1. Linear logistic regression

   (a) The joint density is

$$\begin{aligned} f(x;\alpha,\beta) &= \prod_i \frac{e^{(\alpha+\beta d_i)x_i}}{1+e^{\alpha+\beta d_i}} = \frac{e^{\alpha\sum_i x_i+\beta\sum_i d_i x_i}}{\prod_i(1+e^{\alpha+\beta d_i})} \\ &= e^{\alpha\sum_i x_i+\beta\sum_i d_i x_i-\sum_i \log(1+e^{\alpha+\beta d_i})}. \end{aligned}$$

   This has the form of a canonical exponential family with canonical sufficient statistic $t(X) = (\sum_i X_i, \sum_i X_i d_i)$ and log-normalizing constant

$$c(\alpha,\beta) = \sum_i \log(1+e^{\alpha+\beta d_i}).$$

   (b) The likelihood equation for the parameters α and β is determined by equating the canonical statistics to their expectations. We have

$$\mathbf{E}(X_i) = p_i = \frac{e^{\alpha+\beta d_i}}{1+e^{\alpha+\beta d_i}}$$

   so the likelihood equations are

$$\sum_i x_i = \mathbf{E}(\sum_i X_i) = \sum_i \frac{e^{\alpha+\beta d_i}}{1+e^{\alpha+\beta d_i}}, \quad \sum_i x_i d_i = \mathbf{E}(\sum_i X_i d_i) = \sum_i \frac{d_i e^{\alpha+\beta d_i}}{1+e^{\alpha+\beta d_i}}.$$

   Expressing this in terms of $p_i = \mathbf{E}(X_i)$ yields the alternative expressions

$$\sum_i x_i = \sum_i p_i, \quad \sum_i x_i d_i = \sum_i p_i d_i.$$

   (c) In the case where it is known that $\alpha = 0$ we also have a canonical exponential family with $t_2(X) = \sum_i X_i d_i$ as canonical statistic. The Fisher information for β is found by differentiating the log-normalizing constant $c(0,\beta)$ twice to yield

$$i(\beta) = \sum_i \frac{d_i^2 e^{\beta d_i} - (d_i e^{\beta d_i})^2}{(1+e^{\beta d_i})^2} = \sum_i d_i^2 p_i(1-p_i) = \mathbf{V}(\sum_i X_i d_i)$$

   and the asymptotic variance of $\hat{\beta}$ is $1/i(\beta)$.

   (d) The iterative step for Fisher's scoring method in the case where $\alpha = 0$ is known is

$$\beta \leftarrow \beta + \frac{\sum_i (x_i - p_i) d_i}{\sum_i d_i^2 p_i(1-p_i)}.$$

(e) The iterative step for the Newton-Raphson method, when $\alpha = 0$ is known is identical to that in the method of scoring since this is a canonical exponential family.

(f) To write the iterative step for Fisher's method of scoring in the case where both α and β are unknown we need the full information matrix, obtained by differentiation of the log-normalizing constant $c(\alpha, \beta)$ or—whichever is easier— by calculating the covariance matrix of the canonical statistic. With the short notation introduced earlier, this gives

$$\begin{aligned} i(\alpha, \beta) &= \begin{pmatrix} \mathbf{V}(\sum_i X_i) & \mathrm{Cov}(\sum_i X_i, \sum_i d_i X_i) \\ \mathrm{Cov}(\sum_i X_i, \sum_i d_i X_i) & \mathbf{V}(\sum_i d_i X_i) \end{pmatrix} \\ &= \begin{pmatrix} \sum_i \mathbf{V}(X_i) & \sum_i \mathrm{Cov}(X_i, d_i X_i) \\ \sum_i \mathrm{Cov}(X_i, d_i X_i) & \sum_i \mathbf{V}(d_i X_i) \end{pmatrix} \\ &= \begin{pmatrix} \sum_i p_i(1-p_i) & \sum_i d_i p_i(1-p_i) \\ \sum_i d_i p_i(1-p_i) & \sum_i d_i^2 p_i(1-p_i) \end{pmatrix}. \end{aligned}$$

(g) The asymptotic covariance matrix of $(\hat{\alpha}, \hat{\beta})$ is given by the inverse of the information matrix calculated above. No specific simplication is obtained.

2. Deconvolution of Poisson signals.

(a) The joint density is

$$f(y; \alpha, \beta) = \prod_i \frac{(\beta + \alpha d_i)^{y_i}}{y_i!} e^{-\beta - \alpha d_i}.$$

For $\beta = 1$ we get

$$f(y; \alpha) = \prod_i \frac{(1 + \alpha d_i)^{y_i}}{y_i!} e^{-1 - \alpha d_i}.$$

This is not a canonical exponential family. If we try to give it exponential form we could write

$$f(y; \alpha) = \left(\prod_i \frac{1}{y_i!} \right) e^{\sum_i y_i \log(1 + \alpha d_i) - n - \alpha \sum_i d_i},$$

but $\log(1 + \alpha d_i)$ is not a linear function of any transformation of α.

(b) The score statistic for α is

$$S(\alpha) = \sum_i \frac{Y_i d_i}{1 + \alpha d_i} - \sum_i d_i = \sum_i \frac{d_i(Y_i - 1 - \alpha d_i)}{1 + \alpha d_i}.$$

(c) The Fisher information for α is found by differentiating yet another time and taking expectations

$$j(\alpha) = -S'(\alpha) = \sum_i \frac{Y_i d_i^2}{(1+\alpha d_i)^2}.$$

Taking expectations yields

$$i(\alpha) = \sum_i \frac{d_i^2}{1+\alpha d_i}.$$

(d) The iterative step for solving the likelihood equation by the Newton–Raphson method is

$$\alpha \leftarrow \alpha + \frac{S(\alpha)}{j(\alpha)} = \alpha + \frac{\sum_i \frac{d_i(Y_i - 1 - \alpha d_i)}{1+\alpha d_i}}{\sum_i \frac{Y_i d_i^2}{(1+\alpha d_i)^2}}.$$

(e) The iterative step for solving the likelihood equation by Fisher's scoring method replaces the observed information with the expected

$$\alpha \leftarrow \alpha + \frac{S(\alpha)}{i(\alpha)} = \alpha + \frac{\sum_i \frac{d_i(Y_i - 1 - \alpha d_i)}{1+\alpha d_i}}{\sum_i \frac{d_i^2}{(1+\alpha d_i)}}.$$

(f) For complete data the likelihood function becomes

$$f(b, x; \alpha) = \prod_i \frac{1}{b_i!} e^{-n} \prod_i \frac{(\alpha d_i)^{x_i}}{x_i!} e^{-\alpha d_i} = h(b, x, d) e^{\log \alpha \sum_i X_i - \alpha \sum_i d_i}.$$

This is a linear exponential family with canonical parameter $\theta = \log \alpha$ and the MLE of α is thus found by equating the canonical statistic to its expectation, i.e.

$$\sum_i x_i = \sum_i \alpha d_i; \quad \hat{\alpha} = \frac{\sum_i x_i}{\sum_i d_i}. \tag{1}$$

(g) The complete data log-likelihood is (ignoring an additive constant)

$$\log L(\alpha; X, B) = \log \alpha \sum_i X_i - \alpha \sum_i d_i.$$

The E-step of the EM algorithm finds the conditional expectation of this given the observed data, i.e.

$$\mathbf{E}\{\log L(\alpha; X, B) \,|\, Y\} = \log \alpha \sum_i \mathbf{E}\{X_i \,|\, Y\} - \alpha \sum_i d_i = \log \alpha \sum_i x_i^* - \alpha \sum d_i$$

where $x_i^* = \mathbf{E}(X_i \mid Y_i)$. Conditionally on Y_i, X_i follows a binomial distribution with parameters (Y_i, p_i) where

$$p_i = \frac{\mathbf{E}(X_i)}{\mathbf{E}(Y_i)} = \frac{\alpha d_i}{1 + \alpha d_i}$$

hence

$$x_i^* = \frac{\alpha Y_i d_i}{1 + \alpha d_i}.$$

(h) The E-step calculates x_i^* as above, thus imputing the unobserved values of the radioactive emissions from the source.

The M-step calculates the MLE by replacing x_i by x_i^* in (1), i.e. updates α as

$$\alpha \leftarrow \frac{\sum_i x_i^*}{\sum_i d_i}.$$

(i) If the background intensity β is unknown as well, the complete data log-likelihood becomes

$$\log L(\alpha, \beta; X, B) = \log \beta \sum_i B_i - n\beta + \log \alpha \sum_i X_i - \alpha \sum_i d_i.$$

Thus for the E-step we should calculate

$$x_i^* = \frac{\alpha Y_i d_i}{\beta + \alpha d_i}, \quad b_i^* = Y_i - x_i^*$$

and then update (α, β) as

$$\alpha \leftarrow \frac{\sum_i x_i^*}{\sum_i d_i}, \quad \beta \leftarrow \frac{\sum_i b_i^*}{n}.$$