Exercise 10 The log-likelihood based on n independent observations within the generalised linear model are:

$$l(\beta, \phi) = \log \prod_i f(y_i; \beta, \phi), \quad \beta = \beta_0, \ldots, \beta_p$$

Further we define (with ϕ considered fixed):

$$s_j(\beta, \phi) = \frac{\partial}{\partial \beta_j} l(\beta, \phi), \quad j = 0, \ldots, p$$

$$I_{j,k}(\beta, \phi) = -E\left[\frac{\partial^2}{\partial \beta_j \partial \beta_k} l(\beta, \phi)\right], \quad j, k = 0, \ldots, p$$

Assumtion: derivative and integral are exchangeable. This is true if the derivatives of both the function being integrated with respect to the parameter of interest and the derivative of the integral with respect to this parameter of interest must exist, then the derivative and integral operations are exchangeable

Show that $\text{cov}[s_j(\beta, \phi), s_k(\beta, \phi)] = I_{j,k}(\beta, \phi)$:

i) First look at the situation where $n = 1$:

First result we will use is

$$E_{f(y)}(s_j) = \int s_j f(y) dy = \int \frac{\partial log(f(y))}{\partial \beta_j} f(y) dy = \int \frac{1}{f(y)} \frac{\partial f(y)}{\partial \beta_j} f(y) dy = \int \frac{\partial f(y)}{\partial \beta_j} dy$$
$$= \frac{\partial}{\partial \beta_j} \int f(y) dy = \frac{\partial}{\partial \beta_j} 1 dy = 0$$

The second result comes from taking the derivative again: (is also obviously equal to zero)

$$\begin{aligned}
\frac{\partial}{\partial\beta_k}0 &= \frac{\partial}{\partial\beta_k}E_{f(y)}(s_j)\\
0 &= \frac{\partial}{\partial\beta_k}\int\frac{\partial log(f(y))}{\partial\beta_j}f(y)dy\\
&= \int\left(\frac{\partial^2 log(f(y))}{\partial\beta_j\beta_k}f(y)+\frac{\partial log(f(y))}{\partial\beta_j}\frac{\partial f(y)}{\partial\beta_k}\right)dy\\
&= \left[\frac{\partial g(x)}{\partial x}=\frac{\partial\log g(x)}{\partial x}g(x)\right]=\int\left(\frac{\partial^2 log(f(y))}{\partial\beta_j\beta_k}f(y)+\frac{\partial log(f(y))}{\partial\beta_j}\frac{\partial log(f(y))}{\partial\beta_k}f(y)\right)dy\\
&= \int\frac{\partial^2 log(f(y))}{\partial\beta_j\beta_k}f(y)dy+\int\frac{\partial log(f(y))}{\partial\beta_j}\frac{\partial log(f(y))}{\partial\beta_k}f(y)dy\\
&= E\left(\frac{\partial^2 log(f(y))}{\partial\beta_j\beta_k}\right)+\int s_j s_k f(y)dy\\
&= -I+E(s_j s_k)\\
I &= E(s_j s_k)
\end{aligned}$$

Finally for (n=1):

$$\text{cov}[s_j,s_k]=E[(s_j-E[s_j])(s_k-E[s_k])]=E[s_j,s_k]=I_{jk}$$

Looking at the situation in general, having in mind that we have independent samples:

$$\begin{aligned}
s_j &= \frac{\partial l}{\partial\beta_j}=\frac{\partial\log\prod_i f(y_i)}{\partial\beta_j}=\sum_i\frac{\partial\log f(y_i)}{\partial\beta_j}=\sum_i s_{ij}\\
E(s_j) &= 0+\cdots+0=0
\end{aligned}$$

$$\begin{aligned}
I_{jk} &= E\left[-\frac{\partial^2}{\partial\beta_j\partial\beta_k}l(\beta,\phi)\right]=E\left[-\frac{\partial^2}{\partial\beta_j\partial\beta_k}\log\prod_i f(y_i)\right]=\sum_i E\left[-\frac{\partial^2}{\partial\beta_j\partial\beta_k}\log f(y_i)\right]\\
&= \sum_i I_{ijk}=nI_{ijk}
\end{aligned}$$

Finally (n≥1):

$$\text{cov}[s_j,s_k]=\text{cov}[\sum_i s_{ij},\sum_l s_{lk}]=\sum_i\sum_l cov(s_{ij},s_{lk})=\sum_i cov(s_{ij},s_{ik})=\sum_i I_{ijk}=nI_{ijk}$$

ii) Show that $I=\{I_{jk}(\beta,\phi)\}$ cannot be negative definite.
The score vector for each β_j is

$$\underline{\text{s}}=(s_0,\ldots,s_p)^T.$$

We shown that we can write

$$I = \text{cov}(s) = E(ss^T) \qquad \text{diag(I)} = \text{Var}(s_j) \geq 0.$$

DEF: I is non-negative definite if $a^T I a \geq 0$ for any non-zero a.

$$a^T I a = a^T \text{cov}(s) a = \text{cov}(a^T s) = E\Big[(a^T s)(a^T s)^T\Big] = E\Big[a^T s s^T a\Big] = E\Big[a^T s s^T a\Big] = a^T E\Big[s s^T\Big] a \geq 0$$

These results are general and not specific for generalised linear models!

so in general for any co variance matrix Σ: $\Sigma = \Sigma^T$ meaning that Σ is symmetric.

$$a^T \Sigma a = a^T \text{cov}(x,x) a = a^T E\Big[(x-m)(x-m)^T\Big] a =$$

$$= E\Big[a^T (x-m)(x-m)^T a\Big] = E\Big[((x-m)^T a)^2\Big] \geq 0$$

another way is to show that

$$\Sigma = [eigendecomposition] = V \Lambda V^T$$

where the matrix of $V := [v_1, ..., v_n]$ is orthogonal (that is, $V^T V = V V^T = I_n$), and contains the eigenvectors of Σ, while the diagonal matrix Λ contains the eigenvalues of Σ (see Symmetric eigenvalue decomposition theorem). Thus

$$a^T \Sigma a = a^T V \Lambda V^T a = \geq [\lambda_i \geq 0, V^T V = V V^T = I_n] \geq 0$$