Lecture 5: Newton's Method, the Exponential Family

(Text Sections 4.2, 3.2)

When there is no closed form for the solution of

$$\frac{d}{d\boldsymbol{\theta}} \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) \equiv g(\boldsymbol{\theta}) = 0$$

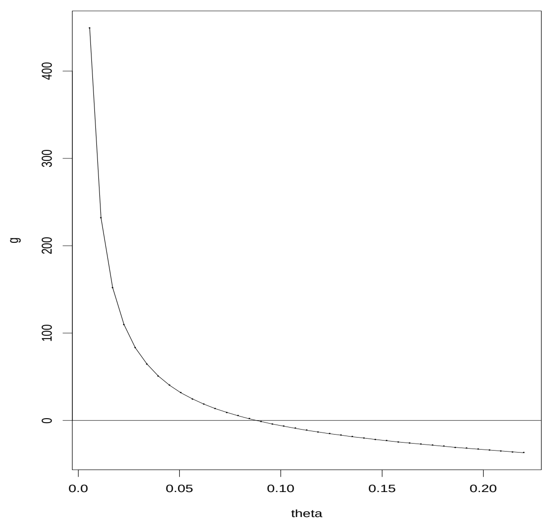
we may use a numerical method to determine the MLE. The Newton-Raphson method is one commonly used choice. (See Section 4 of the text.)

Newton-Raphson method

Assume that θ is 1-dimensional.

1. Choose an initial estimate, $\theta^{(0)}$.
2. Compute $\theta^{(i)} = \theta^{(i-1)} - \frac{g(\theta^{(i-1)})}{g'(\theta^{(i-1)})}$.
3. Repeat (2) until convergence at some iteration k, i.e. where $g(\theta^{(k)}) \approx 0$.

Figure 1: Plot of g vs. θ for $n = 15$, $t = 64$ in the logarithmic distribution.



Explanation:

- $\dfrac{g(\theta^{(i)}) - g(\theta^{(i-1)})}{\theta^{i} - \theta^{(i-1)}} \approx g'(\theta^{(i-1)})$

- Since we expect that $\theta^{(i)}$ is closer to the true parameter value than $\theta^{(i-1)}$

$$g(\theta^{(i)}) - g(\theta^{(i-1)}) \approx 0 - g(\theta^{(i-1)}).$$

- Thus, $\theta^{(i)} \approx \theta^{(i-1)} - \frac{g(\theta^{(i-1)})}{g'(\theta^{(i-1)})}$.

The NR method does *not* guarantee that we will locate the MLE!

- The algorithm might get "trapped" in a local minimum or maximum of g: choose good starting values!
- If the algorithm converges to a point where $g(\theta^{(k)}) \approx 0$, this point may correspond to a minimum, saddlepoint, or even local maximum of the likelihood function.

A graph can be helpful (from Figure 1, $\hat{\theta} \approx 0.09$). Alternatively, multiple (possibly randomly chosen) starting values can be used.

In the case where $\boldsymbol{\theta}$ is p-dimensional, the derivative of the log-likelihood with respect to $\boldsymbol{\theta}$ is also p-dimensional. Denote this vector of derivatives by $\mathbf{g}(\boldsymbol{\theta})$, and denote the matrix of second derivatives of the log-likelihood with respect to $\boldsymbol{\theta}$ by $\mathbf{H}(\boldsymbol{\theta})$. The NR equation generalizes to

$$\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)} - \mathbf{H}^{-1}(\boldsymbol{\theta}^{(i-1)})\mathbf{g}(\boldsymbol{\theta}^{(i-1)}).$$

The Exponential Family

Let θ be 1-dimensional. A distribution that can be written in the form

$$f_Y(y;\theta) = \exp\{a(y)b(\theta) + c(\theta) + d(y)\},$$

where the range of Y does not depend on θ, belongs to the 1-parameter *exponential family* of distributions.

If there are parameters other than θ, they are regarded as *nuisance parameters* forming parts of the functions a, b, c, and d and are treated as known.

If $a(y) = y$, then the distribution is said to be in *canonical form*, and, in this case, $b(\theta)$ is called the *natural parameter* of the distribution. Exponential families which can be written in canonical form have some nice properties.

NOTE: The functions $a(y)$ and $b(\theta)$ are unique only up to a scalar multiple (e.g. we can always write $ka(y)$ and $b(\theta)/k$ for some constant k). Likewise, the functions $c(\theta)$ and $d(y)$ are unique only up to an additive constant (e.g. we can always write $c(\theta) + k$ and $d(y) - k$ for some constant k).

Example: Poisson distribution with mean μ

$$\begin{aligned} f_Y(y;\mu) &= \frac{e^{-\mu}\mu^y}{y!} \\ &= \exp\{y\log\mu - \mu - \log y!\} \end{aligned}$$

This is the canonical form of the Poisson distribution with $a(y) = y$, $b(\mu) = \log\mu$, $c(\mu) = -\mu$, and $d(y) = -\log y!$. The natural parameter is $\log\mu$.

Example: $\mathrm{N}(\mu, \sigma^2)$ distribution, σ^2 known

The density of the $\mathrm{N}(\mu, \sigma^2)$ distribution can be written as

$$\begin{aligned} f_Y(y;\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y-\mu)^2\right\} \\ &= \exp\left\{y\left(\frac{\mu}{\sigma^2}\right) - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)\right\} \end{aligned}$$

This is the canonical form of the normal distribution, with $a(y) = y$, $b(\mu) = \frac{\mu}{\sigma^2}$, $c(\mu) = -\frac{\mu^2}{2\sigma^2}$, and $d(y) = -\frac{y^2}{2\sigma^2} - \frac{1}{2}\log(2\pi\sigma^2)$. The natural parameter is $\frac{\mu}{\sigma^2}$.

Example: Uniform distribution

Let Y have uniform distribution on $[0, \theta]$, i.e.

$$\begin{aligned} f_Y(y) &= \begin{cases} \frac{1}{\theta}, & 0 \le y \le \theta \\ 0, & \text{otherwise} \end{cases} \\ &= \frac{1}{\theta}\,\mathbf{1}_{\{0\le y\le\theta\}} \end{aligned}$$

where

$$\mathbf{1}_{\{0\le y\le\theta\}} = \begin{cases} 1, & 0 \le y \le \theta \\ 0, & \text{otherwise} \end{cases}$$

in an *indicator function*. This distribution is *not* in the exponential family, since the range of Y depends on θ.

If we have n independent observations from this distribution, the likelihood is

$$\begin{aligned} \mathcal{L}(\theta;\mathbf{y}) &= \prod_{i=1}^{n}\frac{1}{\theta}\,\mathbf{1}_{\{0\le y_i\le\theta\}} \\ &= \frac{1}{\theta^n}\mathbf{1}_{\{0\le y_1\le\theta,\ldots,0\le y_n\le\theta\}} \\ &= \frac{1}{\theta^n}\mathbf{1}_{\{0\le\max(y_1,\ldots,y_n)\le\theta\}}. \end{aligned}$$

Q: What is the MLE of θ?