

ST 762, HOMEWORK 1, FALL 2007

This problem is to be turned in on the due date.

1. In your favorite programming language, write two programs to implement the following methods.
 - (i) The 3-step GLS algorithm on pages 51 and 56 of the notes, where the “weight matrix” \mathbf{W} is held fixed at step (iii), as described in Section 3.2 of the notes. Your program should allow the user to choose the number of iterations C .
 - (ii) IRWLS ($C = \infty$) as described in Section 3.4 of the notes.

Do not mimic the programs in Section 3.7; I want you to write all parts of the algorithm (e.g. the Gauss-Newton scheme) yourself.

Both programs should have the following features:

- Allow the user to supply a starting value $\beta_{(0)}$ to get things going.
- Allow the user to choose a maximum number of iterative updates in step (iii) of the fixed- C algorithm and similarly a maximum number of iterative updates of IRWLS; a reasonable choice would be 500. If iteration continues up to this max number, each program should stop and declare that no convergence was reached.
- Use as the convergence criterion both in step (iii) of the fixed- C algorithm and IRWLS the following rule: If 2 successive iterates $\beta_{(a+1)}$ and $\beta_{(a)}$, say, have a *relative* difference of less than some small constant tol , stop and declare $\beta^{(a+1)}$ to be the solution. That is, stop if

$$\max_{\ell=1,\dots,p} |\beta_{\ell,(a+1)} - \beta_{\ell,(a)}| / |\beta_{\ell,(a)}| < tol$$

as on page 59. For your programs, take $tol = 10^{-8}$.

- Compute an estimate of σ^2 based on the final estimated value for β . Use the version of the estimator “adjusted for loss of degrees of freedom” (with the divisor $(n - p)$) on page 65.

To test your programs, consider the following scenario. Interferon- α -2b (IFN) is one of many experimental therapies that has been evaluated for treating patients infected with hepatitis C virus (HCV). A key way of evaluating such treatments is in terms of how they impact the within-patient *dynamics* of the HCV virus. That is, investigators are interested in understanding the behavior of *viral load*, which is roughly a measure of the concentration of virus in the patient’s system, following the start of IFN therapy. Data from one such patient are given in the file `hcv.dat` on the class web page. The first column is time (days) at which a measurement of viral load (copies of HCV RNA per ml), given in the second column, was taken from a blood sample drawn at that time. Here, time 0 is the time of initiation of IFN therapy, where the time 0 measurement was taken immediately prior to start of treatment; measurements were taken over the first two days of therapy.

Mathematicians have developed mathematical models in the form of systems of differential equations to characterize the dynamics, i.e., how the virus behaves in the body and how the body reacts and how IFN therapy may affect these processes over time. In an article in the journal *Science* (Volume 282, October 2, 1998, p. 103), Neumann et al. propose such a model to describe the dynamics over the first two days of IFN therapy. If the differential equations

are solved, they imply that, if $V(x)$ is the viral load at time x (days) following the start of IFN therapy at time $x = 0$,

$$V(x) = V_0[1 - \epsilon + \epsilon \exp\{-c(x - t_0)_+\}], \quad 0 \leq x \leq 2, \quad V_0, c, \epsilon > 0, \quad (1)$$

where $x_+ = x$ for $x > 0$, and 0 otherwise; V_0 is the viral load at time 0; c is the *virion clearance rate*; t_0 is the so-called *pharmacological delay* such that decay in viral load is not seen until IFN has made sufficient progress in distributing through the body; and ϵ is an *efficacy* parameter. The efficacy parameter has the interpretation that the effect of IFN therapy is to reduce the production of new virus particles, or *virions*, in the system by the fraction $(1 - \epsilon)$; if $\epsilon = 1$ then the therapy is said to be “perfect.” Here, we will take the pharmacological delay to be known: $t_0 = 0.10$ days.

Let Y_j be measured viral load at time x_j , $j = 1, \dots, n$. Based on (1), assume the model $E(Y_j|x_j) = f(x_j, \boldsymbol{\beta})$, where x is time and $f(x, \boldsymbol{\beta}) = V(x) \times 10^6$ with $V(x)$ as in (1). To ensure positivity of the parameters in (1), parameterize the model in terms of $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T = \{\log(V_0), \log(c), \log(\epsilon)\}^T$. Take $\text{var}(Y_j|x_j) = \sigma^2 f^{2\theta}(x_j, \boldsymbol{\beta})$.

(a) Make a plot of the data using your favorite software. This will give you a sense of the shape of the viral load-time relationship for this subject. You may want to plot the viral loads divided by 10^6 so that the scale of the vertical axis prints out more prettily.

(b) Assume that the variance model is

$$\text{var}(Y_j|\mathbf{x}_j) = \sigma^2 f^{2\theta}(\mathbf{x}_j, \boldsymbol{\beta}), \quad \theta \text{ known.}$$

For this mean-variance model, try three different fits using each program:

- (i) $\theta = 0.0$ (thus, assuming constant variance and fitting by OLS)
- (ii) $\theta = 0.5$ (“Poisson-like” variance)
- (iii) $\theta = 1.0$ (constant coefficient of variation)

For the three-step GLS algorithm, use $C = 20$ for (ii) and (iii); note that (i) only requires the first step of the algorithm. For IRWLS, note that in (i) $\mathbf{W} = \mathbf{I}_n$. Use the starting value $\boldsymbol{\beta}_{(0)} = \{\log(4.0), \log(3.5), \log(0.85)\}^T$. Please turn in both your programs and their output.

(c) Assuming that HCV loads exhibit approximate constant coefficient of variation, give an estimate of the HCV load at time 0 for this patient.