means $x = b'(\theta)$ at MLE

If $V_{\text{Fisher}} = V_{\text{Newton}}$ at MLE then Fisher scoring is superlinear