1. Let $X = (X_1, \dots, X_n)$ be independent random variables with

$$P(X_i = 1) = 1 - P(X_i = 0) = \frac{e^{\alpha+\beta d_i}}{1 + e^{\alpha+\beta d_i}},$$

where $\alpha$ and $\beta$ are unknown real-valued parameters and $d = (d_1, \dots, d_n)$ is a vector of known numbers.

This model is a *linear logistic regression* model and the vector $d$ might typically describe varying doses of drug administered to different individuals and $X_i$ whether individual $i$ responds or not.

(a) Show that this is a canonical exponential family with canonical sufficient statistic

$$t(x) = \left(\sum_i x_i, \sum_i x_i d_i\right).$$

(b) Write down the likelihood equation for the parameters $\alpha$ and $\beta$.

(c) Consider now the case where $\alpha$ is known and $\alpha = 0$. Find the Fisher information for $\beta$ and the asymptotic variance of $\hat{\beta}$.

(d) The likelihood equation must be solved by iterative methods. Write down the iterative step for Fisher's scoring method in the case where $\alpha = 0$ is known.

(e) What is the iterative step for the Newton-Raphson method, when $\alpha = 0$ is known?

(f) Write the iterative step for Fisher's method of scoring in the case where both $\alpha$ and $\beta$ are unknown.

(g) Find the asymptotic covariance matrix of $(\hat{\alpha}, \hat{\beta})$.

2. Let $Y = (Y_1, \dots, Y_n)$ be independent Poisson random variables with

$$\mathbf{E}(Y_i) = \beta + \alpha d_i$$

where $\alpha$ and $\beta$ are unknown real-valued parameters and $d = (d_1, \dots, d_n)$ is a vector of known numbers.

This could be a model for the measured number of particles emitted by $d_i$ radioactive sources, each emitting particles at the intensity of $\alpha$ (particles per time unit), and $\beta$ is the intensity of the *background* radiation.

Consider now the background intensity as being known and equal to unity $\beta = 1$. We wish to estimate $\alpha$.

(a) Why is this *not* a canonical exponential family?

(b) Find the score statistic for $\alpha$;

(c) Find the Fisher information for $\alpha$;

(d) Write down the iterative step for solving the likelihood equation by the Newton–Raphson method;

(e) Write down the iterative step for solving the likelihood equation by Fisher's scoring method;

(f) Suppose that instead of observing $Y$ as above, we had 'complete data' $(B_i, X_i), i = 1, \ldots, n$, where $B_i$ is counting the number of background emissions and $X_i$ the number of emissions from the radioactive sources, so that each were independent Poisson and

$$\mathbf{E}(B_i) = \beta, \mathbf{E}(X_i) = \alpha d_i, \quad Y_i = B_i + X_i.$$

Find the MLE for estimating $\alpha$ based on complete data, assuming $\beta = 1$ is known.

(g) This setup is ideal for using the EM algorithm in the case where we have only observed $Y$ as above. Show that the E-step can be implemented through calculating $x_i^* = \mathbf{E}(X_i \mid Y_i)$, and find this conditional expectation.

(h) Describe the E-step and M-step of the EM algorithm for estimating $\alpha$, assuming $\beta$ is known.

(i) How are the steps of the EM algorithm if the background intensity $\beta$ is unknown as well?