

ST 762, HOMEWORK 1 EXTRA PROBLEMS, FALL 2007

These problems are from previous years and are for you to work on or not as you choose; they are not to be turned in. You should be familiar with the concepts covered by these problems for the midterm test. Solutions will be posted when the homework problems to be turned in are due.

1. Suppose we have independent pairs (Y_j, \mathbf{x}_j) , $j = 1, \dots, n$.
 - (a) Suppose that $E(Y_j|\mathbf{x}_j) = f(\mathbf{x}_j, \boldsymbol{\beta})$, $j = 1, \dots, n$, and that the conditional distribution of Y_j given \mathbf{x}_j is Poisson. Write down the corresponding loglikelihood (conditional on \mathbf{x}_j) for $\boldsymbol{\beta}$ and derive the resulting estimating equation to be solved to obtain the maximum likelihood estimator for $\boldsymbol{\beta}$.
 - (b) Now suppose that $E(Y_j|\mathbf{x}_j) = k_j f(\mathbf{x}_j, \boldsymbol{\beta})$, $j = 1, \dots, n$, and that Y_j is the number of successes in k_j independent trials, so that Y_j given \mathbf{x}_j follows a binomial distribution.
 - (c) What features do the estimating equations in (a) and (b) under these different distributional assumptions have in common?
2. Suppose we have independent pairs (Y_j, \mathbf{x}_j) , $j = 1, \dots, n$, such that

$$E(Y_j|\mathbf{x}_j) = f(\mathbf{x}_j, \boldsymbol{\beta}), \quad \text{var}(Y_j|\mathbf{x}_j) = \sigma^2 f^2(\mathbf{x}_j, \boldsymbol{\beta}), \quad j = 1, \dots, n,$$

with σ known. For some of the calculations below, you may find it convenient to define

$$\lambda_\beta(\mathbf{x}_j, \boldsymbol{\beta}) = \partial/\partial\boldsymbol{\beta} \log f(\mathbf{x}_j, \boldsymbol{\beta}).$$

- (a) Assuming that the distribution of Y_j given \mathbf{x}_j is gamma, write down the corresponding loglikelihood for $\boldsymbol{\beta}$, and derive the form of the estimating equation for the maximum likelihood estimator for $\boldsymbol{\beta}$.
 - (b) Same as in (a), but now assuming the distribution of $Y_j|\mathbf{x}_j$ is normal.
 - (c) Same as in (a), but now assuming the distribution of $Y_j|\mathbf{x}_j$ is lognormal. (If $Z \sim N(m, \gamma^2)$, then $Y = \exp(Z)$ has a lognormal distribution.)
 - (d) Compare each estimating equation in (a) – (c) to the equations you found in Problem 1(a) and (b), commenting specifically on similarities or differences in the general forms of these equations. Do any of these equations share the common features you noted in 1(c)?
 - (e) Compare the estimating equations you found in (a) and (b) of this problem. State explicitly what features they share and what features are different. Give a possible reason for the differences you observe.
3. *More on transformations for nonlinear models.* An assumption underlying the “Transform Both Sides” (TBS) model discussed on pages 36–38 of the class notes is that a single transformation can be used to achieve both constant variance and normality on the transformed scale. For independent pairs (Y_j, \mathbf{x}_j) , $j = 1, \dots, n$, a generalization of this model has

$$E\{h(Y_j, \lambda)|\mathbf{x}_j\} = h\{f(\mathbf{x}_j, \boldsymbol{\beta}), \lambda\}, \quad \text{var}\{h(Y_j, \lambda)|\mathbf{x}_j\} = \sigma^2 q^2(\mathbf{x}_j, \theta) \quad (1)$$

for some transformation h depending on a scalar parameter λ . Note that this model implies that, after transformation, variance is not constant but depends on a function q of \mathbf{x}_j that is known if θ is known.

(a) By an argument similar to that on page 37 of the notes, show that (1) is roughly equivalent to a certain mean-variance model for $E(Y_j|\mathbf{x}_j)$ and $\text{var}(Y_j|\mathbf{x}_j)$. Give the form of $E(Y_j|\mathbf{x}_j)$ and $\text{var}(Y_j|\mathbf{x}_j)$. *Hint:* Define $e_j = [h(Y_j, \lambda) - h\{f(\mathbf{x}_j, \boldsymbol{\beta}), \lambda\}]/\{q(\mathbf{x}_j, \theta)\}$.

(b) Suppose we have data for which we believe $\text{var}(Y_j|\mathbf{x}_j)$ follows the model in (2.14) on page 40 of the class notes. Would it be possible to use model (1) with a suitable choice of q and the Box-Cox transformation on page 35 to arrive at a model on the transformed scale for which the variance is constant on the transformed scale? Explain.

(c) Consider the usual TBS model on page 36 of the notes with the Box-Cox transformation, for which $g(\mathbf{x}_j, \theta) \equiv 1$. Suppose the response is such that Y takes on only positive values. Is it always possible to transform such a response to normality using (2.11) for any value of λ ? Give an argument justifying your answer.

4. *Still more on transformations for nonlinear models.* The *Michaelis-Menten* (MM) model

$$f(x, \boldsymbol{\beta}) = \frac{Vx}{K + x} = \{\beta_0 + \beta_1/x\}^{-1}, \quad \beta_1 = 1/V, \beta_2 = K/V, \quad \boldsymbol{\beta} = (\beta_1, \beta_2)^T$$

is widely used to model data in biological, biochemical, and other situations. In fisheries research, it is called the Beverton-Holt spawner-recruit model. Often, data that are well represented by this model also exhibit nonconstant variance.

A number of methods have been proposed to estimate $\boldsymbol{\beta}$ when it is assumed that $E(Y|x) = f(x, \boldsymbol{\beta})$. Here, we explore how these may all be viewed as special cases of (1) in the previous problem. Specifically, consider the general TBS model for independent pairs (Y_j, x_j) , $j = 1, \dots, n$ given in (1) for some transformation h depending on a scalar parameter λ with $q(x_j, \theta) = x_j^\theta$. Throughout this problem, take h to be the Box-Cox transformation on page 35 of the notes. Note that x is a scalar here.

(a) Rather than fitting the (nonlinear) MM model to data on the original scale, perhaps modeling nonconstant variance explicitly, it has been traditional to try to “linearize” the model by using some sort of transformation. Two common such approaches are to consider the following models, which are usually written in the “classical” response = mean + error form:

(i) *Lineweaver-Burk:* $1/Y = \beta_0 + \beta_1/x + e$, where e is taken to be mean zero with variance σ^2 .

(ii) *Woolf:* $1/Y = \beta_0 + \beta_1/x + e/x$, where e is taken to be mean zero with variance σ^2 .

Note that in both cases, the model for conditional mean is linear in β_0 and β_1 , which is the appeal of these approaches.

Show that each of (i), (ii) is a special case of the general TBS model described above by finding the values of λ and θ to which they correspond. Using your result to Problem 3, give the approximate mean-variance relationship on the original scale that is being assumed for “small” error.

(b) Another popular approach has been to fit the so-called *Scatchard* model, which is usually written as

$$Y/x = (1/\beta_1) - (\beta_0/\beta_1)Y + e$$

for “errors” e with mean 0 and constant variance σ^2 . Verify that the Scatchard model corresponds to assuming that $Y|x$ has constant coefficient of variation, and find this coefficient

of variation. Show that the Scatchard model may also be regarded approximately, for “small” error, as a special case of the form of (1), finding the corresponding values for λ and θ .

5. On page 62 of the notes, we remarked that one may derive the IRWLS updating scheme by considering the estimating equation (3.12) on page 61, i.e.

$$\sum_{j=1}^n g^{-2}(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x}_j) \{Y_j - f(\mathbf{x}_j, \boldsymbol{\beta})\} f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta}) = \mathbf{0},$$

replacing all of $f(\mathbf{x}_j, \boldsymbol{\beta})$, $f_{\boldsymbol{\beta}}(\mathbf{x}_j, \boldsymbol{\beta})$, and $g^{-2}(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{x}_j)$ by linear approximations about some value $\boldsymbol{\beta}^*$ “close to” $\boldsymbol{\beta}$, and disregarding all “higher-order” terms. Carry out this argument to obtain the updating scheme on page 60, showing explicitly exactly which terms in the approximation you have disregarded and why.

6. “*Self-starting*” regression models. A popular feature in recent software for fitting nonlinear regression models is that of “self-starting” models. With this feature, the user is not required to determine starting values; rather, the form of the nonlinear model is exploited to generate starting values “automatically.” A very simple example to illustrate this idea is the monoexponential model

$$f(x, \boldsymbol{\beta}) = \beta_1 \exp(-\beta_2 x).$$

Under this model, $Y \approx \beta_1 \exp(-\beta_2 x)$; thus, $\log Y \approx \log \beta_1 - \beta_2 x$. Using this observation, starting values can be obtained by a simple linear regression of $\log Y_j$ on x_j . In nonlinear regression software, this popular model might be “built-in,” so that when the user calls it, the starting value procedure is automatically carried out first and the starting values fed into the fitting algorithm. In general, “self-starting” methods are often based on *ad hoc* ideas such as this and often perform quite well.

Here, we will consider a slightly more complex version of this idea. Consider the biexponential model parameterized as in (1.3) on page 8, i.e.

$$f(x, \boldsymbol{\beta}) = \beta_1 \exp(-\beta_2 x) + \beta_3 \exp(-\beta_4 x), \tag{2}$$

where we will assume $\beta_2 > \beta_4$ so that the second component on the right hand side of (2) represents the second phase of decay (see page 58).

- (a) Develop a “self-starting” method for model (2). There is no “right” or “wrong” way to do this; you are free to be as clever as you like. Provide a step-by-step written description of your method.

Hint: First consider only data that appear to be in the region of the “second phase of decay” first to obtain start values for β_3, β_4 . A true “self-starting” function will have a built-in convention for deciding which responses are in the second phase of decay that does not rely on any information from the user. In the context of model (2), this might be done as follows. Start with the observations at the last 3 time points. Behaving as if these were the only observations, use the approach described above for the monoexponential model to estimate β_3, β_4 . Then consider the observation 4th from the last and conduct a formal procedure based on usual regression theory to decide how likely it is that this observation at this time point comes from a normal distribution with mean on the straight line with this intercept and slope. If it is sufficiently unlikely, say less than 10%, stop and use the current estimates of β_3, β_4 as start values; otherwise, repeat this procedure including this 4th-from-last observation, apply

the “test” to the 5th-from-last observation, and so on until an observation is less than 10% likely. Now hold these fixed and obtain start values for β_1, β_2 by considering the form of (2).

(b) Write a program that implements your method. To test it, apply it to the data in the file `biexp.dat` available on the class web page. Using your IRWLS program, assuming $\text{var}(Y_j|x_j) = \sigma^2 f^{2\theta}(x_j, \boldsymbol{\beta})$ with $\theta = 1.0$, use your starting values to fit (2) to these data.