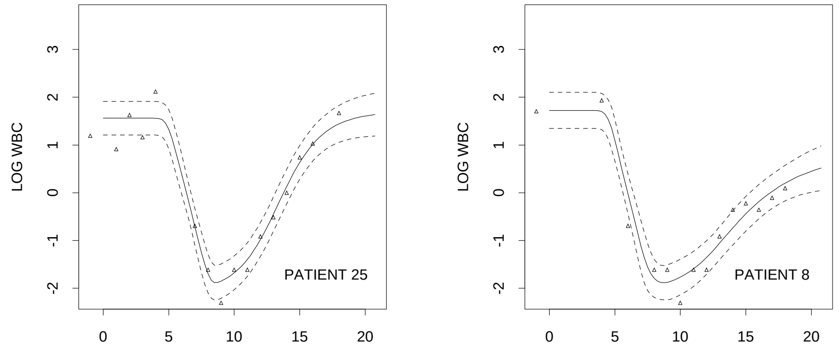
Assignment 7

Remember the example from class: We had data from patients undergoing high dose chemotherapy. The data records log white blood cell count (WBC) over time for $I = 52$ patients $i = 1, \ldots, I$. Here are some typical patients:



Blood cell counts for two typical patients: Data (triangles) and fitted profiles (solid lines).

In this assignement we only consider the data from the recovery phase (when WBC recovers from nadir back to base line level). The data are rescaled to be constrained between 0 and 1.0. Let y_{ij} denote the j-th observation for the i-th patient, taken on day t_{ij}, $i = 1, \ldots, 52$, $j = 1, \ldots, n_i$. We will consider four alternative models:

Model A: Independent models for each patient, i.e., $I = 52$ seperate models, one for each patient. With unknown logistic intercept β_{0i} and slope β_{1i}:

$$
\begin{aligned}
y_{i,j} &= \frac{exp(\beta_{0i} + \beta_{1i} t_{ij})}{1 + exp(\beta_{0i} + \beta_{1i} t_{ij})} + \epsilon_{ij} \quad i = 1, \ldots n \text{ and } j = 1, \ldots n_i \\
\epsilon_{ij} &\sim N(0, \sigma^2).
\end{aligned}
$$

We will estimate Model A by m.l.e. only, i.e., we need no prior probability model.

Model B: A longitudinal data model with normal/linear structure. To make a normal linear model meaningful we consider a transformation of y_{ij} to $z_{ij} = \log[y_{ij}/(1 - y_{ij})]$. For the transformed data we consider the following hierarchical model

$$
\begin{aligned}
z_{ij} &= \gamma_{0i} + \gamma_{1i} t_{ij} + \epsilon_{ij}, \ i = 1, \ldots, I, \quad j = 1, \ldots, n_i \\
&\quad\ \epsilon_{ij} \sim N(0, \sigma^2), \\
\gamma_{0i} &\sim N(\phi_1, \tau_1), \\
\gamma_{1i} | \gamma_{0i} &\sim N(\phi_2 + \phi_3(\gamma_{0i} - \phi_1), \tau_2), \\
\phi_j &\sim N(\mu_j, \rho_j), \ j = 1, 2, 3.
\end{aligned}
$$

Fix the hyperparameters $(\sigma^2, \tau_j, \rho_j, \mu_j)$ at some "reasonable" values. Plotting the transformed data z_{ij} should give you some idea (If you think that fixing prior parameters by looking at the data is bad you are right; but we don't have that oncology expert at hand for good prior elicitation).

Model C: A longitudinal model with a logistic likelihood and a single bivariate normal prior:

$$\begin{aligned} y_{i,j} &= \frac{exp(\beta_{0i} + \beta_{1i} t_{ij})}{1 + exp(\beta_{0i} + \beta_{1i} t_{ij})} + \epsilon_{ij} \quad i = 1, \ldots n \text{ and } j = 1, \ldots n_i \\ & \epsilon_{ij} \sim N(0, \sigma^2), \\ \beta_{0i} &\sim N(\phi_1, \tau_1), \\ \beta_{1i} | \beta_{0i} &\sim N(\phi_2 + \phi_3(\beta_{0i} - \phi_1), \tau_2), \\ \phi_j &\sim N(\mu_j, \rho_j), \;\; j = 1, 2, 3. \end{aligned}$$

Again, fix the hyperparameters $(\sigma^2, \tau_1, \tau_2, \mu_j, \rho_j)$ at some "reasonable" values.

Model D: Same as Model C, but with a mixture of normal prior:

$$\begin{aligned} \beta_{0i} &\sim \sum_{m=1}^{3} w_m N(\phi_{1m}, \tau_1), \\ \beta_{1i} | \beta_{0i} &\sim \sum_{m=1}^{3} w_m N(\phi_{2m} + \phi_{3m}(\beta_{0i} - \phi_1), \tau_2), \\ \phi_{jm} &\sim N(\mu_j, \rho_j), \;\; j = 1, 2, 3, \\ (w_1, w_2, w_3) &\sim Dir(1, 1, 1). \end{aligned}$$

Data:

The data is linked to the 214 homepage:

`Y`: A (52×12) matrix of responses y_{ij} with the i-th row corresponding to the i-th patient and the j-th column corresponding to the j-th observation. If less than 12 observations are recorded for a patient the row is padded with 0's.

`ni`: A (52×1) vector recording the number of observations for each patient, n_i, $i = 1, \ldots, 52$.

`day`: A (52×12) matrix of the sampling times t_{ij}.

`X`: A (52×2) matrix of treatments assigned to the patients. The first column reports dose of the anti-cancer agent (cyclophosphamide, in g/m^2). The second column is another drug given for hematologic support (GM-CSF, in $\mu g/kg$). You will not need these covariates for any of the Models A through D. I included it in case you want to do some exploration on your own.

`214.dta`: A big matrix containing all the above data. The format should be clear from reading the function `read.dta()` in the file `214.s`.

Questions:

In answering questions 1-4 below focus on relative merits and disadvantages of the respective models compared to the alternatives. Discuss assumptions of the respective models, and how each model relaxes some assumptions of more restrictive models. When you show figures include legends to provide a clear and complete description of everything that is on the figure. Describe completely what is graphed in the display; then point out salient features of the dsiplay, and briefly state the importance of these features[1].

1. Estimate Model A by m.l.e. An Splus program is linked to the 214 homepage. Be innovative about summarizing the inference. For example, you should probably make a scatterplot of the estimated pairs $(\hat{\beta}_{0i}, \hat{\beta}_{1i})$.

2a. Fit Model B. Find the marginal posterior distribution $p(\gamma_{0i}, \gamma_{1i}, i = 1, \ldots, 52|y)$ analytically (see the note at the end of this handout). Write down the analytic expression for $p(\gamma_{0i}, \gamma_{1i}, i = 1, \ldots, 52|y)$, and then use Splus or any other program to evaluate them. Consider some interesting ways of summarizing your inference. This should include a scatterplot of the pairs $(\bar{\gamma}_{0i}, \bar{\gamma}_{1i})$ where $\bar{\gamma}_{0i} = E(\gamma_{0i}|y)$ is the posterior mean of γ_{0i}.

2b. By simulation find the posterior predictive distribution for a future patient, $i = 53$. Denote with y the given data for patients $i = 1, \ldots, 52$. Plot $E(y_{ij}|t_{ij} = t, y)$ as a function of t, maybe with plus/minus one posterior predictive standard deviation margin. This should look like the two plots at the beginning of this handout. (*Note:* Easiest is probably to first generate predictive values for z_{ij}, $i = 53$, and then transform to y_{ij}).

3. Fit Model C. Estimate Model C using BUGS. See the `Epil` example on p30 of the BUGS example manual, volume 1, for a similar problem (I have not yet done this. So, please tell me about any problems you run into). Summarize your inference by plotting (a) a scatterplot of the posterior means for the pairs (β_{0i}, β_{1i}) (as in 2a); and (b) The posterior predictive profile for a future patient as in 2b.

4. Model D. Estimate Model D using BUGS. You should be able to use almost the same BUGS code as for Model C, adding only a few extra lines to allow for the mixture. Again, report the posterior means for the pairs (β_{0i}, β_{1i}) and the posterior predictive profile for a future patient, and any other informative summary of the inference which you can think of. Be innovative..

[1] From the ASA "Style Guide"

The Normal/Linear Model

For Question 2. you might need to use the following result (from Lindley, D.V. and Smith, A.F.M. (1972), "Bayes Estimates for the Linear Model", *JRSS, Series B*, pp. 1-18). Consider the normal/linear model

$$\begin{aligned} y &\sim N(A_1\theta, C_1) \\ \theta &\sim N(A_2\phi, C_2) \\ \phi &\sim N(\mu, C_3). \end{aligned}$$

Then the marginal posterior for θ is $p(\theta|y) = N(m, S)$ with

$$S^{-1} = A_1' C_1^{-1} A_1 + \{C_2 + A_2 C_3 A_2'\}^{-1} \text{ and } m = S(A_1' C_1^{-1} y + \{C_2 + A_2 C_3 A_2'\}^{-1} A_2 \mu).$$

The conditional posterior for ϕ given θ is $p(\phi|\theta, y) = p(\phi|\theta) = N(u, T)$ with

$$T^{-1} = A_2' C_2^{-1} A_2 + C_3^{-1} \text{ and } u = T(A_2' C_2^{-1} \theta + C_3^{-1} \mu).$$

Some References

Ask me if you want any of these, or any other related references.

Davidian, M., and Giltinan, D. M. (1995), "Nonlinear Models for Repeated Measurement Data". London: Chapman and Hall.
(This is a review of m.l.e. in longitudinal models).

Dellaportas, P. and Smith, A.M.F. (1993), "Bayesian analysis for generalized linear models via Gibbs sampling." *Applied Statistics* **42**, 443-459.
(Gibbs sampling implementation in models like Model C).

Wakefield, J.C. (1996), "The Bayesian analysis of population pharmacokinetic models," *Journal of the American Statistical Association*, **91**, 62-75.
(An appliction like Model C, in the specific context of pharmacokinetics.)

Wakefield, J.C., Smith, A.F.M., Racine-Poon, A., and Gelfand, A. (1994), "Bayesian analysis of linear and nonlinear population models using the Gibbs sample," *Applied Statistics*, **43**, 201-221.
(A nice review of the general structure in longitudinal and related models. I gave you copies of this paper in class).