1. Consider a sample $(X_1 \ldots, X_n) = (x_1, \ldots, x_n)$ of independent observations from a canonical exponential family with density

$$f(x;\theta) = b(x)e^{\theta^\top t(x) - c(\theta)}.$$

The *reciprocal score* $R(\theta)$ is defined by

$$\mathcal{R}(\theta) = -\frac{\partial}{\partial\theta} L(\theta)^{-1}$$

i.e. the derivative of the reciprocal of the likelihood function rather than its logarithm.

(a) Show that

$$R(\theta) = L(\theta)^{-1}\{t(X) - \tau(\theta)\} = L(\theta)^{-1}S(\theta)$$

where $\tau(\theta) = \mathbf{E}_\theta\{t(X)\}$.

We get

$$L(\theta)^{-1} = b(X)^{-1}e^{-\theta^\top t(X) + c(\theta)}.$$

Differentiating yields

$$R(\theta) = -b(x)^{-1}\{-t(X) + \frac{\partial}{\partial\theta}c(\theta)\}e^{-\theta^\top t(X)+c(\theta)} = L(\theta)^{-1}S(\theta) \qquad (1)$$

as required.

(b) Show that using Newton's method on the reciprocal score equation $R(\theta) = 0$ leads to the iteration

$$\theta \leftarrow \theta + \{v(\theta) + S(\theta)S(\theta)^\top\}^{-1}S(\theta).$$

Newton's method applied to the reciprocal score equation has the form

$$\theta \leftarrow \theta - \left\{\frac{\partial}{\partial\theta^\top}R(\theta)\right\}^{-1} R(\theta).$$

Differentiating further in (1) we get

$$\begin{aligned} -\left\{\frac{\partial}{\partial\theta^\top}R(\theta)\right\} &= L(\theta)^{-1}\frac{\partial}{\partial\theta^\top}\tau(\theta) - S(\theta)\left\{\frac{\partial}{\partial\theta^\top}L(\theta)^{-1}\right\} \\ &= L(\theta)^{-1}v(\theta) + S(\theta)R(\theta)^\top = L(\theta)^{-1}\left\{v(\theta) + S(\theta)S(\theta)^\top\right\} \end{aligned}$$

and the result follows.

(c) Compare this method to the method of scoring.

Using the identity for a non-singular square matrix $A$ and column vectors $x, y$

$$(A + xy^\top)^{-1} = A^{-1} - \frac{A^{-1}xy^\top A^{-1}}{1 + y^\top A^{-1}x}$$

we get

$$\{v(\theta) + S(\theta)S(\theta)^\top\}^{-1} = v(\theta)^{-1} - \frac{v(\theta)^{-1}S(\theta)S(\theta)^\top v(\theta)^{-1}}{1 + S(\theta)^\top v(\theta)^{-1}S(\theta)}$$

leading to

$$\{v(\theta) + S(\theta)S(\theta)^\top\}^{-1}S(\theta) = \frac{v(\theta)^{-1}S(\theta)}{1 + S(\theta)^\top v(\theta)^{-1}S(\theta)}$$

making it easier (?) to see that this iteration takes smaller steps when $S(\theta)$ is large, hence becomes more stable.

2. Let $X = (X_1, \ldots, X_n)$ be a sample from the *Weibull* distribution with individual densities

$$f(x; \theta) = \theta x^{\theta-1} e^{-x^\theta} \text{ for } x > 0$$

where $\theta > 0$ is unknown.

(a) Find the score statistic and the likelihood equation;

The score statistic is

$$S(\theta) = \frac{n}{\theta} + \sum \log x_i - \sum x_i^\theta \log x_i$$

and the likelihood equation can therefore be written.

$$\theta = \frac{n}{\sum x_i^\theta \log x_i - \sum \log x_i}. \tag{2}$$

(b) Show that if the likelihood equation has a solution, it must be the MLE;

The second derivative of the log-likelihood function is

$$S'(\theta) = -j(\theta) = -\frac{n}{\theta^2} - \sum x_i^\theta (\log x_i)^2$$

and this is clearly negative, so a solution of the likelihood equation must necessarily be the MLE.

(c) Describe the Newton–Raphson method for solving the likelihood equation;

The Newton–Raphson iterative step becomes

$$\theta \leftarrow \theta + \frac{S(\theta)}{j(\theta)} = \theta + \frac{n\theta + \theta^2 \sum \log x_i - \theta^2 \sum x_i^\theta \log x_i}{n + \theta^2 \sum x_i^\theta (\log x_i)^2}.$$

(d) Describe the method of scoring for solving the likelihood equation;

For the method of scoring we need to calculate the Fisher information which involves the integral

$$\mathbf{E}\{X^\theta(\log X)^2\} = \int_0^\infty x^\theta(\log x)^2\theta x^{\theta-1}e^{-x^\theta}\,dx.$$

Substituting $u = x^\theta$, $du = \theta x^{\theta-1}$ we get

$$\mathbf{E}\{X^\theta(\log X)^2\} = \theta^{-2}\int_0^\infty u(\log u^2)e^{-u}\,du = \theta^{-2}(\pi^2/6 + \gamma^2 - \gamma)$$

where $\gamma = -0.5772\ldots$ is Euler's constant. This yields the Fisher information

$$i(\theta) = \frac{n}{\theta^2}(1 + \pi^2/6 + \gamma^2 - \gamma)$$

and hence the iterative step in the method of scoring becomes

$$\theta \leftarrow \theta + \frac{S(\theta)}{i(\theta)} = \theta + \frac{n\theta + \theta^2\sum\log x_i - \theta^2\sum x_i^\theta \log x_i}{n(1 + \pi^2/6 + \gamma^2 - \gamma)}.$$

(e) Can you think of other methods for solving the likelihood equation?

It is tempting to use the equation (2) as a basis for an iteration

$$\theta \leftarrow \frac{n}{\sum_i x_i^\theta \log x_i - \sum \log x_i}$$

but its convergence properties are not all that clear.

3. Consider a sample $X = (X_1, \ldots, X_n)$ from a normal distribution $\mathcal{N}(\mu, \mu^2)$, where $\mu > 0$ is unknown. This corresponds to the coefficient of variation $\sqrt{\mathbf{V}(X)}/\mathbf{E}(X)$ being known and equal to 1.

(a) Find the score function for $\mu$;

We get by differentiation of

$$\log f(x_i;\mu) = -\frac{1}{2}\log(2\pi) - \log\mu - \frac{x_i^2}{2\mu^2} + \frac{x_i}{\mu} - \frac{1}{2}$$

and summing over $i$ that

$$S(\mu) = -\frac{n}{\mu} + \frac{\sum_i X_i^2}{\mu^3} - \frac{\sum_i X_i}{\mu^2}.$$

If we let $S = \sum X_i$ and $SS = \sum X_i^2$ this can be rewritten as

$$S(\mu) = -\frac{n}{\mu} + \frac{SS}{\mu^3} - \frac{S}{\mu^2}.$$

(b) Show that the likelihood equation has a unique root $\hat{\mu}$ within the parameter space unless $X_i$ are all equal to zero;
The likelihood equation is obtained by equating the score statistic to 0. Doing this and multiplying with $\mu^3$ yields the equation

$$n\mu^2 + \mu S - SS = 0$$

which has exactly one positive root

$$\hat{\mu} = \frac{-S + \sqrt{S^2 + 4nSS}}{2n}$$

unless $SS = 0$, which implies all $X_i$ are equal to zero.

(c) Show that the observed information at $\hat{\mu}$ is

$$j(\hat{\mu}) = \frac{n}{\hat{\mu}^2} + \frac{\sum_i X_i^2}{\hat{\mu}^4},$$

and use this to argue that the root $\hat{\mu}$ is indeed the MLE of $\mu$;
We get by further differentiation that

$$j(\mu) = -\frac{n}{\mu^2} + \frac{3SS}{\mu^4} - \frac{2S}{\mu^3}.$$

Since $\hat{\mu}$ satisfies the likelihood equation we have

$$\frac{S}{\hat{\mu}^2} = \frac{SS}{\hat{\mu}^3} - \frac{n}{\hat{\mu}}.$$

Inserting this into the expression for $j(\hat{\mu})$ we get

$$j(\hat{\mu}) = \frac{n}{\hat{\mu}^2} + \frac{SS}{\hat{\mu}^4} > 0,$$

so the root of the likelihood equation is the unique local (and therefore global) maximum.

(d) Show that the Fisher information is equal to $(3n)/\mu^2$.
We use that $\mathbf{E}(X^2) = \mathbf{V}(X) + \{\mathbf{E}(X)\}^2 = 2\mu^2$ and take expectations in the expression for $j(\mu)$ to get

$$i(\mu) = \mathbf{E}\{j(\mu)\} = -\frac{n}{\mu^2} + \frac{6n\mu^2}{\mu^4} - \frac{2n\mu}{\mu^3} = \frac{3n}{\mu^2}.$$

4. Let $X = (X_1, \dots, X_n)$ be a sample from the Gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ both unknown, i.e. the distribution with individual densities

$$f(x; \alpha, \beta) = \frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)} e^{-\beta x}, \quad x > 0.$$

(a) Show that the score statistic for $\theta = (\alpha, \beta)$ is equal to

$$S(\alpha, \beta) = \begin{pmatrix} \sum_i \log X_i + n \log \beta - n\psi(\alpha) \\ n\alpha/\beta - \sum_i X_i \end{pmatrix}$$

where $\psi(\alpha) = D \log \Gamma(\alpha)$ is the *digamma* function;
As this family is a canonical exponential family, the score statistic is the difference of the canonical statistic $t(X) = (\sum_i \log X_i, -\sum_i X_i)^\top$ and the derivative of the log-normalizing constant

$$c(\alpha, \beta) = n \log \Gamma(\alpha) - n\alpha \log \beta,$$

which yields the result.

(b) Show that the method of scoring for $\theta$ leads to the iteration

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \leftarrow \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \frac{1}{n\{\alpha\psi'(\alpha) - 1\}} \begin{pmatrix} \alpha & \beta \\ \beta & \beta^2\psi'(\alpha) \end{pmatrix} S(\alpha, \beta),$$

where $\psi'(\alpha)$ is the *trigamma* function.
As this family is a canonical exponential family, the information matrix is determined by double differentiation of the log-normalising constant so we get

$$i(\theta) = n \begin{pmatrix} \psi'(\alpha) & -1/\beta \\ -1/\beta & \alpha/\beta^2 \end{pmatrix}.$$

Taking inverses we get

$$i(\theta)^{-1} = \frac{1}{n\{\alpha\psi'(\alpha) - 1\}} \begin{pmatrix} \alpha & \beta \\ \beta & \beta^2\psi'(\alpha) \end{pmatrix}.$$

(c) Consider a simpler iteration for solving the likelihood equation by first elimination $\beta$ from the equation.
We first eliminate $\beta$ from the equations by solving

$$n\alpha/\beta = \sum_i X_i$$

to get

$$\beta = \alpha/\overline{X}.$$

Inserting this into the equation

$$\sum_i \log X_i + n \log \beta - n\psi(\alpha) = 0$$

and dividing by $n$ now yields

$$\overline{\log X} - \log \bar{X} + \log \alpha - \psi(\alpha) = 0.$$

If we apply Newton's method to this equation, we get the iteration

$$\alpha \leftarrow \alpha + \frac{\overline{\log X} - \log \bar{X} + \log \alpha - \psi(\alpha)}{\psi'(\alpha) - 1/\alpha}.$$