10 Additional Topics

10.1 Introduction

Chapters 1-9 present an overview of the ***fundamental topics*** that ordinarily comprise a first course on longitudinal data analysis. Understanding of these fundamentals is preparation for study of ***additional topics*** connected with modeling and analysis of longitudinal and other correlated/clustered data and for reading the current literature on longitudinal data methods.

In this chapter, we provide an introduction to several of these additional topics.

DATA, RESTATED: As in previous chapters, the observed data are

$$(\boldsymbol{Y}_i, \boldsymbol{z}_i, \boldsymbol{a}_i) = (\boldsymbol{Y}_i, \boldsymbol{x}_i), \quad , i = 1, \ldots, m, \tag{10.1}$$

independent across i, where $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{in_i})^T$, with Y_{ij} recorded at time t_{ij}, $j = 1, \ldots, n_i$ (possibly different times for different individuals); $\boldsymbol{z}_i = (\boldsymbol{z}_{i1}^T, \ldots, \boldsymbol{z}_{in_i}^T)^T$, comprising ***within-individual*** covariate information $\boldsymbol{u}_i$ and the ***times*** t_{ij}; $\boldsymbol{a}_i$ is a vector of ***among-individual*** covariates; and $\boldsymbol{x}_i = (\boldsymbol{z}_i^T, \boldsymbol{a}_i^T)^T$. We take the ***among-individual*** covariates to be ***time-independent*** unless otherwise stated.

10.2 Bayesian formulation of hierarchical models

BAYESIAN VS. FREQUENTIST INFERENCE: We have restricted attention to so-called frequentist-type inference. In particular, for most of the methods discussed, we have considered model parameters to be ***fixed quantities*** and viewed data to be a ***repeatable sample*** arising as a result of some generative data process (e.g., according to an experimental design or data gathering scheme). Inference regarding the (fixed) values of the parameters of interest is by reference to (ideally) an exact, or in most cases, an ***approximate*** (derived via ***large sample theory***), ***sampling distribution***. In this view, probability refers to the conceptual repeated sampling if the generative process were to be repeated infinitely.

Bayesian inference is based on a different point of view in which the data are viewed as ***fixed***, model parameters are viewed as ***random***, and, in the classical Bayesian paradigm, probability refers to "***degree of belief***" regarding the values of the parameters.

Given a ***prior distribution*** specifying degree of belief before the data are observed, ***Bayes theorem*** is used to obtain the ***posterior distribution***, reflecting the ***updated*** degree of belief having seen the data. "Estimation" of parameters is usually based on the ***mode*** of the posterior density, and assessments of uncertainty are based on the variance (standard deviation) of the posterior.

It can be shown that, despite the apparent divergence of these frameworks, they can lead to very ***similar results***. Thus, in modern statistics, they are often viewed as ***complementary strategies*** for framing scientific inquiry. Owing to the ***high-dimensional integration*** involved in applying Bayes theorem in complex models, before the latter half of the twentieth century, it was often ***prohibitive*** to implement Bayesian models. However, with the computational advances that began in the last quarter of the twentieth century, and in particular the use and refinement of ***Markov chain Monte Carlo*** (MCMC) techniques, formulation and implementation of complex statistical models from a Bayesian perspective is now ***commonplace***. Indeed, with ***noninformative prior*** specifications, models of both types overlap, and a Bayesian framework with fitting via MCMC techniques is sometimes viewed as a convenient way to implement frequentist analyses.

HIERARCHICAL MODELS FROM A BAYESIAN PERPSECTIVE: Hierarchical models such as the linear, generalized linear, and nonlinear mixed effects models discussed in Chapters 6 and 9 are placed ***naturally*** in a Bayesian framework, as we now demonstrate. Because the latter models subsume the former, we present this in the case of a general nonlinear mixed effects model.

From the Bayesian perspective, the model is a ***three-stage*** hierarchy, and the model parameters β, γ, and $\boldsymbol{D}$ involved in the usual two-stage hierarchy (9.8)-(9.9) are viewed as ***random vectors***. In a ***classical*** Bayesian formulation, ***full parametric distributional assumptions*** are made at each stage, although this can be relaxed, in particular for the distribution of the random effects discussed briefly later. For our discusion here, we adopt full parametric assumptions, so we write the stage 1 model differently from the general specification in (9.8), as follows.

Stage 1 - Individual model. Given a model f for $E(Y_{ij}|\boldsymbol{z}_{ij}, \beta_i$ possibly ***nonlinear*** in β_i, the random vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, m$, are assumed to satisfy

$$\boldsymbol{Y}_i|\boldsymbol{z}_i, \beta_i \sim p(\boldsymbol{y}_i|\boldsymbol{z}_i, \beta_i; \gamma), \tag{10.2}$$

where $p(\boldsymbol{y}_i|\boldsymbol{z}_i, \beta_i; \gamma)$ is a parametric density such that

$$E(\boldsymbol{Y}_i|\boldsymbol{z}_i, \beta_i) = \boldsymbol{f}_i(\boldsymbol{z}_i, \beta_i), \quad \text{var}(\boldsymbol{Y}_i|\boldsymbol{z}_i, \beta_i) = \boldsymbol{R}_i(\beta_i, \gamma, \boldsymbol{z}_i).$$

For continuous responses, $p(\boldsymbol{y}_i|\boldsymbol{z}_i, \beta_i; \gamma)$ in (10.2) is ordinarily taken to be the $\mathcal{N}\{\boldsymbol{f}_i(\boldsymbol{z}_i, \beta_i), \boldsymbol{R}_i(\beta_i, \gamma, \boldsymbol{z}_i)\}$ density.

For responses of the "generalized linear model type," the Y_{ij}, $j = 1, \ldots, n_i$, are taken to be ***conditionally independent*** given $\boldsymbol{z}_i$ and β_i, so that

$$p(\boldsymbol{y}_i|\boldsymbol{z}_i, \beta_i; \gamma) = \prod_{j=1}^{n_i} p(y_{ij}|\boldsymbol{z}_{ij}, \beta_i; \gamma);$$

and $p(y_{ij}|\boldsymbol{z}_{ij}, \beta_i; \gamma)$ is a ***scaled exponential family*** density for each j appropriate to the form of the response, so that $\boldsymbol{R}_i(\beta_i, \gamma, \boldsymbol{z}_i)$ is of necessity a ***diagonal matrix*** whose diagonal elements are determined by the ***variance function*** corresponding to the particular scaled exponential family density.

Stage 2 - Population model. The individual-specific parameter β_i is assumed to be a function of ***among-individual covariates*** $\boldsymbol{a}_i$, ***fixed effects*** β ($p \times 1$), and ***random effects*** $\boldsymbol{b}_i$ ($q \times 1$), namely,

$$\beta_i = \boldsymbol{d}(\boldsymbol{a}_i, \beta, \boldsymbol{b}_i), \qquad \boldsymbol{b}_i \sim p(\boldsymbol{b}_i|\boldsymbol{D}), \tag{10.3}$$

where $\boldsymbol{d}$ is a k-dimensional vector of possibly ***nonlinear*** functions of $\boldsymbol{a}_i$, β, and $\boldsymbol{b}_i$; and $\boldsymbol{b}_i$ density $p(\boldsymbol{b}_i|\boldsymbol{D})$, usually assumed to be the $\mathcal{N}(\boldsymbol{0}, \boldsymbol{D})$ density. As in Chapter 9, this can be relaxed to allow the distribution of $\boldsymbol{b}_i$ to depend on $\boldsymbol{a}_i$.

In the classical Bayesian literature, the distribution of $\boldsymbol{b}_i$ or that implied for β_i are sometimes confusingly referred to as the ***prior distribution***.

Stage 3 - Hyperprior distribution. $(\beta, \gamma, \boldsymbol{D})$ are assumed to have joint density

$$(\beta, \gamma, \boldsymbol{D}) \sim p(\beta, \gamma, \boldsymbol{D}|\Omega), \tag{10.4}$$

where Ω are ***known hyperparameters*** characterizing this density.

REMARKS:

- From the classical Bayesian viewpoint, the hyperprior reflects ***prior beliefs*** about the values of β, γ, and $\boldsymbol{D}$.
- Ordinarily, in practice, the joint density (10.4) is written as

  $$p(\beta, \gamma, \boldsymbol{D}|\Omega) = p(\beta|\Omega_1)\, p(\gamma|\Omega_2)\, p(\boldsymbol{D}|\Omega_3), \tag{10.5}$$

  so that β, γ, and $\boldsymbol{D}$ are taken to be independent, and hyperpriors for them are specified separately. The individual components (10.5) of the hyperprior (10.4) are usually taken to reflect ***weak knowledge*** of the values of the parameters.

- An advantage of this formulation is that the hyperprior is a natural way to incorporate ***historical*** or other information about the parameters in to the overall model. For example, if it is known from past studies that there is a ***range of plausible values*** for β, the hyperprior for β can be taken to concentrate on that range.

- This has been used advantageously in implementation of ***highly complex*** models for ***pharmacokinetics***, ***viral dynamics***, and other phenomena that can be represented as ***compartmental systems*** involving numerous parameters, as we discuss shortly.

- As an example, in the special case where $p(\boldsymbol{y}_i|\boldsymbol{z}_i, \beta_i; \gamma)$ is the $\mathcal{N}\{\boldsymbol{f}_i(\boldsymbol{z}_i, \beta_i), \boldsymbol{R}_i(\beta_i, \gamma, \boldsymbol{z}_i)\}$ distribution, where $\boldsymbol{R}_i(\beta_i, \gamma, \boldsymbol{z}_i) = \sigma^2 \text{diag}\{f^{2\delta}(\boldsymbol{z}_{i1}, \beta_i), \ldots, f^{2\delta}(\boldsymbol{z}_{in_i}, \beta_i)\}$, and the Y_{ij} are conditionally independent with $\gamma = (\sigma^2, \delta)^T$, common specifications for the components of (10.5) are

$$\beta \sim \mathcal{N}(\beta^*, \boldsymbol{H}), \qquad \sigma^{-2} \sim \text{Ga}(\nu_0/2, \nu_0\tau_0/2),$$

$$\boldsymbol{D}^{-1} \sim \text{Wi}\{(\rho\boldsymbol{D}^*)^{-1}, \rho\}, \qquad \delta \sim U(0, \delta_0), \tag{10.6}$$

  where $\text{Ga}(\cdot, \cdot)$ $\text{Wi}(\cdot, \cdot)$, and $U(\cdot, \cdot)$ denote the gamma, Wishart, and uniform distributions, respectively; and the ***hyperparameters*** β^*, $\boldsymbol{H}$, ν_0, τ_0, ρ, $\boldsymbol{D}^*$, and δ_0 are taken to be known.

- Hyperprior specifications for ***generalized linear mixed effects models*** are discussed by Rabe-Hesketh and Skrondal (2009) and references therein; those for ***more complex*** general nonlinear mixed models, and especially in the context of pharmacokinetics, are discussed by Wakefield et al. (1994), Wakefield (1996), Rosner and Müller (1994), and Müller and Rosner (1997), among many others.

POSTERIOR DISTRIBUTIONS: Given the above hierarchy, it is straightforward to deduce expressions for the ***marginal posterior densities*** of each of the parameter β, γ, and $\boldsymbol{D}$. We can express the stage 1 individual model (10.2) density equivalently by substituting the stage 2 population model (10.3) as

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i, \boldsymbol{b}_i; \beta, \gamma). \tag{10.7}$$

Letting as usual $\boldsymbol{Y}$ and $\boldsymbol{b}$ denote the "stacked" vectors of the individual response vectors $\boldsymbol{Y}_i$ and random effects $\boldsymbol{b}_i$, by the conditional independence of $\boldsymbol{Y}_i$ given $\boldsymbol{z}_i$ and $\boldsymbol{b}_i$ and among the $\boldsymbol{b}_i$, the densities of $\boldsymbol{Y}$ and $\boldsymbol{b}$ are the products of the individual densities in (10.7) and (10.3), which we can write as

$$p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{b}; \beta, \gamma) \quad \text{and} \quad p(\boldsymbol{b}|\boldsymbol{D}).$$

We take the entire analysis to be conditional on $\boldsymbol{x}$, as in prior chapters. Then the ***joint posterior density*** of $(\beta, \gamma, \boldsymbol{D})$, where we implicitly condition on $\boldsymbol{x}$ and recall that the hyperparameters Ω are ***known***, is seen to be

$$p(\beta, \gamma, \boldsymbol{D}|\boldsymbol{y}; \boldsymbol{x}, \Omega) = \frac{p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{b}; \beta, \gamma)\, p(\boldsymbol{b}|\boldsymbol{D})\, p(\beta, \gamma, \boldsymbol{D}|\Omega)}{\int\int\int\int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{b}; \beta, \gamma)\, p(\boldsymbol{b}|\boldsymbol{D})\, p(\beta, \gamma, \boldsymbol{D}|\Omega)\, d\beta\, d\gamma\, d\boldsymbol{b}\, d\boldsymbol{D}}. \tag{10.8}$$

To obtain the ***marginal posterior densities*** of β, γ, and $\boldsymbol{D}$, whose ***modes*** are the Bayesian "***estimators***" it is necessary ***integrate*** (10.8) with respect to the rest of the parameter; e.g., to obtain the posterior for β, $p(\beta|\boldsymbol{y}; \boldsymbol{x}, \Omega)$,

$$p(\beta|\boldsymbol{y}; \boldsymbol{x}, \Omega) = \frac{\int\int\int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{b}; \beta, \gamma)\, p(\boldsymbol{b}|\boldsymbol{D})\, p(\beta, \gamma, \boldsymbol{D}|\Omega)\, d\gamma\, d\boldsymbol{b}\, d\boldsymbol{D}}{\int\int\int\int p(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{b}; \beta, \gamma)\, p(\boldsymbol{b}|\boldsymbol{D})\, p(\beta, \gamma, \boldsymbol{D}|\Omega)\, d\beta\, d\gamma\, d\boldsymbol{b}\, d\boldsymbol{D}} \tag{10.9}$$

- Clearly, the potentially ***high-dimensional integration*** in (10.9) and the analogous expressions for the marginal posteriors of the other parameters is ***not analytically tractable*** in general.

IMPLEMENTATION BY MARKOV CHAIN MONTE CARLO SIMULATION: In the early 1990s, the use of ***MCMC techniques*** as a way to "***do***" the required integrations ***numerically*** was popularized. These techniques employ a clever scheme that leads to ***simulated draws*** from the posterior distribution of the parameters. These simulated values can them be used to ***construct numerically*** any ***functional of the posterior distribution*** desired; e.g., posterior modes for each of the parameters and the posterior variance.

A course in Bayesian inference covers these methods in detail; here, we just remark on the simplest version of this technique, the ***Gibbs sampler***, which relies on the premise that the conditional distributions of each parameter given all the others and the data might have forms from which it is ***straightforward*** to simulate random deviates.

Generically, the basic idea is as follows. Suppose we have J random variables $(U_1, \ldots, U_J)$ with joint density $p(u_1, \ldots, u_J)$, and we would like to find the marginal distributions $p(u_j)$, $j = 1, \ldots, J$. Assume that the joint density is uniquely determined (not automatic!) by the ***full conditional densities***

$$p(u_j|u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_J), \quad j = 1, \ldots, J,$$

from which it is "easy" to sample.

The Gibbs sampler is an ***iterative algorithm*** for obtaining a sample from the joint distribution based on the full conditional distributions that can then be used to obtain the marginals. Given a set of initial values $(u_1^{(0)}, \ldots, u_J^{(0)})$, at the ℓth iteration, one generates random variates from the full conditionals as follows:

$$\begin{aligned} U_1^{(\ell+1)} &\sim p(u_1 | u_2^{(\ell)}, u_3^{(\ell)}, \ldots, u_J^{(\ell)}) \\ U_2^{(\ell+1)} &\sim p(u_2 | u_1^{(\ell+1)}, u_3^{(\ell)}, \ldots, u_J^{(\ell)}) \\ &\vdots \\ U_J^{(\ell+1)} &\sim p(u_J | u_1^{(\ell+1)}, u_2^{(\ell+1)}, \ldots, u_{J-1}^{(\ell+1)}) \end{aligned} \tag{10.10}$$

After T iterations, we have a realization of the random vector $(U_1^{(T)}, \ldots, U_J^{(T)})$. It can be shown that the sequence generated in this way is a ***Markov chain*** with ***stationary distribution*** $p(u_1, \ldots, u_J)$. It thus follows that, as $T \to \infty$, this random vector ***tends in distribution*** to a draw from the joint distribution $p(u_1, \ldots, u_J)$ of interest.

- This suggests that one can obtain a sample of size L from $p(u_1, \ldots, u_J)$ by obtaining ***one long chain*** of length T and, after an initial "***burn-in***" period after which the chain is thought to have "***stabilized***," collecting L ***suitably spaced*** realizations from the chain (to eliminate correlation among them).
- Alternatively, one can perform L ***parallel chains***, each of length T, and take the ***final realization*** from each.
- This final sample of L realizations from the joint distribution can then be used to construct the desired functionals, such as ***marginal posterior summaries*** like the mode, mean, and variance.

This generic scheme is used in the context of ***hierarchical models*** by identifying each element of $(U_1, \ldots, U_J)$ with the parameters of the model; e.g., in the example above, β, γ, $\boldsymbol{D}$, as well as β_i or $\boldsymbol{b}_i$, $i = 1, \ldots, m$. One then derives the full conditional distributions, at least up to a ***proportionality constant***, from the assumptions embodied in the three-stage hierarchy.

In the particular case of the ***normal model with nonconstant variance*** above and ***hyperprior*** specifications in (10.6), it can be shown that the full conditional distributions (10.10) are as follows; the diligent student may want to verify this.

$$(\beta|\boldsymbol{y},\sigma^2,\delta,\boldsymbol{D},\beta_i,\ i=1,\ldots,m)\sim\mathcal{N}\{\boldsymbol{\Lambda}(m\boldsymbol{D}^{-1}\overline{\beta}+\boldsymbol{H}^{-1}\beta^*),\boldsymbol{\Lambda}\},$$

$$\boldsymbol{\Lambda}^{-1}=m\boldsymbol{D}^{-1}+\boldsymbol{H}^{-1},\qquad \overline{\beta}=m^{-1}\sum_{i=1}^{m}\beta_i,$$

$$(\boldsymbol{D}^{-1}|\boldsymbol{y},\beta,\sigma^2,\delta,\beta_i,\ i=1,\ldots,m)\sim \text{Wi}\left[\left\{\sum_{i=1}^{m}(\beta_i-\beta)\beta_i-\beta)^T+\rho\boldsymbol{D}^*\right\}^{-1},m+\rho\right],$$

$$(\sigma^{-2}|\boldsymbol{y},\beta,\delta,\boldsymbol{D},\beta_i,\ i=1,\ldots,m)\sim \text{Ga}\{(\nu_0+N)/2,A_0\},$$

$$A_0=\left[\sum_{i=1}^{m}\{\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{z}_i,\beta_i)\}^T\boldsymbol{R}_i^{-1}(\beta_i,\gamma,\boldsymbol{z}_i)\{\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{z}_i,\beta_i)\}+\nu_0\tau_0\right]/2.$$

The full conditional $(\beta_i|\boldsymbol{y},\beta,\sigma^2,\boldsymbol{D},\beta_k,\ k\neq i)$ is ***proportional to***

$$\exp\left[-\frac{1}{2\sigma^2}\{\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{z}_i,\beta_i)\}^T\boldsymbol{R}_i^{-1}(\beta_i,\gamma,\boldsymbol{z}_i)\{\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{z}_i,\beta_i)\}\right]$$

$$\times\exp\left\{-(\beta_i-\beta)^T\boldsymbol{D}^{-1}(\beta_i-\beta)/2\right\}\sigma|\boldsymbol{R}_i(\beta_i,\gamma,\boldsymbol{z}_i)|^{-1/2}.$$

The full conditional $(\delta|\boldsymbol{y},\beta,\sigma^2,\boldsymbol{D},\beta_i,\ i=1,\ldots,m)$ is proportional to

$$\prod_{i=1}^{m}\exp\left[-\frac{1}{2\sigma^2}\{\boldsymbol{y}_i-\boldsymbol{f}_i(\boldsymbol{z}_i,\beta_i)\}^T\boldsymbol{R}_i^{-1}(\beta_i,\gamma,\boldsymbol{z}_i)\right]\sigma|\boldsymbol{R}_i(\beta_i,\gamma,\boldsymbol{z}_i)|^{-1/2}.$$

Although the idea is simple, there are ***several challenges*** for implementation.

- There is a need to ***monitor convergence*** of the chain(s) to feel confident that they have stabilized, and diagnostic techniques for doing so have been developed. The required length of the ***burn-in period*** is also an issue.
- Gibbs sampling has obvious appeal when it is ***straightforward to sample*** from all the full conditional distributions. However, it is ordinarily the case in ***complex statistical models*** that this is not always possible. In the example above, sampling from the full conditionals for β, σ^2, and $\boldsymbol{D}$ is straightforward using standard procedures for ***random variate generation*** from the normal, gamma, and Wishart distributions, which are available in popular software.
- However, sampling from those for the β_i and δ is ***more challenging*** ; this is a consequence of the ***nonlinearity*** of the model f in β_i. One must resort to techniques such as ***rejection sampling*** for which the target distribution to be sampled is known only up to a proportionality constant. Other methods, such as importance sampling, can also be employed. Alternatively, one can embed embed a random variate sampling scheme based on the ***Metropolis-Hastings algorithm***. Usually, the choice and implementation of method depends on the ***particular model*** and must be tailored to the specific problem.

- Accordingly, implementation of the Bayesian formulation of the general nonlinear mixed effects model can require some ***sophistication*** on the part of the user.

- Software such as ***BUGS*** (Bayesian inference Using Gibbs Sampling) accommodates these situations. An interface to BUGS for population pharmacokinetic (and pharmacodynamic) analysis that has many popular PK models built-in and can accommodate complex individual dosing histories, ***PKBugs***, is also available. However, the user must have ***sufficient background*** in MCMC techniques, and especially appreciation for the issues of practical implementation discussed above.

Further discussion is beyond our scope here. Some classic papers discussing implementation of the Bayesian formulation of nonlinear mixed effects models are Wakefield, Smith,, and Racine-Poon (1994), Wakefield (1996), and Bennett, Racine-Poon, and Wakefield (1996). Rosner and Müller (1994), Wakefield (1996), and Müller and Rosner (1997) discuss specific pharmacokinetic applications.

10.3 Complex nonlinear models

As the quantitative and biological sciences continue to ***converge***, complex ***mathematical models of biological systems*** have become commonplace. ***Nonlinear dynamical systems*** models, of which the one-compartment PK models we have discussed in previous chapters are trivial, simple cases, are used in a number of application areas. It is natural to ***embed*** these complex ***mathematical models*** in the ***statistical*** nonlinear mixed effects model framework to address scientific questions of interest, as the following two examples demonstrate.

PHYSIOLOGICALLY-BASED PHARMACOKINETIC MODELS: Ordinary ***compartmental models*** of pharmacokinetics for the study of the disposition of drugs and biologics in humans are typically ***gross simplifications*** of the physiology involved. Although these models can be ***extraordinary useful approximations***, more sophisticated such models are required to address scientific questions in some key settings.

Toxicokinetics refers to the study of ***pharmacokinetics*** of environmental, chemical, or other agents in the context of assessment of their possible ***toxic effects***.

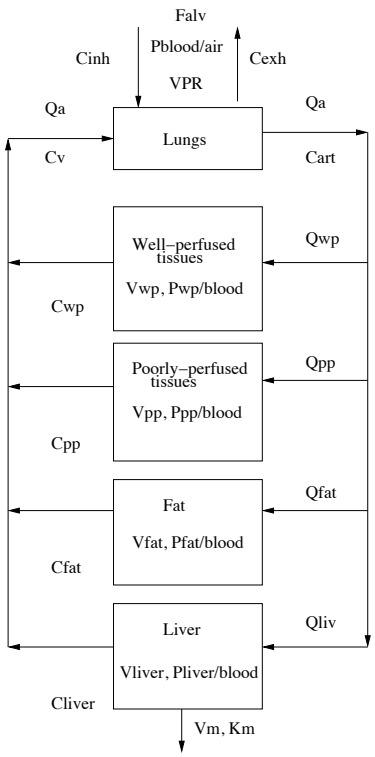

Figure 10.1: *A representative physiologically-based pharmacokinetic model.*

Intensive ***toxicokinetic studies*** are conducted in animal models and involve exposure of each animal to the agent and collection of frequent blood and other samples from which concentrations are ascertained. They are also conducted to a lesser extent in humans, from whom blood, breath, and urine samples are obtained and assayed for concentrations of the agent. Understandably, toxicokinetics is of great interest to ***environmental regulatory agencies*** such as the EPA.

Here, interest focuses on learning about key processes, such as the rate at which the agent is ***metabolized*** in the liver. This information is used in the overall toxicological assessment. Because the scientific questions involve the such organ-specific processes, it is standard to entertain ***more detailed compartmental models*** that represent the body by ***physiologically identifiable compartments*** such as fatty tissues, poorly- and well-perfused tissues, the liver, and so on. Figure 10.1 shows a prototypical such model.

These so-called ***physiologically-based pharmacokinetic*** (PBPK) models generally are complex ***systems of differential equations*** that ***do not admit closed form solutions*** for observable concentrations, so must be solved numerically.

For example, equations corresponding to the model in Figure 10.1 are given by

$$C_{\text{art}} = \frac{F_{\text{card}} C_{\text{ven}} + F_{\text{alv}} C_{\text{inh}}}{F_{\text{card}} + F_{\text{alv}}/P_{\text{blood/air}}}, \quad C_{\text{ven}} = \sum_s \frac{F_s C_s}{F_{\text{card}}}$$

$$C_{\text{exh}} = (1-\delta)\frac{C_{\text{art}}}{P_{\text{blood/air}}} + \delta C_{\text{inh}}$$

$$\frac{dC_s}{dt} = \frac{F_s}{V_s}\left(C_{\text{art}} - \frac{C_s}{P_{s/\text{blood}}}\right), \quad s = \text{wp, pp, fat}$$

$$\frac{dC_{\text{liv}}}{dt} = \frac{F_{\text{liv}}}{V_{\text{liv}}}\left(C_{\text{art}} - \frac{C_{\text{liv}}}{P_{\text{liv/blood}}}\right) - R_{\text{liv}} \ (s = \text{liv}),$$

$$R_{\text{liv}} = \frac{V_{\max} C_{\text{liv}}}{V_{\text{liv}}(K_m + C_{\text{liv}})},$$

In the figure, the Qs are amounts, which are divided by corresponding compartmental ***volume*** parameters V to yield concentrations C. The F parameters are blood flow rates, the P parameters are tissue-over-blood ***partition coefficients***, and $V_{\max}$ and K_m are ***Michaelis-Menten metabolism coefficients*** that describe ***metabolism in the liver***.

Interest thus focuses on these metabolism parameters, their ***typical values*** and how they ***vary in the population***. Thus, the nonlinear hierarchical model we have discussed is a natural framework for addressing this. However, there are some serious ***challenges***.

- The solution to the above system for the ***observable concentration*** (usually in blood or exhaled breath) provides the model f of interest, and it is clear that a closed form expression is extremely unlikely.
- In general, such PBPK models involve ***numerous parameters***, most of which are ***not identifiable*** from longitudinal concentration-time data from these studies.

A naive approach to inference on the ***typical values*** of $V_{\max}$ and K_m has been to ***fix*** all parameters except these at "***literature values***" for all individuals, estimate $V_{\max}$ and K_m for each individual, and base inference on the typical values and variability on these estimates. Clearly, this crude approximation is ***highly suspect***.

Gelman, Bois, and Jiang (1996) were the first to propose formally placing this problem within the ***nonlinear hierarchical framework***. To address the challenge of handling the numerous unidentifiable parameters in a more ***principled*** way, they proposed a ***Bayesian formulation*** as in the previous section, where the hyperprior specifications for each parameter are based on ***historical and literature information***.

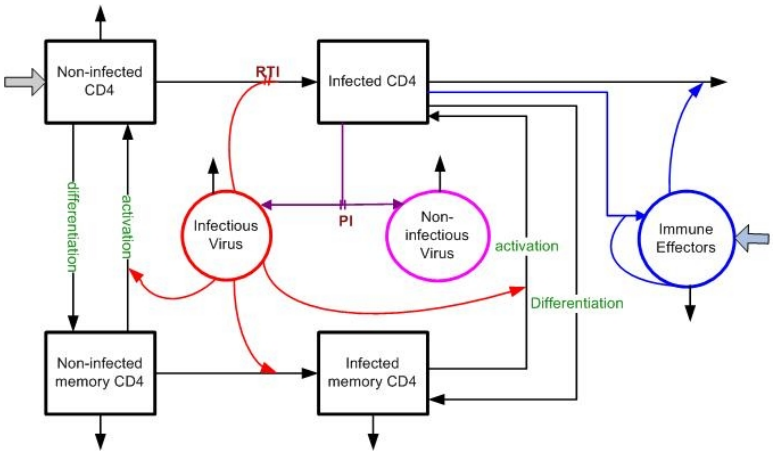

Figure 10.2: *A representative HIV dynamic model.*

These priors are "***informative***" in that they do not reflect ***vague knowledge*** but rather incorporate ***scientific information and judgment***. Implementation of the model is via MCMC techniques.

A software package, ***MCSim***, that implements this general approach, is available. Of course, its appropriate use requires a fairly deep understanding of both the theoretical model and the statistical model in which it is embedded.

Interestingly, the naive approach is still often used.

HIV DYNAMICS: Human immunodeficiency virus Type-1 (HIV) progressively destroys the body's ability to fight infection by killing or damaging cells in the immune system. Since the mid-1990s, there has been considerable interest in developing mathematical models to represent hypothesized mechanisms governing the ***interplay between HIV and the immune system***. These so-called ***HIV dynamic models*** have led to advances in understanding of plausible mechanisms underlying HIV pathogenesis and in developing ***antiretroviral*** treatment strategies for HIV-infected individuals.

As in pharmacokinetics, these models are predicated on representing processes involved in the virus-immune system ***interplay*** via ***hypothetical compartments***, where the compartments characterize different populations of virus; immune system cells targeted by the virus, namely $CD4^+$ T cells; and so on, that interact within a subject. An example is the model in Figure 10.2, which shows a typical such model.

As an example, we show a model studied by Adams et al. (2007), which involves ***compartments*** denoted by T_1, type 1 target cells, e.g., CD4$^+$ cells (cells/μl); T_2, type 2 target cells, such as macrophages (cells/μl); V_I and V_{NI}, infectious and noninfectious free virus, respectively (RNA copies/ml); and E, cytotoxic T-lymphocytes (cells/μl). With a superscript asterisk (*) denoting infected target cells; and, e.g., with $T_1(t)$ = concentration of type 1 target cells at time t, the model is

$$\begin{aligned}
\dot{T}_1 &= \lambda_1 - d_1 T_1 - \{1 - \epsilon_1 u(t)\} k_1 V_I T_1, \\
\dot{T}_2 &= \lambda_2 - d_2 T_2 - \{1 - f\epsilon_1 u(t)\} k_2 V_I T_2, \\
\dot{T}_1^* &= \{1 - \bar{\epsilon}_1(t)\} k_1 V_I T_1 - \delta T_1^* - m_2 E T_1^*, \\
\dot{T}_2^* &= \{1 - f\epsilon_1 u(t)\} k_2 V_I T_2 - \delta T_2^* - m_2 E T_2^*, \\
\dot{V}_I &= \{1 - \epsilon_2 u(t)\} 10^3 N_T \delta (T_1^* + T_2^*) - c V_I - \{1 - \epsilon_1 u(t)\} \rho_1 10^3 k_1 T_1 V_I \\
&\quad - \{1 - f\epsilon_1 u(t)\} \rho_2 10^3 k_2 T_2 V_I, \\
\dot{V}_{NI} &= \epsilon_2 u(t) 10^3 N_T \delta (T_1^* + T_2^*) - c V_{NI}, \\
\dot{E} &= \lambda_E + \frac{b_E (T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_b} E - \frac{d_E (T_1^* + T_2^*)}{(T_1^* + T_2^*) + K_d} E - \delta_E E,
\end{aligned} \tag{10.11}$$

along with an initial condition vector $\{T_1(0), T_2(0), T_1^*(0), T_2^*(0), V_I(0), V_{NI}(0), E(0)\}^T$. In (10.11), most dependence on t is suppressed for brevity, and the factors of 10^3 convert between μl and ml scales. The model depends on ***numerous meaningful parameters*** β that are thought to vary across subjects; e.g., c (1/day), the natural death rate of the virus; δ (1/day), the death rate of infected target cells; and λ_k, $k = 1, 2$, production rates (cells/μl-day) of type 1 and 2 target cells. The function $u(t)$, $0 \leq u(t) \leq 1$, represents ***time-dependent input*** of antiretroviral therapy, with $u(t) = 0$ corresponding to fully off treatment and $u(t) = 1$ to fully on treatment. The parameters ϵ_k, $k = 1, 2$, $0 \leq \epsilon_k \leq 1$, are ***efficacies*** of reverse transcriptase inhibitor treatments for blocking new infections and protease inhibitors for causing infected cells to produce noninfectious virus, respectively. See Adams et al. (2007) for a complete description.

In a typical study of HIV-infected subjects, longitudinal data on ***combinations*** of one or more of these compartments are collected, and interest focuses on ascertaining the typical values of some of these parameters and how they vary across subjects to gain insight into ***viral mechanisms*** and their possible associations with subject characteristics. Usually, longitudinal measurements of total CD4$^+$ cells, $T_1 + T_1^*$, and total viral load, $V_i + V_{NI}$, are available on each subject. Needless to say, solution of (10.11) can only be carried out numerically. A further complication is that total viral load measurements may be ***left-censored*** by the ***lower limit of quantification*** of the assay. Finally, as with PBPK models, the available data ***fail to identify*** all the parameters.

Clearly, the scientific questions can be addressed within the nonlinear mixed effects model framework. There is an ongoing literature on approaches to doing this, which included ***integrating numerical solution of the system of differential equations with estimation***. Not surprisingly, many of the approaches place the problem in a Bayesian formulation.

10.4 Time-dependent covariates in nonlinear mixed effects models

In Chapter 9, we did not address the situation in which ***among-individual covariates change valu***e over the course of observation of an individual. We now remark briefly on the implications of this in the nonlinear mixed effects model context.

GENERALIZATION: When time-dependent among-individual covariates are available, a first thought is to ***modify*** the basic two-stage hierarchy as follows. For this discussion, let $\boldsymbol{a}_{ij}$ denote the values of among-individual covariates at time j for individual i.

Stage 1 - Individual model. Given a model f as in (9.6), the random vectors $\boldsymbol{Y}_i$, $i = 1, \ldots, m$, are assumed to satisfy

$$E(\boldsymbol{Y}_i|\boldsymbol{z}_i, \beta_{ij}) = \boldsymbol{f}_i(\boldsymbol{z}_i, \beta_{ij}), \qquad \text{var}(\boldsymbol{Y}_i|\boldsymbol{z}_i, \beta_{ij}) = \boldsymbol{R}_i(\beta, \gamma, \boldsymbol{x}_i, \boldsymbol{b}_i). \tag{10.12}$$

In (10.12), we allow the individual-specific parameters β_{ij} to ***change with*** j.

Stage 2 - Population model. The individual-specific parameter β_{ij} satisfies

$$\beta_{ij} = \boldsymbol{d}(\boldsymbol{a}_{ij}, \beta, \boldsymbol{b}_i); \tag{10.13}$$

a ***linear*** version of (10.13) is

$$\beta_{ij} = \boldsymbol{A}_{ij}\beta + \boldsymbol{B}_{ij}\boldsymbol{b}_i,$$

where the design matrix $\boldsymbol{A}_{ij}$ changes with changing values $\boldsymbol{a}_{ij}$. Here and in (10.13), the value of β_{ij} thus changes as $\boldsymbol{a}_{ij}$ changes.

There is no ***operational barrier*** to fitting the model in (10.12)- (10.13); e.g., the R function `nlme()` has this capability, which is discussed in Pinheiro and Bates (2000, Section 7.1). Thus, implementing such a model is ***entirely possible***.

The issue is whether or not such a model is ***appropriate or even makes sense*** , as we now discuss.

GENERALIZED LINEAR MIXED EFFECTS MODELS: With ***generalized linear mixed effects models*** , the objective is ordinarily inference on the association between the response and among-individual covariates such as treatment assignment in a randomized clinical trial and/or subject characteristics from a ***subject-specific*** perspective. The model is thus an ***empirical framework*** in which to address these questions. When these covariates are ***time-independent*** , there is ***no conceptual difficulty*** with specifying these models and making the desired inferences, as discussed in Section 8.6 in the case of population-averaged models.

When ***time-dependent*** among-individual covariates are involved, however, the ***same conceptual issues*** discussed in Section 8.6 arise. Modeling involving ***endogenous covariates*** suffers from the ***same difficulties of interpretation*** discussed in that section. Namely, ***time-dependent confounding*** complicates and frankly renders impossible the ability to draw ***causal inferences*** based on these models.

Thus, adopting the model (10.12)-(10.13) when such time-dependent covariates are involved is almost certainly a ***prescription for misleading inference*** and challenging interpretation. As noted in Section 8.6, an entirely ***different approach*** is required.

NONLINEAR MIXED EFFECTS MODELS: In the case of a nonlinear hierarchical model in which the function f representing the within-individual mean trajectory is derived from ***mechanistic***,, ***theoretical considerations***, (10.12)- (10.13) is problematic for ***different reasons***. Such theoretical models are derived under the fundamental assumption that that the ***scientifically meaningful parameters*** involved are ***constants with respect to time***.

For example, in simple compartmental models of the type we have discussed in previous chapters, PK parameters like clearance Cl_i and volume of distribution V_i of necessity are ***fixed*** for individual i, and the ***differential equations*** giving rise to the concentration-time model are predicated on this. Similarly, in the HIV dynamic model (10.11), the numerous parameters are regarded as ***fixed constants*** for each individual. In each case, a practical interpretation is that these fixed parameters are thought to be ***inherent characteristics*** of the individual that govern his/her kinetics or dynamics.

In such systems, if a ***parameter*** is thought to ***vary with time*** , then it is taken in the model to be a ***function of time*** , which might be represented for fitting purposes as a ***parametric or nonparametric*** model. Allowing a mechanistic parameter to vary with time ***fundamentally alters*** the nature of the ***system of differential equations*** and thus the form of the solution.

time (hours)	conc. (mg/L)	dose (mg)	age (years)	weight (kg)	creat. (ml/min)	glyco. (mg/dl)
0.00	–	166	75	108	> 50	69
6.00	–	166	75	108	> 50	69
11.00	–	166	75	108	> 50	69
17.00	–	166	75	108	> 50	69
23.00	–	166	75	108	> 50	69
27.67	0.7	–	75	108	> 50	69
29.00	–	166	75	108	> 50	94
35.00	–	166	75	108	> 50	94
41.00	–	166	75	108	> 50	94
47.00	–	166	75	108	> 50	94
53.00	–	166	75	108	> 50	94
65.00	–	166	75	108	> 50	94
71.00	–	166	75	108	> 50	94
77.00	0.4	–	75	108	> 50	94
161.00	–	166	75	108	> 50	88
168.75	0.6	–	75	108	> 50	88
height=72 inches, Caucasian, smoker, no ethanol abuse, no CHF						

Table 10.1: *Data for a subject in the quinidine study. conc. = quinidine concentration, glyco. =* α_1*-acid glycoprotein concentration, CHF = congestive heart failure.*

Thus, for example, in a one-compartment model, allowing clearance, to vary with the value of an among-individual covariate and writing Cl_{ij}, say, must be carefully considered for ***plausibility*** and can have implications for the ***validity*** of the model used.

In pharmacokinetics, under certain conditions, it is accepted that individual-specific parameters ***can fluctuate*** over observation periods but ***not over*** each observation time *j*. To discuss this, we consider a ***world-famous study*** of the pharmacokinetics of the anti-arrhythmic drug quinidine, which has been cited by numerous authors (e.g., Davidian and Giltinan, 1995, Sections 1.1.2 and 9.3; Wakefield, 1996; Pinheiro and Bates, 2000, Sections 3.4 and 8.2).

The $m = 136$ subjects in the study were hospitalized and undergoing routine treatment with oral quinidine for atrial fibrillation or ventricular arrhythmia. Table 10.1 shows ***abridged data*** for one subject, which typify the information collected. Demographical and physiological characteristics included age; weight; height; ethnicity/race (Caucasian/Black/Hispanic); smoking status (yes/no); ethanol abuse (no/current/previous); congestive heart failure (no or mild/moderate/severe); creatinine clearance, a measure of renal function ($\leq$ 50 ml/min indicates renal impairment); and α_1-acid glycoprotein concentration, the level of a molecule that binds quinidine. In addition, ***dosing history*** (times, amounts) was recorded along with quinidine concentrations.

A ***one-compartment model*** with first-order absorption and elimination has been used to describe the PK of quinidine. With repeated dosing, in addition to the ***principle of superposition*** discussed in Section 9.2, it is also assumed that drug accumulates in the system until a "***steady state***" is reached at which, roughly, rate of administration of drug is ***equal to*** the rate of elimination (e.g., Giltinan, 2014).

Under these conditions, the model can be written as follows. Denoting the ℓth (dose time, amount) by (s_ℓ, D_ℓ) as before, the amount of quinidine in the "***absorption depot***," $A_a(s_\ell)$, and the concentration of quinidine in the blood, $C(s_\ell)$, at dose time s_ℓ for a subject who has ***not yet achieved a steady state*** are given by

$$\begin{aligned} A_a(s_\ell) &= A_a(s_{\ell-1})\exp\{-k_a(s_\ell - s_{\ell-1})\} + D_\ell, \\ C(s_\ell) &= C(s_{\ell-1})\exp\{-k_e(s_\ell - s_{\ell-1})\} + A_a(s_{\ell-1})\frac{k_a}{V(k_a - k_e)} \\ &\quad \times\Big[\exp\{-k_e(s_\ell - s_{\ell-1})\} - \exp\{-k_a(s_\ell - s_{\ell-1})\}\Big], \end{aligned}$$

and the concentration of quinidine at time t in the next dosing interval $(s_\ell, s_{\ell+1})$ is

$$\begin{aligned} C(t) &= C(s_\ell)\exp\{-k_e(t - s_\ell)\} + A_a(s_\ell)\frac{k_a}{V(k_a - k_e)} \\ &\quad \times\Big[\exp\{-k_e(t - s_\ell)\} - \exp\{-k_a(t - s_\ell)\}\Big], \quad s_\ell < t < s_{\ell+1}, \end{aligned} \tag{10.14}$$

where $k_e = Cl/V$. Once a ***steady state*** has been reached, a ***further*** set of equations governs the values of $A_a(s_\ell)$ and $C(s_\ell)$ at dose times, which we exclude here for brevity. The model for concentration at time t that dictates f thus depends on the meaningful parameters $\beta = (k_a, V, Cl)^T$.

The quinidine study is representative of the situation where one or more subject characteristics thought to be ***associated*** with pharmacokinetic behavior ***change*** over the observation period on the subject, as is the case here for α_1-acid glycoprotein concentration. For the subject in Table 10.1, it is likely that α_1-acid glycoprotein concentration was ***measured intermittently*** at times 0, 29, and 161. In this situation, a standard modeling approach is based on the following idea.

If a subject is observed over several treatment intervals, it may be reasonable to expect that, although a basic compartment model with ***static parameters*** applies ***in any interval***, ***fluctuations*** in the values of his/her pharmacokinetic parameters may occur over time that ***show an association*** with ***other characteristics*** that also change. From this point of view, for the quinidine study, the assumption is that the pharmacokinetic parameters in (10.14) for the individual in Table 10.1 are ***constant within the intervals*** 0–29 hours, 29–77 hours, and after 161 hours, but may have ***fluctuated*** over the entire period in a way that is associated with α_1-acid glycoprotein concentration.

Such an assumption is clearly subject to ***scientific debate*** but is often invoked as a practical way to view the problem.

Denoting these intervals by I_k, $k = 1, \ldots, a$ ($a = 3$), a standard modeling approach in the pharmacokinetic literature is to modify the stage 2 population model (10.13) as

$$\beta_{ij} = \boldsymbol{d}(\boldsymbol{a}_{ik}, \beta, \boldsymbol{b}_i), \tag{10.15}$$

where $\boldsymbol{a}_{ik}$ is the value of the subject characteristics for $t_{ij} \in I_k$. In (10.15), the element of β that is the coefficient of α_1-acid glycoprotein concentration is taken to be ***constant*** over all intervals; e.g., for $t_{ij} \in I_k$,

$$\log Cl_{ij} = \beta_0 + \beta_1 a_{ik} + b_i. \tag{10.16}$$

The model (10.15) implies that, within a given individual, "***inter-interval variation***" is ***entirely*** "***explained***," by the change in covariates for that individual. The model can be extended to include ***nested random effects*** that allow for ***unexplained biological variation*** within intervals, as discussed in the next section.

10.5 Multilevel models

Longitudinal data with the structure we have focused on in this course can be viewed as having a ***single level*** of ***clustering***. In particular, responses fall into natural ***clusters*** because they are ascertained longitudinally on ***different individuals***. Responses from the same cluster are naturally viewed as ***correlated*** due to the fact that they are "***more alike***" by virtue of belonging to the same cluster, the phenomenon we have referred to as the ***among-individual source of correlation***. ***Linear and nonlinear mixed effects models*** naturally account for this correlation through the introduction of ***random effects***.

From this point of view, the longitudinal data structure we have focused on in this course is a ***special case*** of a ***general clustered data structure***, in which it is possible to identify ***multiple levels*** of such clustering. In this section, we give a brief introduction to statistical models for this data structure. The models we have discussed are thus a particular case of these models.

MULTIPLE LEVELS OF CLUSTERING: In many settings, the data structure is such that it is possible to identify ***multiple levels of clustering***:

- A classic example is an ***agricultural study*** in which ***plots*** are ***nested*** within experimental ***blocks***, and plots are randomized to treatments within blocks. A single response may be ascertained on each plot at the end of the growing season. Alternatively, in a more complicated version, longitudinal responses are collected on each plot throughout the growing season.

- Similarly, a common situation in ***clinical trials*** is that in which subjects are recruited from a sample of ***clinics***. Each subject within each clinic is randomized to a study treatment. Here, subjects are ***nested*** within clinics. The response or interest may be ascertained on each subject at the end of the study period. Alternatively, each subject may be followed ***longitudinally*** for the response at several clinic visits over the study period.

- This structure occurs frequently in studies in ***public health*** and ***education*** where it is ***not feasible*** to expose participants to treatment or conditions of interest ***individually***. For example, in a study of school-based interventions to prevent smoking, entire ***schools*** might be randomized to receive a particular intervention program; allowing different students within same school to receive different interventions opens the possibility that they could discuss and compare them, which could ***compromise*** the ability to assess their effects.

  Individual instructors might actually ***deliver the interventions*** at the ***classroom level***, so this represents a ***source of variation***, in that particular instructors might be more or less effective at delivering the same intervention. Smoking behavior might then be recorded over time on individual students. Here, students are ***nested*** within classrooms within schools.

  The response, ***score*** on a questionnaire assessing tobacco and health knowledge, might be ascertained at ***baseline***, prior to intervention, and then again at the end of the intervention on each student.

COVARIATES: In the foregoing examples, it is natural that there might be ***covariates*** that recorded at ***different levels*** of the hierarchy.

- In the smoking intervention study, there may be covariates collected at the level of the ***individual student***, such as gender, GPA, socioeconomic status, and so on.

Covariates at the ***classroom level*** , such as those recorded on the ***instructor*** , including gender, previous experience with such interventions, years teaching, etc, might also be collected. Finally, ***school-level*** covariates, such as location, racial/ethnic makeup, proportion of students receiving free/reduced price lunch, etc, might be available.

- Questions of ***scientific interest*** may involve associations between response or level-specific behavior and these covariates. In particular, the goal is often to determine the relative importance of characteristics at ***different levels*** in terms of effects on the response.

MULTILEVEL HIERARCHICAL MODELS: A natural framework in which to place such questions is that of ***multilevel hierarchical models***. An important feature of these models is that they take account of ***correlation*** due to ***clustering*** at each stage of the hierarchy, noted above, as we demonstrate shortly.

The linear and nonlinear mixed effects models we have discussed for longitudinal data are ***special cases*** of this general framework. In the more general model, there ***may or may not*** be a longitudinal aspect. If there is, it corresponds to a ***level*** in the hierarchy.

For simplicity, we consider ***linear models*** , and comment on the obvious generalization to nonlinear and generalized linear models at the end of this section.

LEVELS OF THE HIERARCHY: It is common to identify ***units*** , be they individual subjects or something else, at each ***level*** of the hierarchy. ***Level 1*** units are at the '***lowest*** " level and are the units on which the response is ascertained. ***Level 2*** units are clusters composed of level 1 units. ***Level 3*** units comprise clusters of level 2 units, and so on.

This is most easily appreciated through examples.

- Consider the clinical trial example above, suppose that a single measure of the response is recorded at the end of the study. In this case, level 1 units are the ***individual subjects*** on whom the response measure will obtained. Level 2 units are the ***clinics*** , in which subjects are ***nested***. Thus, this is a two-level structure.
- Consider the same example, but where now each subject will visit the clinic at several occasions during the study period, and the response will be ascertained at each visit. Here, level 1 units are the ***measurement occasions*** , i.e., the longitudinal times at which the response is ascertained. These are ***nested*** within level 2 units, the ***subjects***. Level 3 units are the clinics in which level 2 units, subjects, are ***nested***.

- In the smoking intervention example, if the baseline score is treated as a ***covariate*** , it is natural to view this as a hierarchy with ***three levels***. Level 1 units are the individual students, nested within level 2 units, classrooms, nested within level 3 units, schools. If we view the baseline and final scores as ***longitudinal responses*** instead, level 1 is the measurement occasion. level 2 is the student, level 3 is classroom, and level 4 is school.

- In the longitudinal data structure which we have been concerned in this course, it follows that level 1 units are the ***longitudinal occasions*** at which the response is ascertained, and level 2 units are the ***individuals***.

- This convention obviously can be applied in general.

NOTATIONAL CONVENTION: In the literature, it is ***conventional*** (although there are exceptions) to use i to index level 1 units, j to index level 2 units, and k to index level 3 units in a ***three-level model***.

The ***numbers of units*** are identified as follows. There are n_3 level 3 units, indexed by $k = 1, \ldots, n_3$. Within the kth, there are n_{2k} level 2 units indexed by $j = 1, \ldots, n_{2k}$. Within the jth level 2 unit, there are n_{1jk} level 1 units indexed by $i = 1, \ldots, n_{1jk}$.

In a ***two-level model*** , there are n_2 level 2 units indexed by $j = 1, \ldots, n_2$ and within the jth level 2 unit there are n_{1j} level 1 units indexed by $i = 1, \ldots, n_{1j}$.

- Thus, under this indexing convention, the ***longitudinal data structure*** we have considered would use i to index time points and j to index individuals, exactly ***opposite*** of the standard indexing convention in the longitudinal data literature, which we have used to this point.

  Here, the ***total number of individuals*** m in the notation we have used corresponds to n_2 and the number of observations on each individual, which we denote by n_i for individuals indexed by i, is n_{1j} for the jth individual in this indexing scheme.

- It is ***prudent*** when reading the multilevel modeling literature to pay attention to the indexing convention used.

In the following example, we demonstrate this indexing convention.

MULTICENTER LONGITUDINAL CLINICAL TRIAL: As above, suppose a ***clinical trial*** is conducted comparing two treatments, coded as 0 and 1. The trial involves n_3 clinics (level 3 units) indexed by $k = 1, \ldots, n_3$. Within each clinic k, n_{2k} subjects are recruited, indexed by $j = 1, \ldots, n_{2k}$ (level 2 units), each of whom is randomized to receive treatment 0 or 1. On subject j within clinic k, n_{1jk} responses are ascertained at several clinic visits, indexed by $i = 1 \ldots, n_{1jk}$, where $i = 1$ corresponds to ***baseline***.

The usual perspective is that the clinics represent a ***random sample*** from the ***hypothetical population*** of all possible such clinics in which subjects could be recruited and given the study agents. The subjects within clinics likewise are viewed as ***random samples*** from the hypothetical populations of all possible subjects who could attend each clinic.

Suppose that interest focuses on comparing the patterns of change in the response over the study period for the two treatments, where, for ***any subject*** , the expected longitudinal trajectory over time is expected to show a ***constant rate of change***. From this perspective, it is natural to take a ***subject-specific perspective*** to developing a model, as follows.

Let Y_{ijk} represent the ***response*** on subject j from clinic k at the ith time at which the subject is observed, $i = 1, \ldots, n_{1jk}$. Under the principles we used to develop the ***two-stage linear mixed effect model hierarchy*** in Chapter 6, the following formulation is natural.

Letting t_{ijk} denote the ***observation times*** for subject j in clinic k, represent the responses at the subject level as

$$Y_{ijk} = \beta_{0jk} + \beta_{1jk} t_{ijk} + e_{ijk}, \quad i = 1, \ldots, n_{1jk}, \tag{10.17}$$

where, in (10.17), β_{0jk} and β_{1jk} are the ***subject-specific*** intercept and slope dictating subject (j, k)'s ***inherent longitudinal trajectory***; and e_{ijk} is a ***mean zero*** (conditional on covariates) ***within-subject*** deviation representing the effects of the realization process and measurement error.

Defining $\boldsymbol{Y}_{jk}$ and $\boldsymbol{e}_{jk}$ $(n_{1jk} \times 1)$ in the obvious way, we can write (10.17) succinctly as

$$\boldsymbol{Y}_{jk} = \boldsymbol{C}_{jk}\boldsymbol{\beta}_{jk} + \boldsymbol{e}_{jk}, \quad \boldsymbol{\beta}_{jk} = (\beta_{0jk}, \beta_{1jk})^T. \tag{10.18}$$

Assumptions on $\boldsymbol{e}_{jk}$ would be made as discussed in Chapters 2, 6, 7, and 9.

The next step in formulating the multilevel model is to represent the ***subject-specific intercept and slope*** in terms of ***fixed effects*** , covariates, and ***random effects***.

Given that this is a ***randomized study*** , so that we do not expect baseline response to be associated with treatment assignment, write

$$\beta_{0jk} = \beta_0 + b_{0k} + b_{0jk}, \tag{10.19}$$

where $E(b_{0k}) = 0$, $E(b_{0jk}) = 0$, so that $E(\beta_{0jk}) = \beta_0$, the mean or "typical" value of intercept across all clinics and subjects within them. In (10.19), b_{0k} is a ***random effect*** representing how ***subject-specific intercepts*** for subjects in the kth clinic deviate from the overall mean intercept β_0,, and b_{0jk} is a ***random effect*** representing further how the intercept for the jth subject within that clinic deviates from the ***clinic-specific*** (conditional on clinic) mean intercept (across subjects in the clinic) $\beta_0 + b_{0k}$.

Similarly, a model for the ***subject-specific*** slope is

$$\beta_{1jk} = \beta_1 + \beta_2\delta_{jk} + b_{1k} + b_{1jk}, \tag{10.20}$$

where $\delta_{jk} = 0$ if the jth subject in clinic k received drug 0, and $\delta_{jk} = 1$ if s/he received drug 1; and $E(b_{1k}|\delta_{jk}) = 0$, $E(b_{1jk}|\delta_{jk}) = 0$. Thus, in (10.20), β_1 is the mean or "typical" slope ***across all clinics and subjects*** for subjects receiving drug 0, and $\beta_1 + \beta_2$ is that for drug 1, so that β_2 represents the ***difference in mean or typical rate of change*** between the two drugs.

As for the intercept, b_{1k} is a ***random effect*** representing how subject-specific slopes for subjects in the kth clinic deviate from the overall mean slope for each treatment, and b_{1jk} is a ***random effect*** representing further how the slope for the jth subject within that clinic deviates from the ***clinic-specific*** (conditional on clinic) mean slope $\beta_1 + \beta_2\delta_{jk} + b_{1k}$, depending on the treatment assigned.

As in Chapter 6, we can ***summarize*** (10.19)-(10.20) as

$$\boldsymbol{\beta}_{jk} = \boldsymbol{A}_{jk}\boldsymbol{\beta} + \boldsymbol{B}_k^{(1)}\boldsymbol{b}_k + \boldsymbol{B}_{jk}^{(2)}\boldsymbol{b}_{jk}, \tag{10.21}$$

where

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T, \quad \boldsymbol{b}_{jk} = (b_{0jk}, b_{1jk})^T, \quad \boldsymbol{b}_k = (b_{0k}, b_{1k})^T,$$

$$\boldsymbol{A}_{jk} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \delta_{jk} \end{pmatrix}, \quad \boldsymbol{B}_k^{(1)} = \boldsymbol{I}_2, \quad \boldsymbol{B}_{jk}^{(2)} = \boldsymbol{I}_2.$$

It should be clear that it is possible to modify (10.19) and (10.20) to incorporate dependence on both ***clinic-level*** and ***subject-level covariates***. For example, suppose that w_{jk} is the weight of subject j in clinic k, and s_k is the average number of suffers of the condition the drugs are supposed to treat seen at clinic k over the past five years.

A model that allows subject-specific baseline response to depend on these covariates is

$$\beta_{0jk} = \beta_{00} + \beta_{01} s_k + \beta_{02} w_{jk} + b_{0k} + b_{0jk}. \tag{10.22}$$

In (10.22), inference on β_{01} addresses the ***clinic-level*** issue of whether or not the average number of patients seen is associated with subject-specific baseline response; e.g., do patients with ***worse*** baseline response seek out clinics that treat ***larger numbers*** of patients with this condition? Likewise, β_{02} addresses the ***subject-level*** issue of whether or not weight is associated with baseline response.

The slope specification (10.20) can of course be modified analogously to incorporate dependence on such clinic- and subject-level characteristics. Under (10.22), the definitions of $\boldsymbol{A}_{jk}$ and β above would be revised in the ***obvious way***.

The model is completed by ***assumptions*** on the ***random effects*** $\boldsymbol{b}_k$ and $\boldsymbol{b}_{jk}$. The random effects corresponding to different levels of the model (clinics and subjects within clinics here) are ordinarily assumed to be ***independent*** of one another. Under the assumption that the $\boldsymbol{b}_k$ and $\boldsymbol{b}_{jk}$ are both ***iid*** , so that their distributions do not depend on covariates, the usual assumption is

$$\boldsymbol{b}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}^{(1)}), \qquad \boldsymbol{b}_{jk} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}^{(2)}), \tag{10.23}$$

where, in (10.23), $\boldsymbol{D}^{(1)}$ is a covariance matrix corresponding to the ***level 1*** (clinic) random effect and $\boldsymbol{D}^{(2)}$ is a covariance matrix corresponding to the ***level 2*** (subject) random effect.

- The ***elements*** of $\boldsymbol{D}^{(1)}$ thus characterize variation in intercepts and slope and covariation between them due to ***clinic-level*** phenomena, and those of $\boldsymbol{D}^{(2)}$ thus represent the additional variation and correlation due to ***subject-level*** sources.

GENERAL THREE-LEVEL MODEL: In general, substituting (10.21) in (10.18) yields, analogous to (6.44), a general model of the form

$$\boldsymbol{Y}_{jk} = \boldsymbol{X}_{jk}\beta + \boldsymbol{Z}^{(1)}_{jk}\boldsymbol{b}_k + \boldsymbol{Z}^{(2)}_{jk}\boldsymbol{b}_{jk} + \boldsymbol{e}_{jk}, \tag{10.24}$$

where, letting β be $(p \times 1)$, $\boldsymbol{b}_k$ be $(q_1 \times 1)$, and $\boldsymbol{b}_{jk}$ be $(q_2 \times 1)$,

$$\boldsymbol{X}_{jk} = \boldsymbol{C}_{jk}\boldsymbol{A}_{jk} \ (n_{1jk} \times p), \quad \boldsymbol{Z}^{(1)}_{jk} = \boldsymbol{C}_{jk}\boldsymbol{B}^{(1)}_{k} \ (n_{1jk} \times q_1), \quad \boldsymbol{Z}^{(2)}_{jk} = \boldsymbol{C}_{jk}\boldsymbol{B}^{(2)}_{jk} \ (n_{1jk} \times q_2);$$

and, as in (10.23), the usual assumption is that

$$\boldsymbol{b}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}^{(1)}), \qquad \boldsymbol{b}_{jk} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{D}^{(2)}), \qquad \boldsymbol{e}_{jk} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n_{1jk}}). \tag{10.25}$$

As we have discussed, more general specifications would ***not*** take the random effects and $\boldsymbol{e}_{jk}$ to be ***independent*** of covariates $\boldsymbol{x}_{jk}$, where $\boldsymbol{x}_{jk}$ is the collection of all covariates on unit (j, k) at all levels; would allow ***serial correlation*** among elements of $\boldsymbol{e}_{jk}$ when level 1 units are longitudinal measurement occasions; and would allow the variance of elements of $\boldsymbol{e}_{jk}$ to be ***nonconstant*** and to depend on the random effects.

- Clearly, in principle, the general hierarchical formulation shown here for three levels can be ***extended*** to ***any number*** of levels.
- Likewise, the ***same considerations*** can be invoked to specify ***multilevel generalized linear mixed effects models*** and ***multilevel nonlinear mixed effects models***.

INFERENCE: Given the specification in (10.24)-(10.25), it is possible to write down the corresponding ***normal loglikelihood*** and to ***maximize*** it in β, σ^2, and the distinct elements of $\boldsymbol{D}^{(1)}$ and $\boldsymbol{D}^{(2)}$.

- This is routine in software such as SAS `proc mixed` and `proc glimmix` and in the R packages `nlme` and `lme4`.
- It is also possible to derive and obtain ***empirical Bayes estimates*** of ***posterior modes*** for $\boldsymbol{b}_k$ and $\boldsymbol{b}_{jk}$ and to use these to obtain empirical Bayes estimates of ***subject-specific parameters*** such as β_{jk}.

A practical introduction to multilevel models is given in Chapter 22 of Fitzmaurice, Laird, and Ware (2011). More generally, there is an enormous literature on these models, particularly in the literature on social and behavioral science.

We conclude this section by noting that, in the context of ***pharmacokinetics*** with ***time-dependent among-individual covariates*** as in (10.15), the same considerations can be invoked to specify models for individual-specific PK parameters that are taken to "***fluctuate***" over time intervals where such covariates change value. Reverting to the standard longitudinal data indexing, for example, the model (10.16) for clearance corresponding to $t_{ij} \in I_k$ could be modified as

$$\log Cl_{ij} = \beta_0 + \beta_1 a_{ik} + b_i + b_{ik}$$

to allow this fluctuation to depend on biological variation as well as systematic association with the changing covariate.

10.6 Distribution of random effects

We now return to the two-stage, subject-specific hierarchical linear and nonlinear mixed effects models for the standard longitudinal data structure discussed in Chapters 6 and 9. A standard, default assumption in these models, and one that is embedded in standard software, is that the ***random effects*** $\boldsymbol{b}_i$ are ***normally distributed***. In Chapter 9, we discussed in the context of pharmacokinetics how the individual-level model can be ***reparameterized*** so that individual-specific PK parameters can be taken to have ***skewed distributions*** in the population; however, this formulation is still predicated on normal random effects.

WHY SHOULD RANDOM EFFECTS BE NORMALLY DISTRIBUTED? In general, there is ***no fundamental principle*** that requires that random effects implicated in the distributions of ***individual-specific*** parameters need be normally distributed or, for that matter, have ***unimodal distributions***.

- It may well be that, in the population, individual-specific features such as intercepts, slopes, or PK parameters have ***heavy-tailed*** distributions relative to the normal, so that the extent of spread in the population of these values is greater than that dictated by a normal distribution.
- It may even be possible for such features to have ***multimodal*** distributions. For example, it could be that there are underlying ***subpopulations*** for which values of such features tend to "cluster" around different mean or "typical" values.
- Such ***subpopulations*** might correspond to a particular ***among-individual characteristic***. For example, suppose that in a PK setting values of drug clearance tend to be ***different*** for smokers and nonsmokers. If log clearance values cluster as above according to smoking status, the overall distribution of clearance values, if we could see it, would appear ***bimodal***.

  It could be that, ***within*** each of these subpopulations, the distribution of log clearance ***is approximately normal***. The result is that the overall distribution of log clearance is thus ***a mixture of normal*** distributions. Normal mixtures can be ***bimodal*** if the means each normal component of the mixture are sufficiently far apart.
- This suggests that, in implementation of these models, failure to include such a ***among-individual covariate*** in the population model could lead to evidence of ***nonnormality*** of random effects, perhaps through plots of empirical Bayes estimates.

RELAXING THE NORMALITY ASSUMPTION: There is a large body of work on models and methods for linear and nonlinear mixed effects models that allow the distribution of the random effects to be ***nonnormal***. In these approaches, under assumptions about the ***true random effects distribution***, the distribution itself is ***estimated*** from the data along with other model parameters.

- In one class of models and methods, ***no restrictions*** are placed on the nature of the random effects distribution. That is, the distribution is assumed to lie in the class of ***all probability distributions***; this includes ***discrete*** distributions and distributions that are a mixture of ***continuous*** and discrete components.

  In this case, the ***estimator*** for the random effects distribution is ***fully nonparametric*** and itself a ***discrete distribution***.

- When the random effects correspond to individual-specific features such as drug clearance or slope of an individual-specific trajectory that are naturally regarded as ***continuous***, distributions that are discrete do not seem ***realistic*** or ***plausible models*** for the nature of the random effects. The models and methods above thus include as possibilities distributions that are highly unlikely to be plausible representations of the true density.

  This has led to models and methods that take as a starting point the assumption that the distribution of the random effects has a ***density*** that obeys certain ***smoothness restrictions***; for example, that is has a certain number of continuous derivatives. The assumption that the random effects distribution has such a ***smooth density*** of necessity restricts the class of plausible distributions, but the tradeoffs are that the restricted class is highly likely to contain the true density and that the resulting estimate will itself lead to a smooth distribution.

In the context of mixed effects models, there have been many proposals along both of these lines; see Davidian and Giltinan (1995, Chapter 7), Verbeke and Molenberghs (2000, Chapter 12), and and Zhang and Davidian (2001) for review of these approaches. In a Bayesian formulation of mixed effects model, similar developments have been proposed in which the random effects distribution F_b is taken to be unrestricted or to lie within some restricted class; see, for example, Müller and Rosner (1997).

Here, we briefly describe two popular approaches to representing the distribution of the random effects when it is assumed to have a ***smooth density***.

MIXTURE OF NORMAL DENSITIES: A number of authors have proposed to represent the distribution of random effects by a ***mixture of normal*** distributions; see the above references. A version of this has been called the ***heterogeneity model*** by Verbeke and Molenberghs (2000, Chapter 12). Here, under the assumption of ***iid random effects*** $\boldsymbol{b}_i$ ($q \times 1$), the usual normality assumption is replaced by

$$\boldsymbol{b}_i \sim \sum_{k=1}^{K} p_k \mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{D}_k), \qquad \sum_{k=1}^{K} p_k = 1, \tag{10.26}$$

where the contraint

$$\sum_{k=1}^{K} p_k \boldsymbol{\mu}_k = \boldsymbol{0} \tag{10.27}$$

is imposed to ensure that $E(\boldsymbol{b}_i) = \boldsymbol{0}$.

The representation (10.26) is standard and can be interpreted as saying that the population is a ***combination*** of K ***subpopulations***, where the kth population is a fraction p_k of the overall populations. Under this model, we can view each individual i as being a member of one of the underlying subpopulations.

It is straightforward to show, defining $u_{ik} = 1$ if the random effect $\boldsymbol{b}_i$ for individual i is from the kth subpopulation and = 0 otherwise and using a conditioning argument (try it), that the ***overall covariance matrix*** var($\boldsymbol{b}_i$) is

$$\boldsymbol{D} = \sum_{k=1}^{K} p_k \boldsymbol{\mu}_k \boldsymbol{\mu}_k^T + \sum_{k=1}^{K} p_k \boldsymbol{D}_k. \tag{10.28}$$

IMPLEMENTATION: Under (10.26), it follows that, with an individual model involving within-individual covariance parameters γ and a population model involving fixed parameters β,

$$p(\boldsymbol{y}_i|\boldsymbol{x}_i; \beta, \gamma, \boldsymbol{D}) = \sum_{k=1}^{K} p_k \int p(\boldsymbol{Y}_i|\boldsymbol{x}_i, \boldsymbol{b}_i; \beta, \gamma)\, p_k(\boldsymbol{b}_i; \mu_k, \boldsymbol{D}_k)\, d\boldsymbol{b}_i, \tag{10.29}$$

where $p_k(\,\cdot\,; \mu_k, \boldsymbol{D}_k)$ is the q-variate normal density with mean μ_k and covariance matrix $\boldsymbol{D}_k$.; and $\boldsymbol{D}$ depends on all of these as in (10.28).

- In this formulation, the number of mixture components K can be regarded as a ***tuning parameter*** controlling the ***flexibility*** of the representation. The more components in a mixture, the more flexible it is for representing ***complicated true densities***; however, this is at the expense of ***more parameters*** that need to be estimated; namely, more p_k, $\boldsymbol{\mu}_k$, and $\boldsymbol{D}_k$.

- It is customary to treat K as ***known***, fit the model incorporating (10.26) subject to the constraints (10.27) by maximizing the loglikelihood corresponding to (10.29) for $i = 1, \ldots, m$. Here, $K = 0$ corresponds to the ***usual normality assumption***, and successively increasing K leads to richer parameterizations of the density. One thus fits the model with $K = 0$, and usually $K = 1, 2$, and 3.

- Testing for the number of components K in a mixture as in (10.26) is subject to ***boundary problems*** similar to those we discussed for variance parameters in Section 6.6. Thus, it is not possible to deduce K under the assumption that the mixture model (10.26) is ***correct*** via ***likelihood ratio tests*** comparing models with increasing K, for example.

- An ad hoc approach is to inspect ***information criteria*** and choose the value of K that optimizes a given criterion.

For given K, it is possible to maximize the loglikelihood corresponding to (10.29) via an ***EM algorithm***. See Verbeke and Molenberghs (2000, Chapter 12).

SEMINONPARAMETRIC (SNP) DENSITY REPRESENTATION: An ***alternative approach*** to representing the density of the random effects was proposed for use with the nonlinear mixed effects model by Davidian and Gallant (1993) and represented in an advantageous parameterization by Zhang and Davidian (2001). In this formulation, under the assumption of ***iid random effects*** $\boldsymbol{b}_i$ $(q \times 1)$, the random effects are first written as

$$\boldsymbol{b}_i = \boldsymbol{\mu} + \boldsymbol{S}\boldsymbol{U}_i \quad (q \times 1), \tag{10.30}$$

where $\boldsymbol{\mu}$ is a $(q \times 1)$ vector of parameters, $\boldsymbol{S}$ is a lower triangular matrix, and $\boldsymbol{U}_i$ is a $(q \times 1)$ random vector. If the random vector $\boldsymbol{U}_i$ in (10.30) is taken to be ***standard multivariate normal***, and $\boldsymbol{\mu} = \boldsymbol{0}$, then (10.30) reduces to the usual normality assumption with $\boldsymbol{D} = \boldsymbol{S}\boldsymbol{S}^T$.

Instead, $\boldsymbol{U}_i$ is taken to have a ***smooth density*** that falls in a class of such densities that can be represented by an ***infinite series expansion***. As proposed by Davidian and Gallant (1993), the idea is to approximate this density, and thus the density of $\boldsymbol{b}_i$ in (10.30), by a ***truncated series expansion***; this has been referred to as the ***seminonparametric*** (SNP) approximation.

That is, the density of $\boldsymbol{U}_i$ is represented as

$$h_K(\boldsymbol{u}, \boldsymbol{a}) = P_K^2(\boldsymbol{u}, \boldsymbol{a})\, \varphi(\boldsymbol{u}) = \left\{ \sum_{|\lambda| \le K} a_\lambda \boldsymbol{u}^\lambda \right\}^2 \varphi(\boldsymbol{u}), \tag{10.31}$$

where $\varphi(\boldsymbol{u})$ is the standard q-variate normal density; $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)^T$ is a vector of nonnegative integers; $\boldsymbol{u}^\lambda = u_1^{\lambda_1} \cdots u_q^{\lambda_q}$, the ***monomial*** of order $|\boldsymbol{\lambda}| = \sum_{\ell=1}^q \lambda_\ell$, and $P_K(\boldsymbol{u}, \boldsymbol{a})$ is thus a polynomial of order K with coefficients collected in the vector $\boldsymbol{a}$. For example, when $K = 2$, $q = 2$,

$$P_K(\boldsymbol{u}) = a_{00} + a_{10} u_1 + a_{01} u_2 + a_{20} u_1^2 + a_{11} u_1 u_2 + a_{02} u_2^2,$$

and $\boldsymbol{a} = (a_{00}, a_{10}, a_{01}, a_{20}, a_{11}, a_{02})^T$.

In the representation (10.31), as in the ***normal mixture*** representation (10.26), the degree of the polynomial K plays the role of a a ***tuning parameter*** controlling the ***flexibility*** of the representation. The higher the order of the polynomial, the more flexible it is for representing ***complicated true densities***; however, this is again at the expense of ***more parameters*** that need to be estimated. Vast experience shows that $K = 1$ or 2 is often sufficient to approximate complex shapes, including ***multimodality and skewness***.

For (10.31) to be a ***legitimate density***, the coefficients in the polynomial $P_K(\boldsymbol{u}, \boldsymbol{a})$ must be chosen so that

$$\int h_K(\boldsymbol{u}, \boldsymbol{a})\, d\boldsymbol{u} = 1. \tag{10.32}$$

For (10.32) to hold, then, it is necessary to ***impose a constraint*** on the coefficients in $P_K(\boldsymbol{u}, \boldsymbol{a})$. For example, when $K = 0$, it must be that $a_{00} = 1$, so that $h_K(\boldsymbol{u})$ reduces to a standard normal density, and thus from (10.30) $\boldsymbol{b}_i \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{S}\boldsymbol{S}^T)$. A special case is to take $\boldsymbol{\mu} = \boldsymbol{0}$; otherwise, in ***linear population models*** for which the kth element of $\boldsymbol{\beta}_i$ is of the form

$$\beta_{ki} = \beta_{k0} + \boldsymbol{\beta}_{k1}^T h(\boldsymbol{a}_i) + b_{ki},$$

say, with "***intercept***" β_{k0}, where $h(\cdot)$ is a function of among-individual covariates $\boldsymbol{a}_i$, with arbitrary $\boldsymbol{\mu}$, one would write instead

$$\beta_{ki} = \boldsymbol{\beta}_{k1}^T h(\boldsymbol{a}_i) + b_{ki},$$

so that μ_k, the kth element of $\boldsymbol{\mu}$, plays the role of the population model "intercept." In the following, we take $\boldsymbol{\mu} = \boldsymbol{0}$.

More generally, to ensure (10.32), note from (10.31) that it must be that

$$E\{P_K^2(\boldsymbol{U}, \boldsymbol{a})\} = 1, \quad \text{where } \boldsymbol{U} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_q). \tag{10.33}$$

It is possible to show that the expectation in (10.33) can be written as a ***quadratic form*** involving the coefficients in $P_K^2(\boldsymbol{U}, \boldsymbol{a})$ and ***moments of a standard normal random variable***, as follows. For example, for $q = 2$, there are $d = (K+1)(K+2)/2$ distinct coefficients; for $K = 3$,

$$\boldsymbol{a} = (a_{00}, a_{10}, a_{01}, a_{20}, a_{11}, a_{02}, a_{30}, a_{21}, a_{12}, a_{03})^T.$$

is the $(d \times 1)$ vector of these coefficients. Writing the ℓth element of $\boldsymbol{a}$ as $a_{\ell_1 \ell_2}$, $\ell = 1, \ldots, d$, let $\boldsymbol{U}_a$ be the random vector whose ℓth element is $U_1^{\ell_1} U_2^{\ell_2}$, where U_1 and U_2 are independent, standard normal random variables, $\ell = 1, \ldots, d$. Then it can be shown that $P_K(\boldsymbol{U}, \boldsymbol{a}) = \boldsymbol{a}^T \boldsymbol{U}_a$, so that

$$E\{P_K^2(\boldsymbol{U}) = \boldsymbol{a}^T E(\boldsymbol{U}_a \boldsymbol{U}_a^T)\boldsymbol{a} = \boldsymbol{a}^T \boldsymbol{A}\boldsymbol{a}, \tag{10.34}$$

where $\boldsymbol{A} = E(\boldsymbol{U}_a \boldsymbol{U}_a^T)$ is the matrix with (ℓ, ℓ') element $E(U_1^{\ell_1+\ell'_1})E(U_2^{\ell_2+\ell'_2})$. These moments are readily available via ***standard recursive formulæ***. This formulation can be ***generalized*** to any q and K. It follows from (10.34) that (10.33) can be written as

$$\boldsymbol{a}^T \boldsymbol{A}\boldsymbol{a} = 1. \tag{10.35}$$

Thus, from (10.30) with $\boldsymbol{\mu} = \boldsymbol{0}$ and (10.31), we can represent the density of $\boldsymbol{b}_i$ for fixed K as

$$h_K(\boldsymbol{b}_i, \boldsymbol{a}) = P_K^2(\boldsymbol{S}^{-1}\boldsymbol{b}_i, \boldsymbol{a})\, p(\boldsymbol{b}_i; \boldsymbol{0}, \boldsymbol{S}\boldsymbol{S}^T), \tag{10.36}$$

where $p(\cdot\,; \boldsymbol{0}, \boldsymbol{S}\boldsymbol{S}^T)$ is the $\mathcal{N}(\boldsymbol{0}, \boldsymbol{S}\boldsymbol{S}^T)$ density, and $\boldsymbol{a}$ satisfies the ***constraint*** (10.35). It follows from (10.36) that, with an individual model involving within-individual covariance parameters γ and a population model involving fixed parameters β,

$$p(\boldsymbol{y}_i | \boldsymbol{x}_i; \beta, \gamma, \boldsymbol{D}) = \int p(\boldsymbol{Y}_i | \boldsymbol{x}_i, \boldsymbol{b}_i; \beta, \gamma)\, P_K^2(\boldsymbol{S}^{-1}\boldsymbol{b}_i, \boldsymbol{a})\, p(\boldsymbol{b}_i; \boldsymbol{0}, \boldsymbol{S}\boldsymbol{S}^T)\, d\boldsymbol{b}_i, \tag{10.37}$$

where $\boldsymbol{D}$ is now determined by $\boldsymbol{S}$ and $\boldsymbol{a}$. Using the representation (10.37), one can write down loglikelihood for β, γ, $\boldsymbol{a}$, and $\boldsymbol{S}$ and maximize in these parameters ***subject to the constraint*** (10.35).

- Zhang and Davidian (2001) show that, for a ***linear mixed effects model***, the loglikelihood corresponding to (10.37) can be expressed in a ***closed form***.

- For general ***nonlinear*** models, the loglikeliood is ***not*** available in a closed form in general. In this case, Davidian and Gallant (1993) suggest "doing" the integrals in (10.36) using ***Gaussian quadrature*** or ***Monte Carlo integration***, taking advantage of the fact that they depend on integration against a ***normal density***.

- Zhang and Davidian (2001) show further that, regardless of the form of the model, the constraint (10.35) can be imposed "***automatically***" by a ***reparameterization*** via a ***spherical transformation*** of (10.35).

- As with the mixture of normals representation, it is suggested to choose K by fitting the model first for $K = 0$ corresponding to normal random effects, and followed by successive fits with $K = 1, 2$, and 3, and selecting K via inspection of ***information criteria***.

PHARMACOKINETICS OF PHENOBARBITAL IN NEONATES, continued. We conclude with a brief look at an analysis of the phenobarbital PK study using the SNP representation for the random effects; a full account is in Chapter 7 of Davidian and Giltinan (1997). Assuming the ***individual model*** (9.98) with repeated dosing and the ***population model***

$$(ii) \quad \beta_{1i} = \beta_1 + \beta_3 w_i + b_{1i}, \quad \beta_{1i} = \beta_2 + \beta_4 w_i + b_{2i},$$

the density of $\boldsymbol{b}_i = (b_{1i}, b_{2i})^T$ is represented by the SNP approximation (10.30) and (10.31).

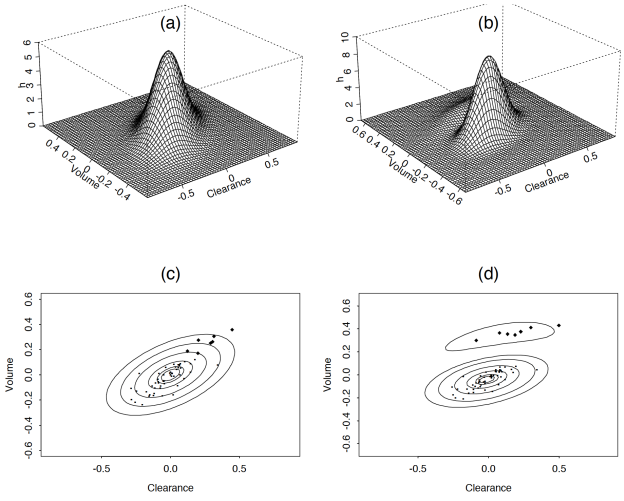

Figure 10.3: *Estimated random effects densities for $\boldsymbol{b}_i$ using the SNP approach. The left hand column shows the estimated density and contour plot with empirical Bayes estimates of random effects superimposed when $K = 0$, corresponding to normal random effects. The right hand column shows the same when $K = 2$ in the SNP representation.*

Figure 10.3 shows plots of the ***estimated random effects density*** with ***empirical Bayes estimates*** of the random effects superimposed for $K = 0$, corresponding to the usual assumption of normal random effects, and $K = 2$, the preferred fit based on inspection of ***information criteria***. The estimated density for $K = 2$ shows a ***clear second mode***.

The random effects estimates for seven infants are shown as diamonds; these seven infants appear to correspond to the second mode and thus seem to arise from a separate ***subpopulation*** relative to the rest of the infants. These infants can be seen upon inspection of the data to have low measured phenobarbital concentrations after the loading dose. Because the initial concentration measurement is highly influential for determining the estimate of log ***volume of distribution*** $\log V_i$, the observed pattern makes sense. Given that birthweight is already accounted for in the population model and Apgar score does not seem associated with either $\log Cl_i$ or $\log V_i$ from the analysis in Section 9.7, it is not possible to explain this by a possible systematic association with an infant characteristic. It may well be that a relevant, ***unmeasured attribute*** is being reflected here.