APTS Statistical Computing: Assessment

The work provided here is intended to take students up to half a week to complete. Students should talk to their supervisors to find out whether or not their department requires this work as part of any formal accreditation process (APTS itself has no resources to assess or certify students). It is anticipated that departments will decide on the appropriate *level* of assessment locally, and may choose to drop some (or indeed all) of the parts, accordingly.

Consider a study investigating the efficacy of a sleeping pill. Patients reporting trouble sleeping are randomly allocated one of several treatment doses or a placebo, and have several covariates such as, weight, age and gender measured at the outset. Over the following months patients are contacted more or less regularly (but never more than once a week), and asked whether or not they slept well the previous night ('well' being defined in qualitative, but reasonably precise, terms).

The study therefore provides binary response data y_i on whether (1) or not (0) patients slept well at a series of times. A reasonable exploratory model of these data might have the form:

$$y_i|\mathbf{b} \sim \text{Bernouilli}(p_i)$$

where

$$\log\{p_i/(1-p_i)\} = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}$$

and $\mathbf{b} \sim N(\mathbf{0}, \mathbf{I}\sigma_b^2)$, there being one b_j for each subject. $\mathbf{X}_i$ is a row of a model matrix summarizing the dependence of the response on the covariates, and here $\mathbf{Z}_{ij}$ is 1 if y_i relates to the j^{th} subject and 0 otherwise. The idea here is that the random effects, b_j, represent intrinsic patient to patient variability (which we would not want to model as fixed effects, since that would make it difficult to generalize beyond the particular set of patients in the study, even if they were representative of a larger population of interest).

1. Show that
$$\frac{\partial p_i}{\partial b_j} = p_i(1-p_i)Z_{ij}$$

2. Write an R function to evaluate the log of the joint p.f. of $\mathbf{y}$ and $\mathbf{b}$, $\log\{f(\mathbf{y}, \mathbf{b})\}$, given $\mathbf{y}$, $\mathbf{b}$, $\mathbf{Z}$, $\mathbf{X}$, $\boldsymbol{\beta}$ and σ_b^2.

3. Write an R function to evaluate the gradient vector of derivatives of $\log\{f(\mathbf{y}, \mathbf{b})\}$ w.r.t. the elements of $\mathbf{b}$.

4. Write code to evaluate the log likelihood, $l(\boldsymbol{\beta}, \sigma_b^2)$, by Laplace approximation, and by importance sampling.

5. Write code to estimate $\boldsymbol{\beta}$ and σ_b^2 by MLE, given data $\mathbf{y}$, and the model matrices $\mathbf{X}$ and $\mathbf{Z}$.

6. Test your code with suitable simulated data.