1. **Unbiasedness of the weighted generalized estimating equation.** In Section 8.7 of the notes, we saw that, with missing data, the usual linear estimating equation (8.70) for estimation of $\beta$ in a population-averaged model of the form (8.3) with exogenous among-individual covariates is not unbiased unless the missingness mechanism is missing completely at random (MCAR) and thus need not lead to a consistent estimator for the true value of $\beta$ under the assumption of a missing at random (MAR) mechanism.

One approach to modifying the equation (8.70) to render it unbiased under a MAR mechanism is to modify it by weighting the contribution to the equation for each individual $i$ at each time point by the reciprocal of the probability of having an observed response at that time point, given her/his past history, as in (8.81), that is

$$\sum_{i=1}^{m} X_{i}^{T}(x_{i}, \beta) V_{i}^{-1}(\beta, \xi, x_{i}) W_{i} \begin{pmatrix} Z_{i1} - f_{1}(a_{i}, \beta) \\ \vdots \\ Z_{in} - f_{n}(a_{i}, \beta) \end{pmatrix} = 0, \quad (1)$$

where

$$W_{i} = \text{diag} \left( \frac{R_{i1}}{\pi_{1}}, \frac{R_{i2}}{\pi_{2}(H_{i1})}, \ldots, \frac{R_{in}}{\pi_{n}(H_{i,n-1})} \right),$$

and $X_{i}^{T}(x_{i}/\beta)$ and $V_{i}(\beta, \xi, x_{i})$ are defined in and below (8.69).

(a) For an individual for whom $R_{i1} = 1, \ldots, R_{ij} = 1, R_{ij+1} = 0, \ldots, R_{in} = 0$, so who drops out at time $j + 1$, express the summand in (1) in terms of $Y_{i1}, \ldots, Y_{ij}$, defining any additional symbols necessary.

(b) Assuming that the cause-specific hazard functions $\lambda_{j}(H_{ij-1})$ are correctly specified and the missing mechanism is MAR, show that (1) is an unbiased estimating equation.

2. **Laplace approximation.** Consider the nonlinear mixed effects model with first stage individual model with second stage model substituted, of the form

$$E(Y_{i} \mid x_{i}, b_{i}) = f_{i}(x_{i}, \beta, b_{i}), \quad \text{var}(Y_{i} \mid x_{i}, b_{i}) = R_{i}(\gamma, x_{i}), \quad (2)$$

so that the within-individual covariance matrix does not depend on $\beta_{i}$ and thus does not depend on $\beta$ or $b_{i}$. Assume further that $b_{i}$ are independent of $x_{i}$ and satisfy $b_{i} \sim \mathcal{N}(0, D)$ and that that the distribution of $Y_{i}$ given $x_{i}$ and $b_{i}$ is normal with moments given in (2).

Consider the derivation of the approximate expression for $p(Y_{i} \mid x_{i}; \beta, \gamma, D)$ in (9.85) using Laplace’s approximation. Using (9.86) and your matrix algebra prowess, show that (9.85) can be expressed equivalently as (9.87). Thus, you will have shown that the contribution of individual $i$ to the likelihood (9.46) can be approximated by a normal density with mean and covariance matrix given in the first bullet on page 322.

**Hints:** You will probably have to look up some relationships having to do with determinants. It will behoove you to define

$$u_{i} = Y_{i} + Z_{i}(x_{i}, \beta, \hat{b}_{i}) \hat{b}_{i}$$

and to use shorthand notation such as $Z_{i} = Z_{i}(x_{i}, \beta, \hat{b}_{i})$, $f_{i} = f_{i}(x_{i}, \beta, \hat{b}_{i})$, etc.
3. **Onychomycosis clinical trial.** In the file `toenail.dat` you will find data from a clinical trial involving \( m = 294 \) patients suffering from onychomycosis, otherwise known as toenail fungus infection. Patients were randomly assigned to one of two oral antifungal treatments (Itraconazole and Terbinafine). Patients were evaluated for the degree of onycholysis (the degree of separation of the nail plate from the nail bed) at baseline (week 0) and at weeks 4, 8, 12, 24, 36, and 48 thereafter. The onycholysis outcome variable is binary (0=none or mild versus 1=moderate or severe). In the data set, the variable month denotes the exact timing of measurements in months, and the variable visit denotes the visit number (visit numbers 1-7 correspond to scheduled visits at 0, 4, 8, 12, 24, 36, and 48 weeks). The columns are as follows:

1. subject ID
2. response (0=none or mild, 1=moderate or severe)
3. treatment ((0=Itraconazole, 1=Terbinafine)
4. month
5. visit

Let \( Y_{ij} \) be the response for subject \( i \) at month \( t_{ij} \), \( i = 1, \ldots, m \), and let \( Y_i = (Y_{i1}, \ldots, Y_{im})^T \). Let \( \delta_i = 0 \) if \( i \) was assigned to placebo and \( \delta_i = 1 \) if \( i \) was assigned to Itraconazole. Here, there are no within-individual covariates, so that \( z_{ij} = t_{ij} \), and the among-individual covariate \( a_i = \delta_i \), so that \( x_{ij} = (t_{ij}, \delta_i)^T \) and \( x_i = (t_{i1}, \ldots, t_{im}, \delta_i)^T \).

Consider the following two models for these data:

1. Assume that for each \( i \) and \( j = 1, \ldots, n_i \)

   \[
   E(Y_{ij}|z_{ij}, \beta_i) = \frac{\exp(\beta_{0i} + \beta_{1i}t_{ij})}{1 + \exp(\beta_{0i} + \beta_{1i}t_{ij})},
   \]

   where
   \[
   \beta_{0i} = \beta_0 + b_{1i} \\
   \beta_{1i} = \beta_1 + \beta_2 \delta_i;
   \]

   \( b_{1i} \sim N(0, D) \); and \( Y_{ij}, j = 1, \ldots, n_i \), are mutually independent conditional on \( z_i \) and \( b_{1i} \).

2. Assume that for each \( i \) and \( j = 1, \ldots, n_i \)

   \[
   E(Y_{ij}|x_i) = E(Y_{ij}|x_{ij}) = \frac{\exp(\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij} \delta_i)}{1 + \exp(\beta_0 + \beta_1 t_{ij} + \beta_2 t_{ij} \delta_i)},
   \]

   \[
   \text{var}(Y_{ij}|x_i) = \text{var}(Y_{ij}|x_{ij}) = E(Y_{ij}|x_{ij})\{1 - E(Y_{ij}|x_{ij})\}, \quad \text{corr}(Y_{ij}|x_i) = \Gamma_i(\alpha, x_i),
   \]

   where \( \Gamma(\alpha, x_i) \) is some correlation matrix.

(a) Give interpretations of the parameters \( \beta_0 \), \( \beta_1 \), and \( \beta_2 \) in model (1).

(b) Give interpretations of the parameters \( \beta_0 \), \( \beta_1 \), and \( \beta_2 \) in model (2).

(c) Consider model (1). An alternative to (4) is

   \[
   \beta_{0i} = \beta_0 + b_{1i} \\
   \beta_{1i} = \beta_1 + \beta_2 \delta_i + b_{2i},
   \]

   \( (7) \)
where now \( \mathbf{b}_i = (b_{1i}, b_{2i})^T \sim \mathcal{N}(\mathbf{0}, \mathbf{D}) \), with \( Y_{ij}, j = 1, \ldots, n_i \), are mutually independent conditional on \( \mathbf{z}_i \) and \( \mathbf{b}_i \).

In this part of this problem, you will get to observe how well various methods for fitting models of the form (3)-(4) and (3)-(7) based on different analytical and numerical ways of approximating the likelihood agree.

Fit each of the models defined by (3)-(4) and (3)-(7) using the following methods:

- An analytical approximation to the likelihood based on a linear approximation of the model about the random effects equal to zero.
- An analytical approximation to the likelihood based on a linear approximation of the model about the random effects equal to current empirical Bayes estimates.
- A numerical approximation to the likelihood in which the integrals are approximated using a full Laplace approximation.
- A numerical approximation to the likelihood in which the integrals are approximated using adaptive Gaussian quadrature.

Compare the resulting estimates of \( \mathbf{\beta} = (\beta_0, \beta_1, \beta_2)^T \) and \( \mathbf{D}/\mathbf{D} \) in each case and comment the results. Which of models (4) or (7) would you prefer to adopt, and which method would you feel most comfortable using?

**Hint:** It may well be that you will find it difficult to implement all of the fits in (a), especially for second stage model (7).

(d) Fit model (2) in (5)-(6) taking \( \Gamma_i(\alpha, \mathbf{x}_i) \) to be a compound symmetric matrix. Comment on how the results compare to those obtained in (c).

4. **Pharmacokinetics of Itraconazole.** The Itraconazole dosing regimen used in the toenail fungus clinical trial in the previous problem was based on a previously conducted study of the pharmacokinetics of Itraconazole in \( m = 100 \) onychomycosis sufferers. In this study, each subject was given an oral dose of 200 mg of Itraconazole, and blood samples were drawn and assayed for Itraconazole concentrations at several time points during the next 36 hours. These data are in the file `itrapk.dat`, with columns

1. subject ID
2. gender (0=female, 1=male)
3. creatinine clearance (ml/min)
4. weight (kg)
5. age (years)
6. Itraconazole concentration (µg/ml)
7. time (hours)

Creatinine clearance is a common measure of kidney (renal) function; a value < 50 ml/min is usually considered to indicate renal impairment. Renal impairment typically is associated with the elimination characteristics of a drug; individuals with renal impairment tend to eliminate the drug at a slower rate than healthy subjects.

As we have discussed, it is well established that drug concentrations at the individual patient level exhibit nonconstant variance that is often represented the “power of the mean” variance model. It is also well established that the distributions of pharmacokinetic parameters in the population tend to be right skewed.
Based on this study, the investigators were interested in characterizing the typical or average pharmacokinetic properties of absorption, distribution, and elimination for Itraconazole in the onychomycosis patient population. They also were interested in whether or not the data contain evidence that these pharmacokinetic characteristics are systematically associated with subject characteristics (gender, renal impairment, weight, or age here) and, if so, with which characteristics. Finally, they hoped to obtain estimates of these characteristics for each patient in the study, which they would then use to evaluate appropriate individualized dosing regimens for these patients.

Identify an appropriate pharmacokinetic model and, using it, carry out analyses to address these questions and write a brief report summarizing what you did and the results. Be sure to describe how you interpreted and formalized the questions of interest. Comment on how confident you feel about the reliability of the inferences and conclusions.

Please turn in your code and output along with your report (you can edit the output to include only the portions that pertain directly to your report and embed it in your report if you like).