Chapter 6

Repeated Measures Data and Random Parameter Models

January 13, 2001

Part of *Beyond Traditional Statistical Methods*

Developed as part of NSF/ILI grant DUE9751644.

Objectives

This chapter explains

- Applications of growth curve models to describe the results of statistical studies in which repeated measures are made on a sample of units from some population or process.
- Different kinds of models for repeated measure data:
  - Empirical
  - Mechanistic (e.g., from systems of differential equations)
- How to fit and interpret nonlinear regression models.
- Different sources of variability in repeated measures data.
- The use of random parameter models to describe variability in repeated measures data.
- Methods for data analysis and inference for repeated measure data.

Overview

6.1 Introduction

6.1.1 Repeated measures data

The term "repeated measures data" data refers to taking repeated measurements or observations on a single unit. In some applications such measurements are taken over time. In other applications, the readings may be spatial (as in a number of observations across a plot of land or across a specimen of material).

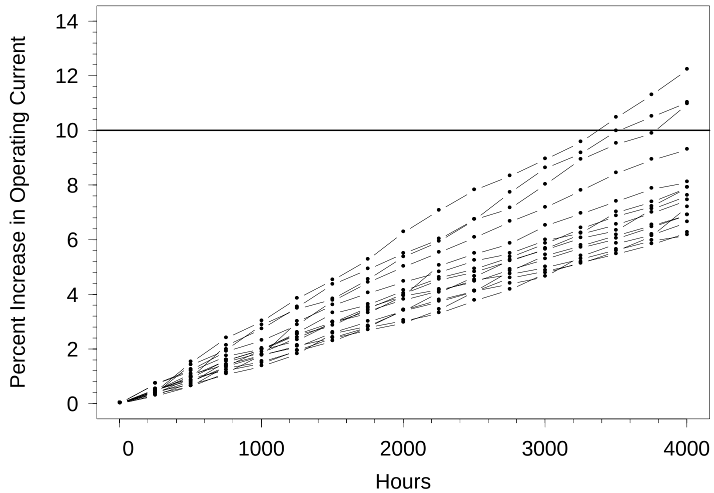


Figure 6.1: Measured percentage increase in operating current over time for GaAs lasers tested at 80°C.

In pharmacokinetic studies, a drug is introduced into a biological unit (e.g., an animal or human) and blood plasma concentrations at different points in time. Another important application arises in the study of growth (e.g., of trees, humans, fatigue cracks)

In some reliability studies, it is possible to measure the physical degradation on units (either in a laboratory test or on units that are in service) as a function of time (e.g., tire wear). In other applications actual physical degradation cannot be observed directly, but measures of product performance degradation (e.g., power output) may be available. Both kinds of data are generically referred to as "degradation data" and we will follow this convention. Modeling performance degradation may be useful, but could be complicated because performance may be affected by more than one underlying degradation process. Depending on the application, degradation data may be available continuously or at specific points in time where measurements are taken.

6.2 Applications

Example 6.1 Laser Operating Current as a Function of Time. Over the life of some laser devices, degradation causes a decrease in light output. Other lasers, however, contain a feedback mechanism that will maintain nearly constant light output by increasing operating current as the laser degrades. When operating current gets too high, the device is considered to have failed. Figure 6.1 shows the increase in operating current over time for a sample of GaAs lasers tested at 80°C (this temperature, much higher than the use temperature, was used to accelerate the failure mechanism so that degradation information would be obtained more rapidly—see Chapter 18 of Meeker and Escobar (1998).

∎

Example 6.2 Plasma concentrations of indomethicin following intravenous injection. Figure 6.2 shows blood plasma concentrations of indomethicin following bolus intravenous injection for six human volunteers. The data were originally published in Kwan et al. (1976) and were reanalyzed by Davidian and Giltinan (1995, page 18). Pharmacokinetic researchers need to know the rate of passing to elimination and the time (or expected time) until a specified level of indomethicin is left. Interest centers on the identification of demographic and physiological characteristics explaining the variability in drug response. This kind of information is important for choosing effective dosage regimes.

∎

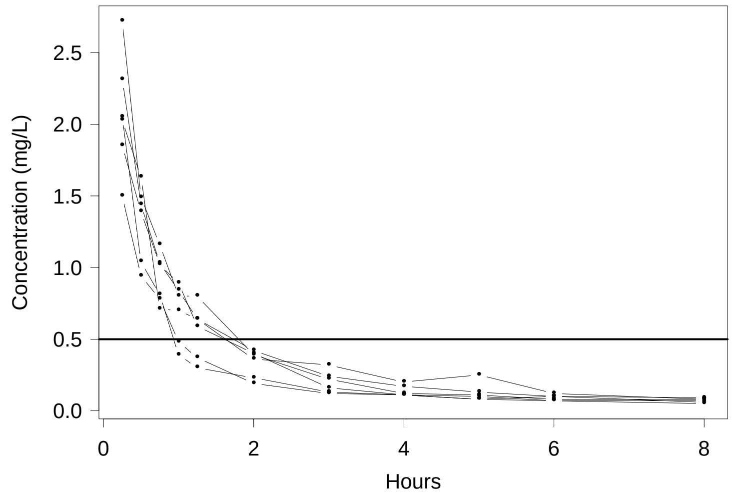


Figure 6.2: Plasma concentrations of indomethicin following intravenous injection.

Example 6.3 Serum concentrations of theophylline following oral administration.
The data in Figure 6.3 show serum concentrations of theophylline following oral administration (theophylline is an anti-asthmatic agent). Data collected on twelve subjects. Serum concentrations were measured at 11 time points from 15 minutes to to 24 hours. Dosage was varied in proportion to body weight. The data were first analyzed by Boeckmann et al. (1992) and reanalyzed by Davidian and Giltinan (1995). Note that when compared to the intravenous injection in Example 6.2, with the oral administration, there is a longer delay period before serum concentrations increase to a high level.

Example 6.4 Growth of conducting filaments in printed circuit boards. Meeker and LuValle (1995) describe models for growth of failure-causing conducting filaments of chlorine-copper compounds in printed-circuit boards. These filaments cause failure when they reach from one copper-plated through-hole to another. In their model, $A_1(t)$ is the amount of chlorine available for reaction and $A_2(t)$ is proportional to the amount of failure-causing chlorine-copper compounds at time t. Under appropriate conditions of temperature, humidity, and electrical charge, copper combines with chlorine A_1 to produce the chlorine-copper compound A_2 with rate constant k_1. Diagrammatically,

$$A_1 \xrightarrow{k_1} A_2.$$

The rate equations for this process are

$$\frac{dA_1}{dt} = -k_1 A_1 \quad \text{and} \quad \frac{dA_2}{dt} = k_1 A_1. \tag{6.1}$$

The solution of this system of differential equations gives

$$A_1(t) = A_1(0) \times \exp(-k_1 t) \tag{6.2}$$

$$A_2(t) = A_2(0) + A_1(0) \times [1 - \exp(-k_1 t)] \tag{6.3}$$

where $A_1(0)$ and $A_2(0)$ are initial amounts. To simplify notation, let $A_2(\infty) = A_1(0) + A_2(0)$. Then if $A_2(0) = 0$, the solution for $A_2(t)$ (the quantity of primary interest) can be expressed as

$$A_2(t) = A_2(\infty) \times [1 - \exp(-k_1 t)]. \tag{6.4}$$

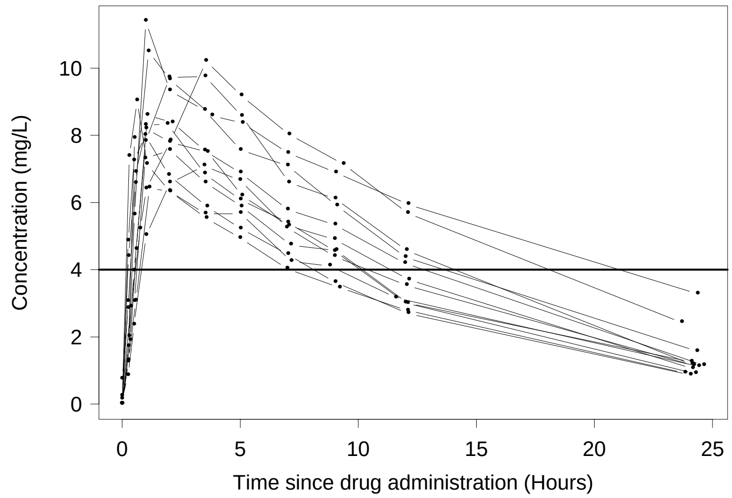


Figure 6.3: Serum concentrations of theophylline following oral administration.

6.3 Models for Repeated Measures Data

6.3.1 Simple models for change over time.

The simplest growth model is a model for linear growth.

Example 6.5 Constant-rate growth model.

For some types of growth, rate will be approximately constant. In the form of a differential equation

$$\frac{dy(t)}{dt} = \beta_1$$

With $y(0) = \beta_0$, this simple differential equation has the solution

$$y(t) = \beta_0 + \beta_1 t$$

where β_0 is the intercept and β_1 is the slope (growth rate). This linear model is useful for describing certain kinds of growth patterns.

6.3.2 Compartmental models for change over time.

Example 6.6 A two-compartment open model.

A two-compartment open model illustrated in Figure 6.5 is commonly used to describe the concentration in the blood over time.

The following differential equations describe the two compartment model

$$\frac{dA_1(t)}{dt} = -k_1 A_1(t), \quad \text{and} \quad \frac{dA_2(t)}{dt} = k_1 A_1(t) - k_2 A_2(t).$$

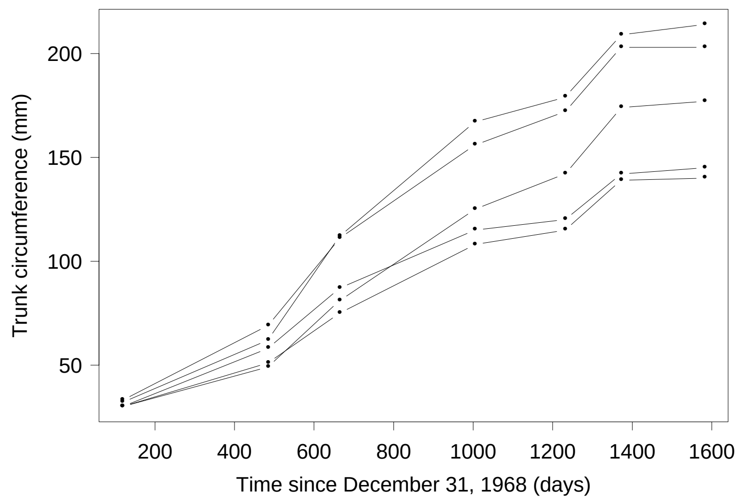


Figure 6.4: Orange Tree Circumference Growth.

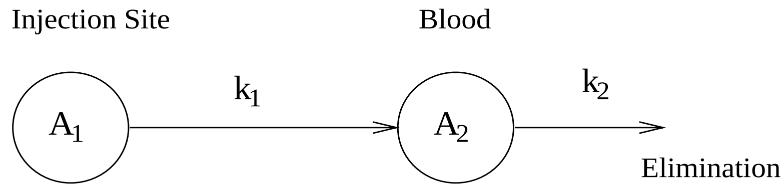


Figure 6.5: A two-compartment open model.

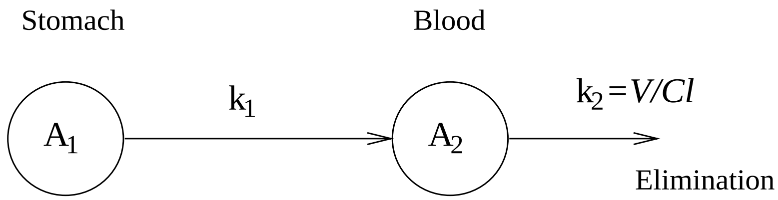


Figure 6.6: A two-compartment open model.

Then, using standard methods for solving systems of linear differential equations (or a mathematics computer program like Maple of Mathematics), the concentration in the blood at time t is

$$\begin{aligned} A_2(t) &= \theta_2 \exp(-k_2 t) - \theta_1 \exp(-k_1 t) \\ \theta_1 &= A_1(0) k_1/(k_1 - k_2) \\ \theta_2 &= \theta_1 + A_2(0) \end{aligned}$$

Splus has the `biexponential` function for this model. ▮

Example 6.7 A first order log model A two-compartment (gut and blood) model is used to describe the concentration in the blood as a function of time for oral administration of a drug. This model is illustrates in Figure 6.5.

$$\frac{dA_1(t)}{dt} = -k_1 A_1(t), \quad \text{and} \quad \frac{dA_2(t)}{dt} = k_1 A_1(t) - \frac{Cl}{V} A_2(t)$$

where $A_1(t)$ is the amount in the stomach at time t and $A_2(t)$ is the concentration in the blood at time t. The rate constant k_1 has units 1/hour, clearance Cl has units L/hr/kg, and volume of distribution V has units L/kg. Thus the elimination rate constant $k_2 = Cl/V$ has units 1/hour.

▮

6.3.3 Models for variation in repeated measures data

If all units were identical, exposed to exactly the same environment and/or conditions, and in exactly the same environment then, according to the simple deterministic models above, all units would follow exactly the same path. Of course, there is some degree of variability in all of these model factors as well as in various factors that are not in the model. These factors combine to cause variability in the observed sample paths.

Unit-to-unit variability. The following are examples of sources of unit-to-unit variability:

- **Initial conditions.** Individual units will vary with respect to the amount of material available to wear, initial level of degradation, amount of harmful degradation-causing material, and so on. Figure 6.7 shows the Paris model for growth of fatigue cracks, with simulated variability in the size of the initial crack, but with all other of the unit's Paris model characteristics and other factors held constant.

- **Material properties**. Figure 6.8 shows the Paris model for growth of fatigue cracks, allowing for unit-to-unit variability in the material properties parameters C and m and the size of the initial crack. In this case, as shown in the Paris model in (**??**), the rate of growth depends on C and m which differ from unit to unit. This yields crossing of the crack-growth curves (typical of what is observed in actual fatigue testing).

- **Component geometry or dimensions.** Unit-to-unit variability in component geometry or dimensions can, for example, cause additional unit-to-unit variability in degradation rates [e.g., through the $\Delta K(a)$ function in (**??**)].

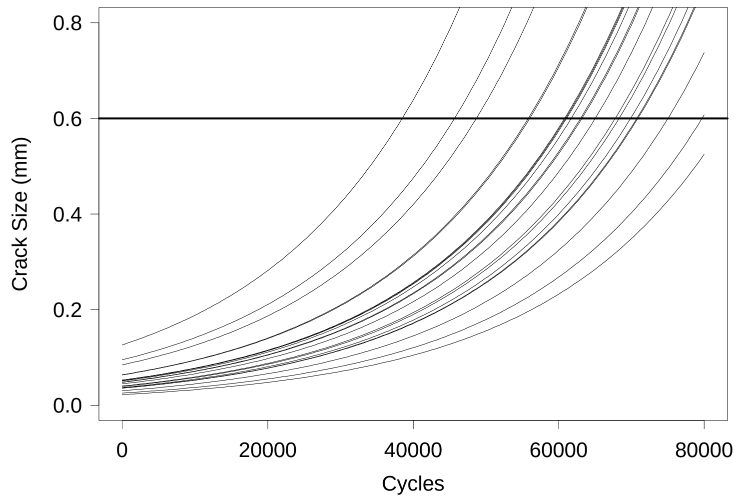


Figure 6.7: Plot of Paris model for growth of fatigue cracks with unit-to-unit variability in the initial crack size a_0 but with constant materials parameters (C and m) and constant stress.

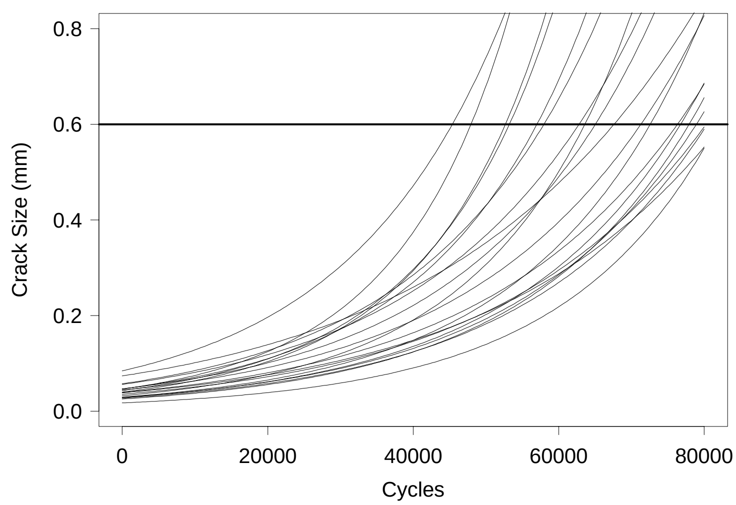


Figure 6.8: Plot of Paris model for growth of fatigue cracks with unit-to-unit variability in the initial crack size and in materials parameters C and m, but with constant stress.

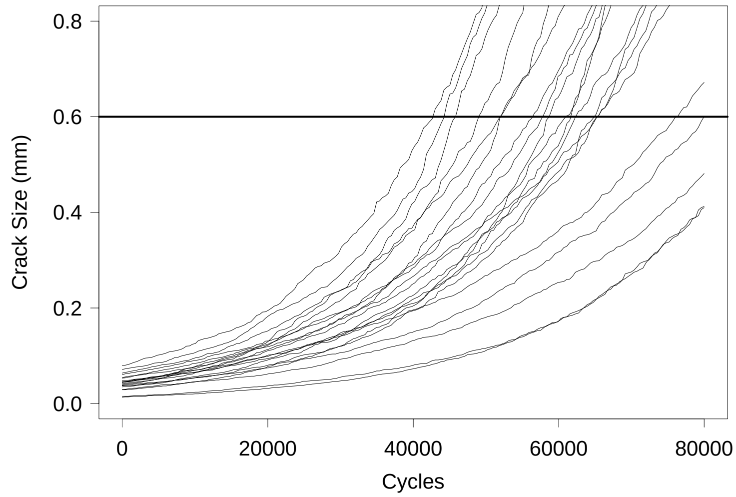


Figure 6.9: Plot of Paris model for growth of fatigue cracks with unit-to-unit variability in the initial crack size and materials parameters C and m, and with a stochastic process model for the changes in stress over the life of the unit.

- **Within-unit variability.** Often there will be spatial variability in materials properties within a unit (e.g., defects).

Variability due to operating and environmental conditions. Besides the materials properties described above, the rate of degradation will depend on operating and environmental conditions. For example, $K(a)$ in the Paris model (**??**) depends on the amount of applied stress and the Paris parameters can depend on temperature. In laboratory fatigue tests, the stress is either fixed or changing in a systematic manner (e.g., to keep $K(a)$ nearly constant as a increases). In actual operation of most components, stress would generally be a complicated function over time. Such variations possibly described by a stochastic process model. Figure 6.9 shows the Paris model with degradation rate varying due to variations in stress that might have been caused, for example, by variation in driving conditions encountered, over time, by an automobile. In some applications, shocks that occur randomly in time can dominate other sources of variability in a failure-causing process.

The models described here are simple relative to more exact theory of failure-causing processes that, almost certainly exist (but may not be known or understood). For some purposes, however, such simple first-order descriptions are useful.

6.3.4 Model for change over time

The actual path of a particular unit over time is denoted by $\mathcal{D}(t), t > 0$. In applications, values of $\mathcal{D}(t)$ are sampled, for a particular unit, at discrete points in time $t_1, t_2, \ldots$. The times of observation may differ from unit to unit. The observed sample degradation y_{ij} of unit i at time t_{ij} is

$$y_{ij} = \mathcal{D}_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, n, \quad j = 1, \ldots, m_i \tag{6.5}$$

where $\mathcal{D}_{ij} = \mathcal{D}(t_{ij}, \beta_{1i}, \ldots, \beta_{ki})$ is the actual path of the unit i at time t_{ij} (the times need not be the same for all units) and $\epsilon_{ij} \sim \text{NOR}\,(0, \sigma_\epsilon)$ is a residual deviation for unit i at time t_j. The total number of measurements on unit i is denoted by m_i. For the ith unit, $\beta_{1i}, \ldots, \beta_{ki}$ is a vector of k unknown parameters. Typically sample paths have $k = 1$, 2, 3 or 4 parameters. As described in Section 6.3.3, some of the $\beta_1, \ldots, \beta_k$ parameters will be random from unit-to-unit. One or more of the $\beta_1, \ldots, \beta_k$ parameters could, however, be modeled as common across all units.

The scales of y and t can be chosen, as suggested by physical theory and the data, to simplify the form of $\mathcal{D}(t, \beta_1, \ldots, \beta_k)$. For example, the relationship between the logarithm of degradation and the logarithm of time might be modeled by the additive relationship in (6.5). Model choice requires not only specification of the form of the $\mathcal{D}(t, \beta_1, \ldots, \beta_k)$ function, but also specification of which of the $\beta_1, \ldots, \beta_k$ are random (differing from unit to unit) and which are fixed (common to all units). Because of the flexibility in specifying the form of $\mathcal{D}(t, \beta_1, \ldots, \beta_k)$, and of the way in which the $\beta_1, \ldots, \beta_k$ come into this form, we can, for simplicity, model the unit-to-unit variability in $\beta_1, \ldots, \beta_k$ with a multivariate normal distribution with mean vector $\boldsymbol{\mu_\beta}$ and covariance matrix $\Sigma_{\boldsymbol{\beta}}$.

It is generally assumed that the random $\beta_1, \ldots, \beta_k$ are independent of the ϵ_{ij} deviations. Another common assumption is that σ_ϵ is constant. The adequacy of this assumption can be affected by transforming $\mathcal{D}(t)$. Because the y_{ij} are taken serially on a unit, however, there is potential for autocorrelation among the $\epsilon_{ij}, j = 1, \ldots, m_i$ values, especially if there are many closely-spaced readings. In many practical applications involving inference on the degradation of units from a population or process, however, the correlation is weak and, moreover, dominated by the unit-to-unit variability in the $\beta_1, \ldots, \beta_k$ values and thus can be ignored. In situations where autocorrelation cannot be ignored, one can use a time series model for the residual term along with appropriate estimation methods. This subject is discussed in Section xx.

6.3.5 Model parameters for individual units

Although the values of $\beta_1, \ldots, \beta_k$ for the individual units may be of interest in some applications (e.g., to predict the future path of a particular unit, based on a few early readings), subsequent development in this chapter will concentrate on the use of repeated measures data to make inferences about the population or process or predictions about future units. In this case, the underlying model parameters are $\boldsymbol{\mu_\beta}$ and $\Sigma_{\boldsymbol{\beta}}$, as well as the residual standard deviation σ_ϵ. For shorthand, we will use $\boldsymbol{\theta_\beta} = (\boldsymbol{\mu_\beta}, \Sigma_{\boldsymbol{\beta}})$ to denote the overall population/process parameters.

6.4 Estimation of Population Model Parameters

The likelihood for the random-parameter growth curve model in Section 6.3.4 can be expressed as

$$L(\boldsymbol{\theta_\beta}, \sigma_\epsilon | \text{DATA}) = \prod_{i=1}^{n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left[\prod_{j=1}^{m_i} \frac{1}{\sigma_\epsilon} \phi_{\text{nor}}(\zeta_{ij}) \right] f_{\boldsymbol{\beta}}(\beta_{1i}, \ldots, \beta_{ki}; \boldsymbol{\theta_\beta}) d\beta_{1i}, \ldots, d\beta_{ki} \tag{6.6}$$

where $\zeta_{ij} = [y_{ij} - \mathcal{D}(t_{ij}, \beta_{1i}, \ldots, \beta_{ki})]/\sigma_\epsilon$ and $f_{\boldsymbol{\beta}}(\beta_{1i}, \ldots, \beta_{ki}; \boldsymbol{\theta_\beta})$ is the multivariate normal distribution density function. Each evaluation of (6.6) will, in general, require numerical approximation of n integrals of dimension k (where n is the number of sample paths and k is the number of random parameters in each path). Maximizing (6.6) with respect to $(\boldsymbol{\mu_\beta}, \Sigma_{\boldsymbol{\beta}}, \sigma_\epsilon)$ directly, even with today's computational capabilities, is extremely difficult unless $\mathcal{D}(t)$ is a linear function. Pinheiro and Bates (1995a) describe and compare estimation schemes that provide approximate ML estimates of $\boldsymbol{\theta_\beta} = (\boldsymbol{\mu_\beta}, \Sigma_{\boldsymbol{\beta}})$ and σ_ϵ, as well as the unit-specific components in $\beta_{1i}, \ldots, \beta_{ki}, i = 1, \ldots, n$. Pinheiro and Bates (1995b) implement a modification of the method of Lindstrom and Bates (1990). The examples in this chapter were computed with the Pinheiro and Bates (1995b) program implemented in S-PLUS as function `nlme`.

Example 6.8 Estimates of fatigue data model parameters for Alloy-A. Continuing with Example **??**, we fit the model in (6.5) with $\mathcal{D}_{ij} = a(t)$ in (**??**), $a(0) = .9$, Stress=1, $\beta_1 = m$ and $\beta_2 = C$, modeling (β_1, β_2) with a bivariate normal distribution. The program of Pinheiro and Bates (1995b) gives the following approximate ML estimates.

$$\widehat{\boldsymbol{\mu}}_{\boldsymbol{\beta}} = \begin{pmatrix} 5.17 \\ 3.73 \end{pmatrix}, \qquad \widehat{\Sigma}_{\boldsymbol{\beta}} = \begin{pmatrix} .251 & -.194 \\ -.194 & .519 \end{pmatrix}.$$

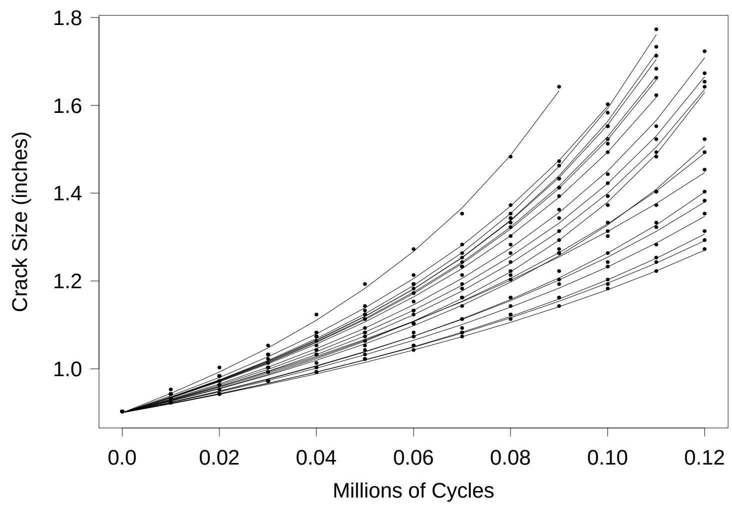


Figure 6.10: Alloy-A fatigue crack size observations and fitted Paris-rule model.

and $\widehat{\sigma}_\epsilon = .0034$. Figure 6.10 shows the fitted Paris relationship for each of the sample paths (indicated by the points on the plot) for the Alloy-A fatigue-crack data. Figure 6.11 is a scatter plot of the estimates of the Paris relationship parameters for each of the 21 sample paths, indicating the reasonableness of the bivariate normal distribution model for this random-coefficients model. ∎

6.5 Evaluating Crossing-Time Distributions

6.5.1 Soft failures: specified degradation level

For some applications there is interest in the distribution of time to reach a certain level the measurement . Then failure would be defined (in a somewhat arbitrary, but purposeful, manner) at a specified level of degradation (e.g., 60% of initial output). We call this a "soft failure" definition.

A fixed value of $\mathcal{D}_\mathrm{f}$ will be used to denote the critical level for the path. Fro example a failure time T could be defined as the time when the actual path $\mathcal{D}(t)$ crosses the critical degradation level $\mathcal{D}_\mathrm{f}$. We use t_c to denote the planned stopping time in the experiment (as illustrated in Figure **??**). Inferences are desired on the crossing-time distribution of a particular product or material.

6.6 Evaluation of $F(t)$

Bibliographic Notes

Davidian and Giltinan (1995) provide an excellent description and development of methods for fitting statistical models to nonlinear degradation-type models (also known as "growth curves") and estimating random parameters.

References

Bates, D. M., and Watts, D. G. (1988), *Nonlinear Regression Analysis and its Applications,* New York, NY: John Wiley & Sons, Inc.

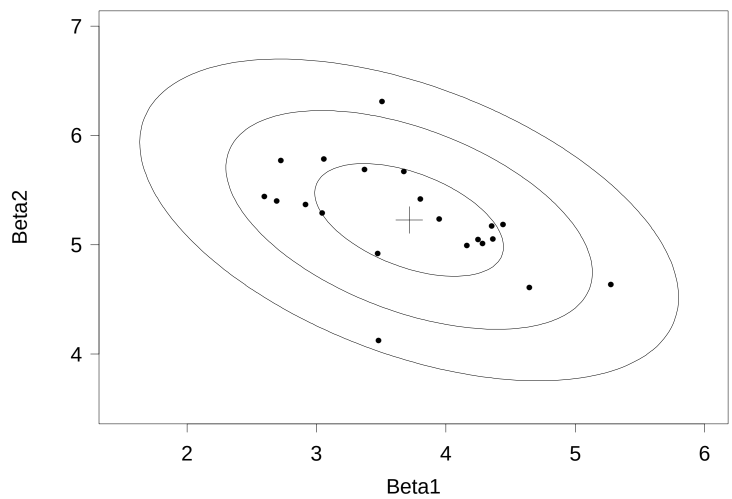


Figure 6.11: Plot of $\widehat{\beta}_{1i}$ versus $\widehat{\beta}_{2i}$ for the $i = 1, \ldots, 21$ sample paths from the Alloy-A fatigue crack size data. The contour lines represent the fitted bivariate normal distribution for β_1 and β_2.

Davidian, M., and Giltinan, D. M. (1995), *Nonlinear Models for Repeated Measurement Data*, London: Chapman & Hall.

Dowling, N. E. (1993), *Mechanical Behavior of Materials*, Englewood Cliffs, NJ: Prentice Hall.

Efron, B. (1985), Bootstrap confidence intervals for a class of parametric problems, *Biometrika* **72**, 45–58.

Kwan et al. (1976), Kinetics of indomethicin absorption, elimination, and enterohepatic circulation in man, *Journal of Pharmacokinetics and Biopharmaceutics*, 4, 255-280.

Lindstrom, M. J., and Bates, D. M. (1990), Nonlinear mixed effects models for repeated measures data, *Biometrics* **46**, 673–687.

Lu, C. J., and Meeker, W. Q. (1993), Using degradation measures to estimate a time-to-failure distribution. *Technometrics* **34**, 161-174.

Meeker, W. Q., and LuValle, M. J. (1995), An accelerated life test model based on reliability kinetics, *Technometrics* **37**, 133-146.

Murray, W. P. (1993), Archival life expectancy of 3M magneto-optic media, *Journal of the Magnetics Society of Japan* **17**, Supplement S1, 309-314.

Palmer, M. J., Phillips, B. F., and Smith, G. T. (1991), Application of nonlinear models with random coefficients to growth data, *Biometrics* **47**, 623-635.

Pinheiro, J. C., and Bates, D. M. (1995a), Approximations to the loglikelihood function in the nonlinear mixed effects model, *Journal of Computational and Graphical Statistics* **4**, 12-35.

Pinheiro, J. C., and Bates, D. M. (1995b), Mixed effects models, methods, and classes for S and SPLUS. Department of Statistics, University of Wisconsin. Available from Statlib.

Seber, G. A. F., and Wild, C. J. (1989), *Nonlinear Regression*, New York: John Wiley & Sons.

Venables, W. N. and Ripley, B. D. (1997), *Modern Applied Statistics with S-Plus*, Second Edition, New York: Springer.

Ziegel, E. R. and Gorman, J. W. (1980), Kinetic modeling with multiresponse data," *Technometrics* 22, 139-151.

Exercises

6.1 In file echapter06.q, you will find commands for fitting nonlinear mixed effect models for the serum concentrations of theophylline following oral administration data. Use these data to do the following

(a). Examine carefully the different plots of the data. Look particularly at the results of the `plot(Theoph.rmd)` plot. What conclusions can your draw from this plot?

(b). Use the `nlsList` command to fit the "first order log" model separately to the data from each each patient. Create a matrix of the coefficients for the fits. This matrix can be considered as a "data set" where the coefficients are the observed data. Compute the means and standard deviations, and the "raw" covariance and correlation matrices for these "data." What do the results tell you?

Note that the function `two-stage` does these computations automatically. You can see the inside of `two-stage` by just typing this name at the Splus prompt.

(c). Under the assumption that the random effects lCl, lka, and lke have a multivariate normal distribution. The naive estimate of the population covariance matrix for the random effects provided by the simple covariance matrix computed by `two-stage` is, in general biased. Why? Hint: imagine what happens if there is a substantial amount of measurement error in the data.

(d). Use the `Theoph.nlme <- nlme(Theoph.lis)` command to compute the approximate ML estimates of the nonlinear mixed effect model. Examine the summary of the results.

(e). Compare the estimates of the means of the fixed effects with the estimates of the means from the `two-stage` command.

(f). Examine the correlations between the random effects and the plot given by `plot(Theoph.nlme)`. Notice the correspondence. Think of the biological system being modeled. Can you explain any of these correlations in terms of this biological system?

(g). Compare the estimates of the standard deviations of the random effects with the estimates of the standard deviations from the `two-stage` command. What do you notice. Can you explain why the standard deviations are larger with the `two-stage` command?

(h). The `summary(Theoph.nlme)` command gives only estimates of standard deviations and correlations between the random effects. It does not print out a covariance matrix (similar to the one from `two-stage`. Why? The model was parameterized in terms of log clearance (lCl) and log rates (lka and lke for $\log(k_a)$ and $\log(k_e)$, respectively). One question of interest is the average rates among all individuals. Use information in the summary to obtain 95% confidence intervals for these rates (not log rates).

(i). Use the output from the `summary(Theoph.nlme)` command to compute a prediction for the concentration in Subject 2 at time 5.10. Compare this with the output from the `predict(Theoph.nlme,data=Theoph,clus` `Subject)` command.

(j). How would you define residuals for this model? Are residuals useful for analyzing date with `nlme` models? Explain.

6.2 Refer to the output from the `plot(Theoph.rmd)` and the `plot.fits` commands. One important question to investigators is the length of time that the concentration stays above a certain level, say 4 mcg/mL. This time is a random variable.

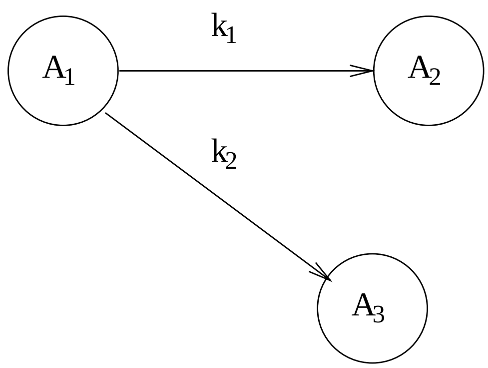
(a). Describe some of the possible sources of variability in this quantity. Order these according to the amount of variability that they contribute.

(b). Suppose that you wanted to estimate the distribution of this random variable using the results of the analysis of the available data. List the steps that you would follow to compute a simulation-based estimator (as if you were writing these down for a computer programmer who understands statistics, probability, and random number generation).

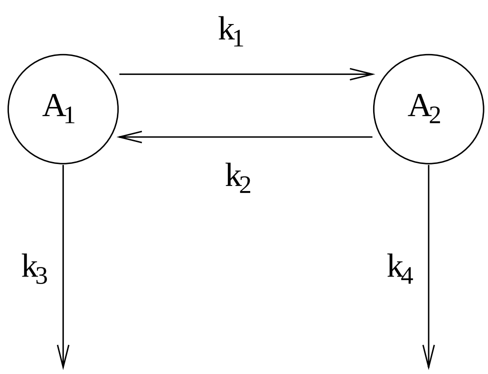
(c). The estimate that you compute from the above simulation-based method will have some sampling error, resulting from the small number of subjects that were used in the study (there will also be some "Monte Carlo error," but this can be made arbitrarily small by increasing the number of simulation trials). Describe how one might compute a confidence interval that could be used to quantify this sampling error.

6.3 Consider the following diagrams illustrating some compartmental models. For each, write down the appropriate system of differential equations. Use Maple to solve the differential equations. Translate the Maple output into well-formed-formulas for the specified compartment(s) the diagram, using the specified initial conditions.

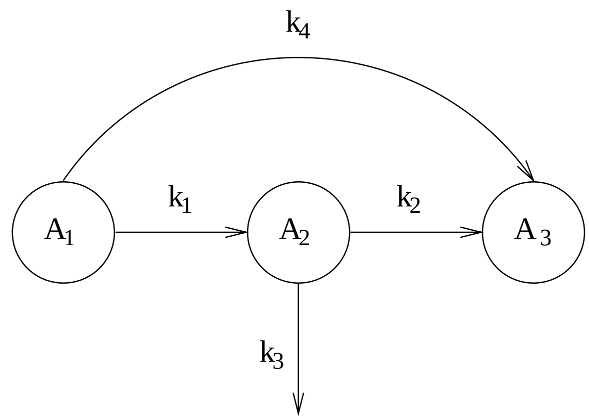
(a). The following diagram might be called a "competing" diffusion process. Interest centers on A_2 as a function of t. Suppose that $A_1(0) = C$ and $A_2(0) = 0$.



(b). The following diagram is a two-compartment open system. Interestingly, if only A_1 is observed, it is impossible to identify the different rates. Interest centers on A_2 as a function of t. Suppose that $A_1(0) = C$ and $A_2(0) = 0$.



(c). The following diagram has been used to model a pyrolysis process used to extract oil from oil shale rock (Ziegel and Gorman 1980). Interest centers on A_1 and A_2 as a function of t. Suppose that $A_1(0) = C$, $A_2(0) = 0$ and $A_3(0) = 0$.



6.4 Show that (6.2) is the solution to (6.1).

6.5 Use the degradation equation (6.4) to obtain an expression for the time that $A_2(t)$ crosses a specified A_{2f}.

6.6 Determine the value of N needed in Algorithm **??** to evaluate $F(t)$ at the point where $F(t) = .01$ so that the probability that Monte Carlo error in evaluation is less than .0001, with probability approximately .95. Use the normal distribution approximation to the binomial probability.

6.7 Appendix Table C.18 of Meeker and Escobar (1998) gives degradation data on block error rates (the ratio of number of bytes with errors to the total number of bytes tested) of magneto-optic data storage disks tested for 2000 hours at 80°C and 85% relative humidity. Use the simple analysis method described in Section **??** to estimate the failure-time distribution of these disks at 80°C and 85% relative humidity where failure is defined as the time that it takes to reach an error rate of 50×10^{-5} (a safe level at which error correcting codes can be expected to correct errors).

6.8 As an electronic device ages, its power output decreases. Because the degradation results from a simple one-step chemical reaction with a limited amount of harmful material available for reaction, the decrease in power can be described by the degradation model $\mathcal{D}(t) = \beta_2[1 - \exp(-\beta_1 t)]$ where $\mathcal{D}(t)$ is the power output at time t, $\beta_2 < 0$ is nearly the same for all units, and β_1, comes from a lognormal distribution. System performance degrades noticeably when $\mathcal{D}(t)$ falls below $\mathcal{D}_{\mathrm{f}}$. Thus we define a failure as the point in time when $\mathcal{D}(t) < \mathcal{D}_{\mathrm{f}}$.

(a). Describe some possible physical reasons for the asymptotic behavior of $\mathcal{D}(t)$.

(b). What happens, in the long run, if $\beta_2 > \mathcal{D}_{\mathrm{f}}$?

(c). Assuming that $\mathcal{D}_{\mathrm{f}}$ is a fixed constant and that $\beta_2 < \mathcal{D}_{\mathrm{f}}$, derive an expression for $F(t)$, the failure time cdf.

6.9 Suppose the actual degradation path of a particular unit is given by $\mathcal{D}(t) = \beta_1 t$ where $-\log(\beta_1)$ varies from unit to unit according to a $\mathrm{SEV}(\mu, \sigma)$ distribution. If failure occurs when $\mathcal{D}(t) > \mathcal{D}_{\mathrm{f}}$, where $\mathcal{D}_{\mathrm{f}}$ is a fixed constant, show that the failure time distribution is Weibull.

▲ **6.10** Suppose the actual degradation path of a particular unit is given by $\mathcal{D}(t) = \beta_1 t$ where β_1 varies from unit to unit according to a $\mathrm{LOGNOR}(\mu_1, \sigma_1)$ distribution. Also suppose that failure occurs when $\mathcal{D}(t) > \mathcal{D}_{\mathrm{f}}$, and $\mathcal{D}_{\mathrm{f}}$ has a $\mathrm{LOGNOR}(\mu_2, \sigma_2)$ distribution. Derive an expression for $F(t)$, the failure time cdf.

▲ **6.11** Suppose the actual degradation path of a particular unit is given by $\mathcal{D}(t) = \beta_1 + \beta_2 t$ where (β_1, β_2) vary from unit to unit according to a bivariate normal distribution with parameters μ_{β_1}, μ_{β_2}, σ_{β_1}, σ_{β_2}, and $\rho_{\beta_1\beta_2}$.

(a). Assuming that failure occurs when $\mathcal{D}(t) > \mathcal{D}_{\mathrm{f}}$, derive an expression for the failure time distribution $F(t) = \Pr(T \leq t) = \Pr[\mathcal{D}(t) > \mathcal{D}_{\mathrm{f}}]$.

(b). Explain why $\lim_{t \to \infty} F(t) < 1$.