Chapter 1
A Simple, Linear, Mixed-effects Model

In this book we describe the theory behind a type of statistical model called *mixed-effects* models and the practice of fitting and analyzing such models using the `lme4` package for R. These models are used in many different disciplines. Because the descriptions of the models can vary markedly between disciplines, we begin by describing what mixed-effects models are and by exploring a very simple example of one type of mixed model, the *linear mixed model*.

This simple example allows us to illustrate the use of the `lmer` function in the `lme4` package for fitting such models and for analyzing the fitted model. We also describe methods of assessing the precision of the parameter estimates and of visualizing the conditional distribution of the random effects, given the observed data.

1.1 Mixed-effects Models

Mixed-effects models, like many other types of statistical models, describe a relationship between a *response* variable and some of the *covariates* that have been measured or observed along with the response. In mixed-effects models at least one of the covariates is a *categorical* covariate representing experimental or observational "units" in the data set. In the example from the chemical industry that is given in this chapter, the observational unit is the batch of an intermediate product used in production of a dye. In medical and social sciences the observational units are often the human or animal subjects in the study. In agriculture the experimental units may be the plots of land or the specific plants being studied.

In all of these cases the categorical covariate or covariates are observed at a set of discrete *levels*. We may use numbers, such as subject identifiers, to designate the particular levels that we observed but these numbers are simply labels. The important characteristic of a categorical covariate is that, at each

observed value of the response, the covariate takes on the value of one of a set of distinct levels.

Parameters associated with the particular levels of a covariate are sometimes called the "effects" of the levels. If the set of possible levels of the covariate is fixed and reproducible we model the covariate using *fixed-effects* parameters. If the levels that we observed represent a random sample from the set of all possible levels we incorporate *random effects* in the model.

There are two things to notice about this distinction between fixed-effects parameters and random effects. First, the names are misleading because the distinction between fixed and random is more a property of the levels of the categorical covariate than a property of the effects associated with them. Secondly, we distinguish between "fixed-effects parameters", which are indeed parameters in the statistical model, and "random effects", which, strictly speaking, are not parameters. As we will see shortly, random effects are unobserved random variables.

To make the distinction more concrete, suppose that we wish to model the annual reading test scores for students in a school district and that the covariates recorded with the score include a student identifier and the student's gender. Both of these are categorical covariates. The levels of the gender covariate, male and female, are fixed. If we consider data from another school district or we incorporate scores from earlier tests, we will not change those levels. On the other hand, the students whose scores we observed would generally be regarded as a sample from the set of all possible students whom we could have observed. Adding more data, either from more school districts or from results on previous or subsequent tests, will increase the number of distinct levels of the student identifier.

Mixed-effects models or, more simply, *mixed models* are statistical models that incorporate both fixed-effects parameters and random effects. Because of the way that we will define random effects, a model with random effects always includes at least one fixed-effects parameter. Thus, any model with random effects is a mixed model.

We characterize the statistical model in terms of two random variables: a q-dimensional vector of random effects represented by the random variable $\mathscr{B}$ and an n-dimensional response vector represented by the random variable $\mathscr{Y}$. (We use upper-case "script" characters to denote random variables. The corresponding lower-case upright letter denotes a particular value of the random variable.) We observe the value, $\mathbf{y}$, of $\mathscr{Y}$. We do not observe the value, $\mathbf{b}$, of $\mathscr{B}$.

When formulating the model we describe the unconditional distribution of $\mathscr{B}$ and the conditional distribution, $(\mathscr{Y}|\mathscr{B} = \mathbf{b})$. The descriptions of the distributions involve the form of the distribution and the values of certain parameters. We use the observed values of the response and the covariates to estimate these parameters and to make inferences about them.

That's the big picture. Now let's make this more concrete by describing a particular, versatile class of mixed models called linear mixed models and by

studying a simple example of such a model. First we describe the data in the example.

1.2 The Dyestuff and Dyestuff2 Data

Models with random effects have been in use for a long time. The first edition of the classic book, *Statistical Methods in Research and Production*, edited by O.L. Davies, was published in 1947 and contained examples of the use of random effects to characterize batch-to-batch variability in chemical processes. The data from one of these examples are available as the `Dyestuff` data in the `lme4` package. In this section we describe and plot these data and introduce a second example, the `Dyestuff2` data, described in Box and Tiao [1973].

1.2.1 The Dyestuff *Data*

The `Dyestuff` data are described in Davies and Goldsmith [1972, Table˜6.3, p.˜131], the fourth edition of the book mentioned above, as coming from

> an investigation to find out how much the variation from batch to batch in the quality of an intermediate product (H-acid) contributes to the variation in the yield of the dyestuff (Naphthalene Black 12B) made from it. In the experiment six samples of the intermediate, representing different batches of works manufacture, were obtained, and five preparations of the dyestuff were made in the laboratory from each sample. The equivalent yield of each preparation as grams of standard colour was determined by dye-trial.

To access these data within R we must first attach the `lme4` package to our session using

```
> library(lme4)
```

Note that the ">" symbol in the line shown is the prompt in R and not part of what the user types. The `lme4` package must be attached before any of the data sets or functions in the package can be used. If typing this line results in an error report stating that there is no package by this name then you must first install the package.

In what follows, we will assume that the `lme4` package has been installed and that it has been attached to the R session before any of the code shown has been run.

The `str` function in R provides a concise description of the structure of the data

```
> str(Dyestuff)
```

```
'data.frame':        30 obs. of  2 variables:
 $ Batch: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 2 2 2 2 2 ...
 $ Yield: num  1545 1440 1440 1520 1580 ...
```

from which we see that it consists of 30 observations of the `Yield`, the response variable, and of the covariate, `Batch`, which is a categorical variable stored as a `factor` object. If the labels for the factor levels are arbitrary, as they are here, we will use letters instead of numbers for the labels. That is, we label the batches as `"A"` through `"F"` rather than `"1"` through `"6"`. When the labels are letters it is clear that the variable is categorical. When the labels are numbers a categorical covariate can be mistaken for a numeric covariate, with unintended consequences.

It is a good practice to apply `str` to any data frame the first time you work with it and to check carefully that any categorical variables are indeed represented as factors.

The data in a data frame are viewed as a table with columns corresponding to variables and rows to observations. The functions `head` and `tail` print the first or last few rows (the default value of "few" happens to be 6 but we can specify another value if we so choose)

```
> head(Dyestuff)
```

```
  Batch Yield
1     A  1545
2     A  1440
3     A  1440
4     A  1520
5     A  1580
6     B  1540
```

or we could ask for a `summary` of the data

```
> summary(Dyestuff)
```

```
 Batch     Yield
 A:5   Min.   :1440
 B:5   1st Qu.:1469
 C:5   Median :1530
 D:5   Mean   :1528
 E:5   3rd Qu.:1575
 F:5   Max.   :1635
```

Although the `summary` does show us an important property of the data, namely that there are exactly 5 observations on each batch — a property that we will describe by saying that the data are *balanced* with respect to `Batch` — we usually learn much more about the structure of such data from plots like Fig.˜1.1 than we do from numerical summaries.

In Fig.˜1.1 we can see that there is considerable variability in yield, even for preparations from the same batch, but there is also noticeable batch-to-batch variability. For example, four of the five preparations from batch F provided lower yields than did any of the preparations from batches C and E.

This plot, and essentially all the other plots in this book, were created using Deepayan Sarkar's `lattice` package for R. In Sarkar [2008] he describes

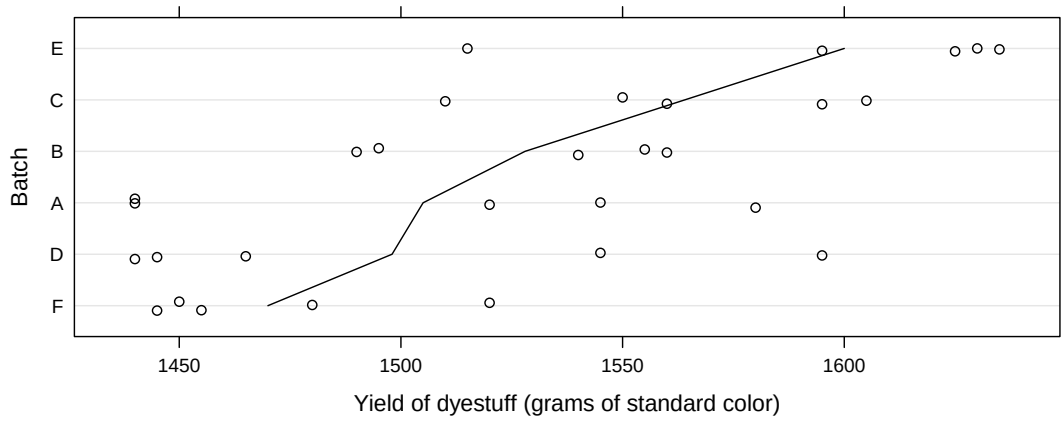

Fig. 1.1 Yield of dyestuff (Napthalene Black 12B) for 5 preparations from each of 6 batches of an intermediate product (H-acid). The line joins the mean yields from the batches, which have been ordered by increasing mean yield. The vertical positions are "jittered" slightly to avoid over-plotting. Notice that the lowest yield for batch A was observed for two distinct preparations from that batch.

how one would create such a plot. Because this book was created using Sweave [Leisch, 2002], the exact code used to create the plot, as well as the code for all the other figures and calculations in the book, is available on the web site for the book. In Sect.˜**??** we review some of the principles of lattice graphics, such as reordering the levels of the `Batch` factor by increasing mean response, that enhance the informativeness of the plot. At this point we will concentrate on the information conveyed by the plot and not on how the plot is created.

In Sect.˜1.3.1 we will use mixed models to quantify the variability in yield between batches. For the time being let us just note that the particular batches used in this experiment are a selection or sample from the set of all batches that we wish to consider. Furthermore, the extent to which one particular batch tends to increase or decrease the mean yield of the process — in other words, the "effect" of that particular batch on the yield — is not as interesting to us as is the extent of the variability between batches. For the purposes of designing, monitoring and controlling a process we want to predict the yield from future batches, taking into account the batch-to-batch variability and the within-batch variability. Being able to estimate the extent to which a particular batch in the past increased or decreased the yield is not usually an important goal for us. We will model the effects of the batches as random effects rather than as fixed-effects parameters.

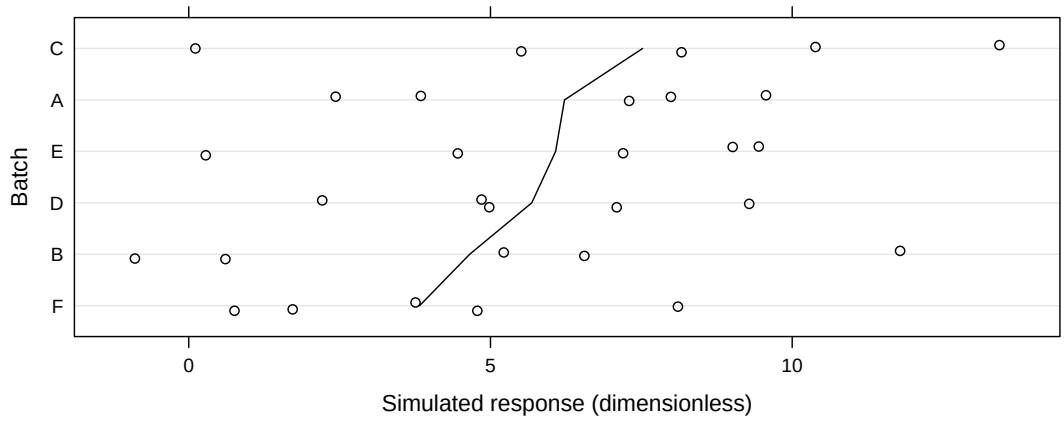

Fig. 1.2 Simulated data presented in Box and Tiao [1973] with a structure similar to that of the `Dyestuff` data. These data represent a case where the batch-to-batch variability is small relative to the within-batch variability.

1.2.2 The `Dyestuff2` *Data*

The `Dyestuff2` data are simulated data presented in Box and Tiao [1973, Table 5.1.4, p. 247] where the authors state

> These data had to be constructed for although examples of this sort undoubtedly occur in practice they seem to be rarely published.

The structure and summary

```
> str(Dyestuff2)

'data.frame':        30 obs. of  2 variables:
 $ Batch: Factor w/ 6 levels "A","B","C","D",..: 1 1 1 1 1 2 2 2 2 2 ...
 $ Yield: num  7.3 3.85 2.43 9.57 7.99 ...

> summary(Dyestuff2)

 Batch     Yield
 A:5   Min.   :-0.892
 B:5   1st Qu.: 2.765
 C:5   Median : 5.365
 D:5   Mean   : 5.666
 E:5   3rd Qu.: 8.151
 F:5   Max.   :13.434
```

are intentionally similar to those of the `Dyestuff` data. As can be seen in Fig.˜1.2 the batch-to-batch variability in these data is small compared to the within-batch variability. In some approaches to mixed models it can be difficult to fit models to such data. Paradoxically, small "variance components" can be more difficult to estimate than large variance components.

The methods we will present are not compromised when estimating small variance components.

1.3 Fitting Linear Mixed Models

Before we formally define a linear mixed model, let's go ahead and fit models to these data sets using `lmer`. Like most model-fitting functions in R, `lmer` takes, as its first two arguments, a *formula* specifying the model and the *data* with which to evaluate the formula. This second argument, `data`, is optional but recommended. It is usually the name of a data frame, such as those we examined in the last section. Throughout this book all model specifications will be given in this formula/data format.

We will explain the structure of the formula after we have considered an example.

1.3.1 A Model For the `Dyestuff` *Data*

We fit a model to the `Dyestuff` data allowing for an overall level of the `Yield` and for an additive random effect for each level of `Batch`

```
> fm01 <- lmer(Yield ~ 1 + (1|Batch), Dyestuff)
> print(fm01)

Linear mixed model fit by REML ['merMod']
Formula: Yield ~ 1 + (1 | Batch)
   Data: Dyestuff
REML criterion at convergence: 319.6543
Random effects:
 Groups   Name        Variance Std.Dev.
 Batch    (Intercept) 1764     42.00
 Residual             2451     49.51
Number of obs: 30, groups: Batch, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1527.50      19.38    78.8
```

In the first line we call the `lmer` function to fit a model with formula

```
Yield ~ 1 + (1 | Batch)
```

applied to the `Dyestuff` data and assign the result to the name `fm01`. (The name is arbitrary. I happen to use names that start with `fm`, indicating "fitted model".)

As is customary in R, there is no output shown after this assignment. We have simply saved the fitted model as an object named `fm01`. In the second line we display some information about the fitted model by applying `print` to `fm01`. In later examples we will condense these two steps into one but here it helps to emphasize that we save the result of fitting a model then apply various *extractor* functions to the fitted model to get a brief summary of the model fit or to obtain the values of some of the estimated quantities.

1.3.1.1 Details of the Printed Display

The printed display of a model fit with `lmer` has four major sections: a description of the model that was fit, some statistics characterizing the model fit, a summary of properties of the random effects and a summary of the fixed-effects parameter estimates. We consider each of these sections in turn.

The description section states that this is a linear mixed model in which the parameters have been estimated as those that minimize the REML criterion (described in Sect.˜5.5). The `formula` and `data` arguments are displayed for later reference. If other, optional arguments affecting the fit, such as a `subset` specification, were used, they too will be displayed here.

For models fit by the REML criterion the only statistic describing the model fit is the value of the REML criterion itself. An alternative set of parameter estimates, the maximum likelihood estimates (MLEs), are obtained by specifying the optional argument `REML=FALSE`.

```
> (fm01ML <- lmer(Yield ~ 1 + (1|Batch), Dyestuff, REML=FALSE))

Linear mixed model fit by maximum likelihood ['merMod']
Formula: Yield ~ 1 + (1 | Batch)
   Data: Dyestuff
      AIC       BIC    logLik  deviance
 333.3271  337.5307 -163.6635  327.3271
Random effects:
 Groups   Name        Variance Std.Dev.
 Batch    (Intercept) 1388     37.26
 Residual             2451     49.51
Number of obs: 30, groups: Batch, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)  1527.50      17.69   86.33
```

(Notice that this code fragment also illustrates a way to condense the assignment and the display of the fitted model into a single step. The redundant set of parentheses surrounding the assignment causes the result of the assignment to be displayed. We will use this device often in what follows.)

The display of a model fit by maximum likelihood (ML) provides several other model-fit statistics such as Akaike's Information Criterion (`AIC`)˜[Sakamoto et˜al., 1986], Schwarz's Bayesian Information Criterion (`BIC`)˜[Schwarz, 1978], the log-likelihood (`logLik`) at the parameter estimates, and the deviance (negative twice the log-likelihood) at the parameter estimates. These are all statistics related to the model fit and are used to compare different models fit to the same data.

At this point the important thing to note is that the default estimation criterion is the REML criterion. Generally the REML estimates of variance components are preferred to the ML estimates. When comparing models, however, we will use *likelihood ratio tests*, for which the test statistic is the difference in the deviance of the fitted models. (Recall that the deviance

is negative twice the log-likelihood, hence a ratio of likelihoods corresponds to the difference in deviance values.) Therefore, when building and assessing models we will often use maximum likelihood fits. As described in Sect.˜2.2.4, the function that performs a likelihood-ratio test will accept models fit by REML but it adds an extra step of refitting the models to obtain the maximum likelihood estimates (MLEs).

The third section is the table of estimates of parameters associated with the random effects. There are two sources of variability in the model we have fit, a batch-to-batch variability in the level of the response and the residual or per-observation variability — also called the within-batch variability. The name "residual" is used in statistical modeling to denote the part of the variability that cannot be explained or modeled with the other terms. It is the variation in the observed data that is "left over" after we have determined the estimates of the parameters in the other parts of the model.

Some of the variability in the response is associated with the fixed-effects terms. In this model there is only one such term, labeled as the `(Intercept)`. The name "intercept", which is better suited to models based on straight lines written in a slope/intercept form, should be understood to represent an overall "typical" or mean level of the response in this case. (In case you are wondering about the parentheses around the name, they are included so that you can't accidentally create a variable with a name that conflicts with this name.) The line labeled `Batch` in the random effects table shows that the random effects added to the `(Intercept)` term, one for each level of the `Batch` factor, are modeled as random variables whose unconditional variance is estimated as 1764.05 g^2 in the REML fit and as 1388.33 g^2 in the ML fit. The corresponding standard deviations are 42.00 g for the REML fit and 37.26 g for the ML fit.

Note that the last column in the random effects summary table is the estimate of the variability expressed as a standard deviation rather than as a variance. These are provided because it is usually easier to visualize standard deviations, which are on the scale of the response, than it is to visualize the magnitude of a variance. The values in this column are a simple re-expression (the square root) of the estimated variances. Do not confuse them with the standard errors of the variance estimators, which are not given here. In Sect.˜1.5 we explain why we do not provide standard errors of variance estimates.

The line labeled `Residual` in this table gives the estimate of the variance of the residuals (also in g^2) and its corresponding standard deviation. For the REML fit the estimated standard deviation of the residuals is 49.51 g and for the ML fit it is also 49.51 g. (Generally these estimates do not need to be equal. They happen to be equal in this case because of the simple model form and the balanced data set.)

The last line in the random effects table states the number of observations to which the model was fit and the number of levels of any "grouping factors" for the random effects. In this case we have a single random effects term,

`(1|Batch)`, in the model formula and the grouping factor for that term is `Batch`. There will be a total of six random effects, one for each level of `Batch`.

The final part of the printed display gives the estimates and standard errors of any fixed-effects parameters in the model. The only fixed-effects term in the model formula is the `1`, denoting a constant which, as explained above, is labeled as `(Intercept)`. For both the REML and the ML estimation criterion the estimate of this parameter is 1527.5 g (equality is again a consequence of the simple model and balanced data set). The standard error of the intercept estimate is 19.38 g for the REML fit and 17.69 g for the ML fit.

1.3.2 A Model For the `Dyestuff2` *Data*

Fitting a similar model to the `Dyestuff2` data produces an estimate $\widehat{\sigma}_1 = 0$ in both the REML

```
> (fm02 <- lmer(Yield ~ 1 + (1|Batch), Dyestuff2))

Linear mixed model fit by REML ['merMod']
Formula: Yield ~ 1 + (1 | Batch)
   Data: Dyestuff2
REML criterion at convergence: 161.8283
Random effects:
 Groups   Name        Variance Std.Dev.
 Batch    (Intercept)  0.00    0.000
 Residual             13.81    3.716
Number of obs: 30, groups: Batch, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)   5.6656     0.6784   8.352
```

and the ML fits.

```
> (fm02ML <- update(fm02, REML=FALSE))

Linear mixed model fit by maximum likelihood ['merMod']
Formula: Yield ~ 1 + (1 | Batch)
   Data: Dyestuff2
     AIC      BIC   logLik deviance
168.8730 173.0766 -81.4365 162.8730
Random effects:
 Groups   Name        Variance Std.Dev.
 Batch    (Intercept)  0.00    0.000
 Residual             13.35    3.653
Number of obs: 30, groups: Batch, 6

Fixed effects:
            Estimate Std. Error t value
(Intercept)    5.666      0.667   8.494
```

(Note the use of the `update` function to re-fit a model changing some of the arguments. In a case like this, where the call to fit the original model is not very complicated, the use of `update` is not that much simpler than repeating the original call to `lmer` with extra arguments. For complicated model fits it can be.)

An estimate of 0 for σ_1 does not mean that there is no variation between the groups. Indeed Fig. ˜1.2 shows that there is some small amount of variability between the groups. The estimate, $\widehat{\sigma}_1 = 0$, simply indicates that the level of "between-group" variability is not sufficient to warrant incorporating random effects in the model.

The important point to take away from this example is that we must allow for the estimates of variance components to be zero. We describe such a model as being degenerate, in the sense that it corresponds to a linear model in which we have removed the random effects associated with `Batch`. Degenerate models can and do occur in practice. Even when the final fitted model is not degenerate, we must allow for such models when determining the parameter estimates through numerical optimization.

To reiterate, the model `fm02` corresponds to the linear model

```
> summary(fm02a <- lm(Yield ~ 1, Dyestuff2))

Call:
lm(formula = Yield ~ 1, data = Dyestuff2)
Residuals:
    Min      1Q  Median      3Q     Max
-6.5576 -2.9006 -0.3006  2.4854  7.7684

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.6656     0.6784   8.352 3.32e-09

Residual standard error: 3.716 on 29 degrees of freedom
```

because the random effects are inert, in the sense that they have a variance of zero, and hence can be removed.

Notice that the estimate of σ from the linear model (called the `Residual standard error` in the summary) corresponds to the estimate in the REML fit (`fm02`) but not that from the ML fit (`fm02ML`). The fact that the REML estimates of variance components in mixed models generalize the estimate of the variance used in linear models, in the sense that these estimates coincide in the degenerate case, is part of the motivation for the use of the REML criterion for fitting mixed-effects models.

1.3.3 Further Assessment of the Fitted Models

The parameter estimates in a statistical model represent our "best guess" at the unknown values of the model parameters and, as such, are important

results in statistical modeling. However, they are not the whole story. Statistical models characterize the variability in the data and we must assess the effect of this variability on the parameter estimates and on the precision of predictions made from the model.

In Sect.˜1.5 we introduce a method of assessing variability in parameter estimates using the "profiled deviance" and in Sect.˜1.6 we show methods of characterizing the conditional distribution of the random effects given the data. Before we get to these sections, however, we should state in some detail the probability model for linear mixed-effects and establish some definitions and notation. In particular, before we can discuss profiling the deviance, we should define the deviance. We do that in the next section.

1.4 The Linear Mixed-effects Probability Model

In explaining some of parameter estimates related to the random effects we have used terms such as "unconditional distribution" from the theory of probability. Before proceeding further we clarify the linear mixed-effects probability model and define several terms and concepts that will be used throughout the book. Readers who are more interested in practical results than in the statistical theory should feel free to skip this section.

1.4.1 Definitions and Results

In this section we provide some definitions and formulas without derivation and with minimal explanation, so that we can use these terms in what follows. In Chap.˜5 we revisit these definitions providing derivations and more explanation.

As mentioned in Sect.˜1.1, a mixed model incorporates two random variables: $\mathscr{B}$, the q-dimensional vector of random effects, and $\mathscr{Y}$, the n-dimensional response vector. In a linear mixed model the unconditional distribution of $\mathscr{B}$ and the conditional distribution, $(\mathscr{Y}|\mathscr{B}=\mathbf{b})$, are both multivariate Gaussian (or "normal") distributions,

$$\begin{aligned}(\mathscr{Y}|\mathscr{B}=\mathbf{b}) &\sim \mathscr{N}(\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b},\sigma^2\mathbf{I})\\ \mathscr{B} &\sim \mathscr{N}(\mathbf{0},\Sigma_\theta).\end{aligned} \tag{1.1}$$

The *conditional mean* of $\mathscr{Y}$, given $\mathscr{B}=\mathbf{b}$, is the *linear predictor*, $\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\mathbf{b}$, which depends on the p-dimensional *fixed-effects parameter*, $\boldsymbol{\beta}$, and on $\mathbf{b}$. The *model matrices*, $\mathbf{X}$ and $\mathbf{Z}$, of dimension $n\times p$ and $n\times q$, respectively, are determined from the formula for the model and the values of covariates.

Although the matrix $\mathbf{Z}$ can be large (i.e. both n and q can be large), it is sparse (i.e. most of the elements in the matrix are zero).

The *relative covariance factor*, Λ_θ, is a $q \times q$ matrix, depending on the *variance-component parameter*, θ, and generating the symmetric $q \times q$ variance-covariance matrix, Σ_θ, according to

$$\Sigma_\theta = \sigma^2 \Lambda_\theta \Lambda_\theta^\mathsf{T}. \tag{1.2}$$

The *spherical random effects*, $\mathscr{U} \sim \mathscr{N}(\mathbf{0}, \sigma^2 \mathbf{I}_q)$, determine $\mathscr{B}$ according to

$$\mathscr{B} = \Lambda_\theta \mathscr{U}.$$

The *penalized residual sum of squares* (PRSS),

$$r^2(\theta, \beta, \mathbf{u}) = \|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\Lambda_\theta \mathbf{u}\|^2 + \|\mathbf{u}\|^2, \tag{1.3}$$

is the sum of the residual sum of squares, measuring fidelity of the model to the data, and a penalty on the size of $\mathbf{u}$, measuring the complexity of the model. Minimizing r^2 with respect to $\mathbf{u}$,

$$r^2_{\beta,\theta} = \min_{\mathbf{u}} \left\{ \|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\Lambda_\theta \mathbf{u}\|^2 + \|\mathbf{u}\|^2 \right\} \tag{1.4}$$

is a direct (i.e. non-iterative) computation during which we calculate the *sparse Cholesky factor*, $\mathbf{L}_\theta$, which is a lower triangular $q \times q$ matrix satisfying

$$\mathbf{L}_\theta \mathbf{L}_\theta^\mathsf{T} = \Lambda_\theta^\mathsf{T} \mathbf{Z}^\mathsf{T} \mathbf{Z} \Lambda_\theta + \mathbf{I}_q. \tag{1.5}$$

where $\mathbf{I}_q$ is the $q \times q$ *identity matrix*.

The *deviance* (negative twice the log-likelihood) of the parameters, given the data, $\mathbf{y}$, is

$$d(\theta, \beta, \sigma | \mathbf{y}) = n \log(2\pi\sigma^2) + \log(|\mathbf{L}_\theta|^2) + \frac{r^2_{\beta,\theta}}{\sigma^2}. \tag{1.6}$$

where $|\mathbf{L}_\theta|$ denotes the *determinant* of $\mathbf{L}_\theta$. Because $\mathbf{L}_\theta$ is triangular, its determinant is the product of its diagonal elements.

Because the conditional mean, $\mu_{\mathscr{Y}|\mathscr{B}=\mathbf{b}} = \mathbf{X}\beta + \mathbf{Z}\Lambda_\theta \mathbf{u}$, is a linear function of both β and $\mathbf{u}$, minimization of the PRSS with respect to both β and $\mathbf{u}$ to produce

$$r^2_\theta = \min_{\beta,\mathbf{u}} \left\{ \|\mathbf{y} - \mathbf{X}\beta - \mathbf{Z}\Lambda_\theta \mathbf{u}\|^2 + \|\mathbf{u}\|^2 \right\} \tag{1.7}$$

is also a direct calculation. The values of $\mathbf{u}$ and β that provide this minimum are called, respectively, the *conditional mode*, $\tilde{\mathbf{u}}_\theta$, of the spherical random effects and the conditional estimate, $\widehat{\beta}_\theta$, of the fixed effects. At the conditional estimate of the fixed effects the deviance is

$$d(\theta,\widehat{\beta}_\theta,\sigma|\mathbf{y}) = n\log(2\pi\sigma^2) + \log(|\mathbf{L}_\theta|^2) + \frac{r_\theta^2}{\sigma^2}. \tag{1.8}$$

Minimizing this expression with respect to σ^2 produces the conditional estimate

$$\widehat{\sigma^2}_\theta = \frac{r_\theta^2}{n} \tag{1.9}$$

which provides the *profiled deviance*,

$$\tilde{d}(\theta|\mathbf{y}) = d(\theta,\widehat{\beta}_\theta,\widehat{\sigma}_\theta|\mathbf{y}) = \log(|\mathbf{L}_\theta|^2) + n\left[1+\log\left(\frac{2\pi r_\theta^2}{n}\right)\right], \tag{1.10}$$

a function of θ alone.

The MLE of θ, written $\widehat{\theta}$, is the value that minimizes the profiled deviance˜(1.10). We determine this value by numerical optimization. In the process of evaluating $\tilde{d}(\widehat{\theta}|\mathbf{y})$ we determine $\widehat{\beta} = \widehat{\beta}_{\widehat{\theta}}$, $\tilde{\mathbf{u}}_{\widehat{\theta}}$ and $r_{\widehat{\theta}}^2$, from which we can evaluate $\widehat{\sigma} = \sqrt{r_{\widehat{\theta}}^2/n}$.

The elements of the conditional mode of $\mathscr{B}$, evaluated at the parameter estimates,

$$\tilde{b}_{\widehat{\theta}} = \Lambda_{\widehat{\theta}}\tilde{u}_{\widehat{\theta}} \tag{1.11}$$

are sometimes called the *best linear unbiased predictors* or BLUPs of the random effects. Although it has an appealing acronym, I don't find the term particularly instructive (what is a "linear unbiased predictor" and in what sense are these the "best"?) and prefer the term "conditional mode", which is explained in Sect.˜1.6.

1.4.2 Matrices and Vectors in the Fitted Model Object

The optional argument, `verbose=1`, in a call to `lmer` produces output showing the progress of the iterative optimization of $\tilde{d}(\theta|\mathbf{y})$.

```
> fm01ML <- lmer(Yield ~ 1|Batch, Dyestuff, REML=FALSE, verbose=1)

npt = 3 , n =  1
rhobeg =  0.2 , rhoend =  2e-07
   0.020:   4:      327.347;0.800000
  0.0020:   6:      327.328;0.764659
 0.00020:   9:      327.327;0.752808
 2.0e-05:  10:      327.327;0.752583
 2.0e-06:  12:      327.327;0.752583
 2.0e-07:  14:      327.327;0.752581
At return
 17:     327.32706: 0.752581
```

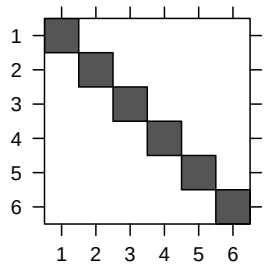

Fig. 1.3 Image of the relative covariance factor, $\Lambda_{\hat{\theta}}$ for model `fm01ML`. The non-zero elements are shown as darkened squares. The zero elements are blank.

The algorithm converges after 17 function evaluations to a profiled deviance of 327.32706 at θ =0.752581. In this model the scalar parameter θ is the ratio σ_1/σ.

The actual values of many of the matrices and vectors defined above are available as *slots* of the fitted model object. In general the object returned by a model-fitting function from the `lme4` package has three upper-level slots: `re`, consisting of information related to the random effects, `fe`, consisting of information related to the fixed effects, and `resp`, consisting of information related to the response.

In practice we rarely access the slots of a fitted model object directly, preferring instead to use some of the *accessor functions* that, as the name implies, provide access to some of the properties of the model. However, in this section we will access some of the slots directly for the purposes of illustration.

For example, $\Lambda_{\hat{\theta}}$ is stored as

```
> fm01ML@re@Lambda

6 x 6 sparse Matrix of class "dgCMatrix"
[1,] 0.7525807 .         .         .         .         .
[2,] .         0.7525807 .         .         .         .
[3,] .         .         0.7525807 .         .         .
[4,] .         .         .         0.7525807 .         .
[5,] .         .         .         .         0.7525807 .
[6,] .         .         .         .         .         0.7525807
```

Often we will show the structure of sparse matrices as an image (Fig.˜1.3). Especially for large sparse matrices, the image conveys the structure more compactly than does the printed representation.

In this simple model $\Lambda = \theta \mathbf{I}_6$ is a multiple of the identity matrix and the 30×6 model matrix $\mathbf{Z}$, whose transpose is shown in Fig.˜1.4, consists of the indicator columns for `Batch`. Because the data are balanced with respect to `Batch`, the Cholesky factor, $\mathbf{L}$ is also a multiple of the identity (use `image(fm01ML@re@L)` to check if you wish). The vector $\mathbf{u}$ is available in `fm01ML@re@u`. The vector β and the model matrix $\mathbf{X}$ are in `fm01ML@fe`.

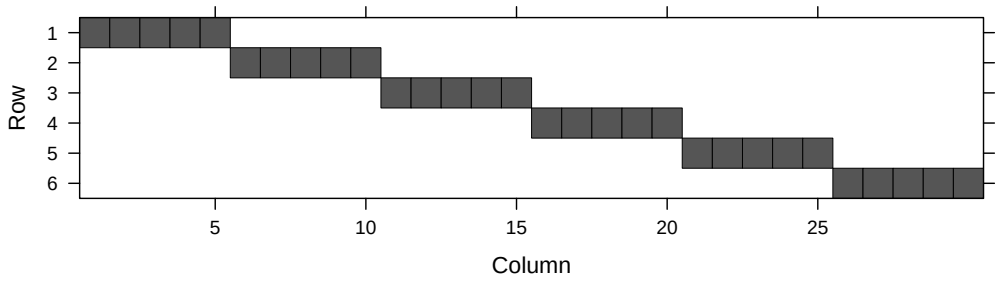

Fig. 1.4 Image of the transpose of the random-effects model matrix, **Z**, for model `fm01`. The non-zero elements, which are all unity, are shown as darkened squares. The zero elements are blank.

1.5 Assessing the Variability of the Parameter Estimates

In this section we show how to create a *profile deviance* object from a fitted linear mixed model and how to use this object to evaluate confidence intervals on the parameters. We also discuss the construction and interpretation of *profile zeta* plots for the parameters. In Chap.˜**??** we discuss the use of the deviance profiles to produce likelihood contours for pairs of parameters.

1.5.1 Confidence Intervals on the Parameters

The mixed-effects model fit as `fm01` or `fm01ML` has three parameters for which we obtained estimates. These parameters are σ_1, the standard deviation of the random effects, σ, the standard deviation of the residual or "per-observation" noise term and β_0, the fixed-effects parameter that is labeled as `(Intercept)`.

The `profile` function systematically varies the parameters in a model, assessing the best possible fit that can be obtained with one parameter fixed at a specific value and comparing this fit to the *globally optimal fit*, which is the original model fit that allowed all the parameters to vary. The models are compared according to the change in the deviance, which is the *likelihood ratio test* (LRT) statistic. We apply a *signed square root* transformation to this statistic and plot the resulting function, called ζ, versus the parameter value. A ζ value can be compared to the quantiles of the *standard normal distribution*, $\mathscr{Z} \sim \mathscr{N}(0,1)$. For example, a 95% profile deviance confidence interval on the parameter consists of those values for which $-1.960 < \zeta < 1.960$.

Because the process of profiling a fitted model, which involves re-fitting the model many times, can be computationally intensive, one should exercise caution with complex models fit to very large data sets. Because the statistic of interest is a likelihood ratio, the model is re-fit according to the maximum

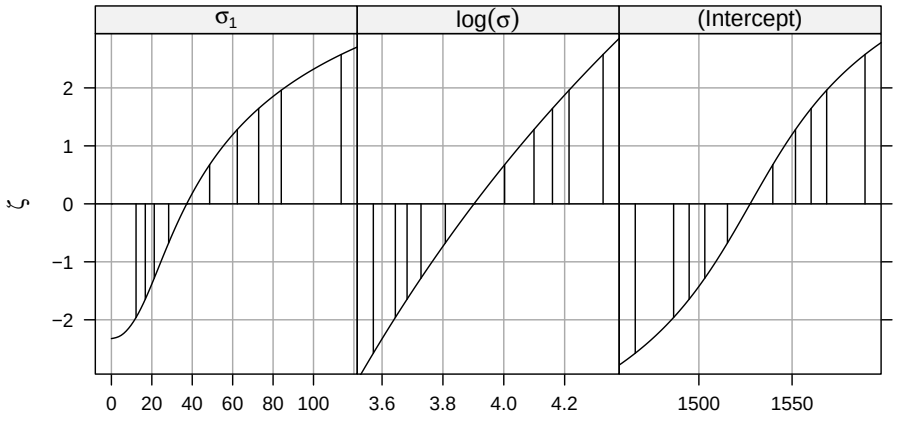

Fig. 1.5 Signed square root, ζ, of the likelihood ratio test statistic for each of the parameters in model `fm01ML`. The vertical lines are the endpoints of 50%, 80%, 90%, 95% and 99% confidence intervals derived from this test statistic.

likelihood criterion, even if the original fit is a REML fit. Thus, there is a slight advantage in starting with an ML fit.

```
> pr01 <- profile(fm01ML)
```

Plots of ζ versus the parameter being profiled (Fig.˜1.5) are obtained with

```
> xyplot(pr01, aspect = 1.3)
```

We will refer to such plots as *profile zeta* plots. I usually adjust the aspect ratio of the panels in profile zeta plots to, say, `aspect = 1.3` and frequently set the layout so the panels form a single row (`layout = c(3,1)`, in this case).

The vertical lines in the panels delimit the 50%, 80%, 90%, 95% and 99% confidence intervals, when these intervals can be calculated. Numerical values of the endpoints are returned by the `confint` extractor.

```
> confint(pr01)
```

```
                   2.5 %      97.5 %
.sig01         12.197461   84.063361
.lsig           3.643624    4.214461
(Intercept) 1486.451506 1568.548494
```

By default the 95% confidence interval is returned. The optional argument, `level`, is used to obtain other confidence levels.

```
> confint(pr01, level = 0.99)
```

```
                   0.5 %      99.5 %
.sig01                NA  113.690280
.lsig           3.571290    4.326337
(Intercept) 1465.872875 1589.127125
```

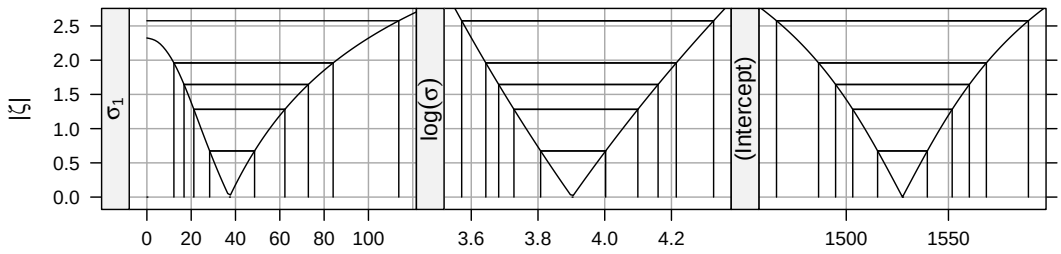

Fig. 1.6 Profiled deviance, on the scale $|\zeta|$, the square root of the change in the deviance, for each of the parameters in model `fm01ML`. The intervals shown are 50%, 80%, 90%, 95% and 99% confidence intervals based on the profile likelihood.

Notice that the lower bound on the 99% confidence interval for σ_1 is not defined. Also notice that we profile $\log(\sigma)$ instead of σ, the residual standard deviation.

A plot of $|\zeta|$, the absolute value of ζ, versus the parameter (Fig.˜1.6), obtained by adding the optional argument `absVal = TRUE` to the call to `xyplot`, can be more effective for visualizing the confidence intervals.

1.5.2 Interpreting the Profile Zeta Plot

A profile zeta plot, such as Fig.˜1.5, shows us the sensitivity of the model fit to changes in the value of particular parameters. Although this is not quite the same as describing the distribution of an estimator, it is a similar idea and we will use some of the terminology from distributions when describing these plots. Essentially we view the patterns in the plots as we would those in a normal probability plot of data values or of residuals from a model.

Ideally the profile zeta plot will be close to a straight line over the region of interest, in which case we can perform reliable statistical inference based on the parameter's estimate, its standard error and quantiles of the standard normal distribution. We will describe such a situation as providing a good normal approximation for inference. The common practice of quoting a parameter estimate and its standard error assumes that this is always the case.

In Fig.˜1.5 the profile zeta plot for $\log(\sigma)$ is reasonably straight so $\log(\sigma)$ has a good normal approximation. But this does not mean that there is a good normal approximation for σ^2 or even for σ. As shown in Fig.˜1.7 the profile zeta plot for $\log(\sigma)$ is slightly skewed, that for σ is moderately skewed and the profile zeta plot for σ^2 is highly skewed. Deviance-based confidence intervals on σ^2 are quite asymmetric, of the form "estimate minus a little, plus a lot".

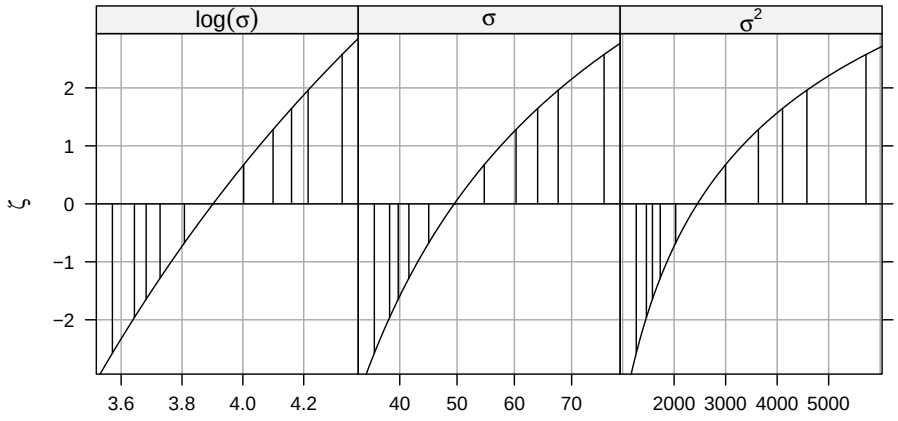

Fig. 1.7 Signed square root, ζ, of the likelihood ratio test statistic as a function of $\log(\sigma)$, of σ and of σ^2. The vertical lines are the endpoints of 50%, 80%, 90%, 95% and 99% confidence intervals.

This should not come as a surprise to anyone who learned in an introductory statistics course that, given a random sample of data assumed to come from a Gaussian distribution, we use a χ^2 distribution, which can be quite skewed, to form a confidence interval on σ^2. Yet somehow there is a widespread belief that the distribution of variance estimators in much more complex situations should be well approximated by a normal distribution. It is nonsensical to believe that. In most cases summarizing the precision of a variance component estimate by giving an approximate standard error is woefully inadequate.

The pattern in the profile plot for β_0 is sigmoidal (i.e. an elongated "S"-shape). The pattern is symmetric about the estimate but curved in such a way that the profile-based confidence intervals are wider than those based on a normal approximation. We characterize this pattern as symmetric but over-dispersed (relative to a normal distribution). Again, this pattern is not unexpected. Estimators of the coefficients in a linear model without random effects have a distribution which is a scaled Student's T distribution. That is, they follow a symmetric distribution that is over-dispersed relative to the normal.

The pattern in the profile zeta plot for σ_1 is more complex. Fig.˜1.8 shows the profile zeta plot on the scale of $\log(\sigma_1)$, σ_1 and σ_1^2. Notice that the profile zeta plot for $\log(\sigma_1)$ is very close to linear to the right of the estimate but flattens out on the left. That is, σ_1 behaves like σ in that its profile zeta plot is more-or-less a straight line on the logarithmic scale, except when σ_1 is close to zero. The model loses sensitivity to values of σ_1 that are close to zero. If, as in this case, zero is within the "region of interest" then we should expect that the profile zeta plot will flatten out on the left hand side.

Notice that the profile zeta plot of σ_1^2 in Fig.˜1.8 is dramatically skewed. If reporting the estimate, $\widehat{\sigma^2}_1$, and its standard error, as many statistical

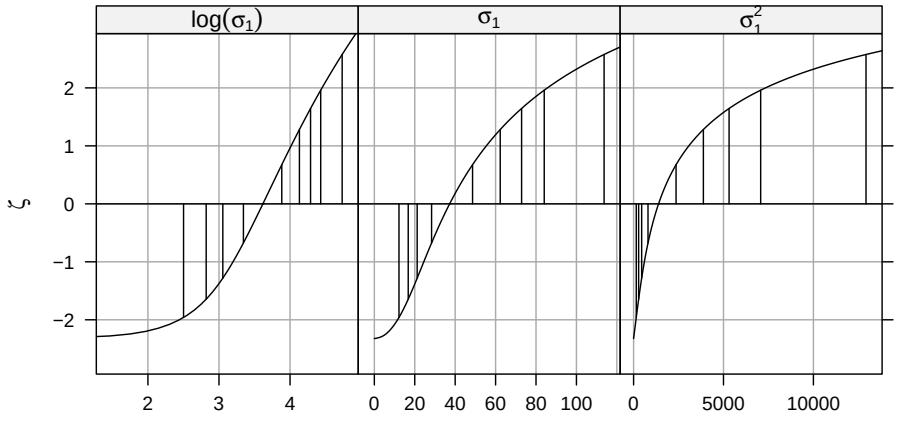

Fig. 1.8 Signed square root, ζ, of the likelihood ratio test statistic as a function of $\log(\sigma_1)$, of σ_1 and of σ_1^2. The vertical lines are the endpoints of 50%, 80%, 90%, 95% and 99% confidence intervals.

software packages do, were to be an adequate description of the variability in this estimate then this profile zeta plot should be a straight line. It's nowhere close to being a straight line in this and in many other model fits, which is why we don't report standard errors for variance estimates.

1.5.3 Deriving densities from the profile

In the profile zeta plots we show ζ as a function of a parameter. We can use the function shown there, which we will call the *profile zeta function*, to generate a corresponding distribution by setting the cumulative distribution function (c.d.f) to be $\Phi(\zeta)$ where Φ is the c.d.f. of the standard normal distribution. From this we can derive a density.

This is not quite the same as evaluating the distribution of the estimator of the parameter, which for mixed-effects models can be very difficult, but it gives us a good indication of what the distribution of the estimator would be. Fig.˜1.9, created as

```
> densityplot(pr01)
```

shows the densities corresponding to the profiles in Fig.˜1.5. We see that the density for σ_1 is quite skewed.

If we had plotted the densities corresponding to the profiles of the variance components instead, we would get Fig.˜1.10 which, of course, just accentuates the skewness in the distribution of these variance components.

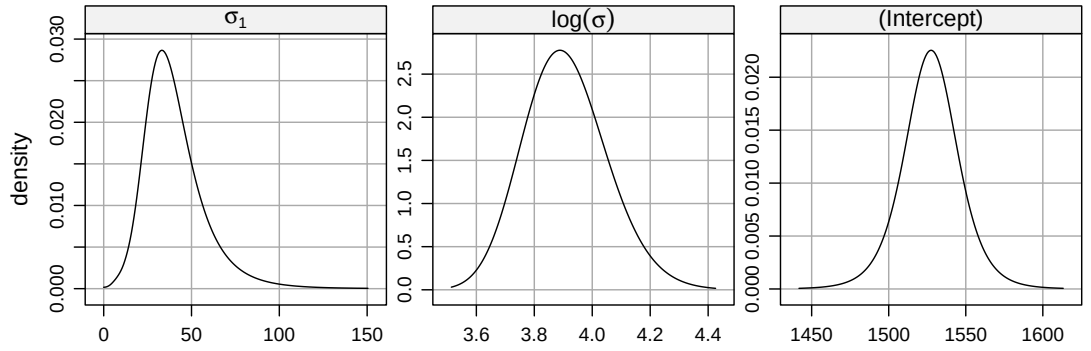

Fig. 1.9 Density plot derived from the profile of model `fm01ML`

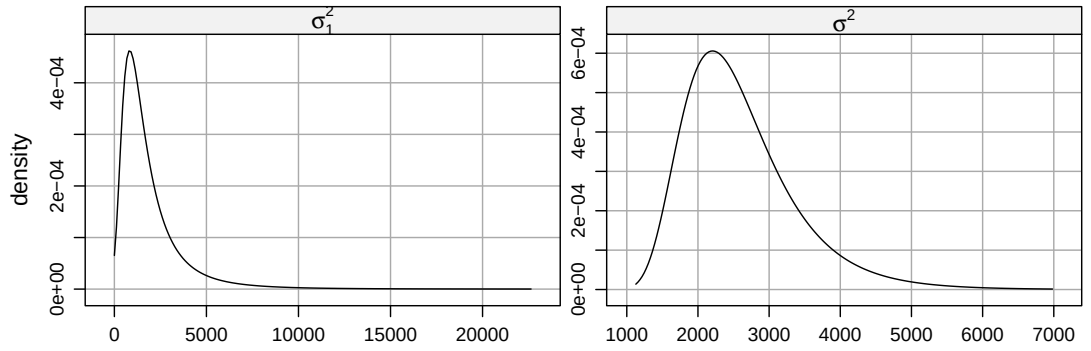

Fig. 1.10 Densities of the variance components, σ_1^2 and σ^2 for model `fm01ML` derived from the profile, `pr01`.

1.6 Assessing the Random Effects

In Sect.˜1.4.1 we mentioned that what are sometimes called the BLUPs (or best linear unbiased predictors) of the random effects, $\mathscr{B}$, are the conditional modes evaluated at the parameter estimates, calculated as $\tilde{b}_{\widehat{\theta}} = \Lambda_{\widehat{\theta}}\tilde{u}_{\widehat{\theta}}$.

These values are often considered as some sort of "estimates" of the random effects. It can be helpful to think of them this way but it can also be misleading. As we have stated, the random effects are not, strictly speaking, parameters—they are unobserved random variables. We don't estimate the random effects in the same sense that we estimate parameters. Instead, we consider the conditional distribution of $\mathscr{B}$ given the observed data, $(\mathscr{B}|\mathscr{Y} = \mathbf{y})$.

Because the unconditional distribution, $\mathscr{B} \sim \mathscr{N}(\mathbf{0}, \Sigma_\theta)$ is continuous, the conditional distribution, $(\mathscr{B}|\mathscr{Y} = \mathbf{y})$ will also be continuous. In general, the mode of a probability density is the point of maximum density, so the phrase "conditional mode" refers to the point at which this conditional density is maximized. Because this definition relates to the probability model, the values

of the parameters are assumed to be known. In practice, of course, we don't know the values of the parameters (if we did there would be no purpose in forming the parameter estimates), so we use the estimated values of the parameters to evaluate the conditional modes.

Those who are familiar with the multivariate Gaussian distribution may recognize that, because both $\mathscr{B}$ and $(\mathscr{Y}|\mathscr{B}=\mathbf{b})$ are multivariate Gaussian, $(\mathscr{B}|\mathscr{Y}=\mathbf{y})$ will also be multivariate Gaussian and the conditional mode will also be the conditional mean of $\mathscr{B}$, given $\mathscr{Y}=\mathbf{y}$. This is the case for a linear mixed model but it does not carry over to other forms of mixed models. In the general case all we can say about $\tilde{\mathbf{u}}$ or $\tilde{\mathbf{b}}$ is that they maximize a conditional density, which is why we use the term "conditional mode" to describe these values. We will only use the term "conditional mean" and the symbol, μ, in reference to $\mathrm{E}(\mathscr{Y}|\mathscr{B}=\mathbf{b})$, which is the conditional mean of $\mathscr{Y}$ given $\mathscr{B}$, and an important part of the formulation of all types of mixed-effects models.

The `ranef` extractor returns the conditional modes.

```
> ranef(fm01ML)

$Batch
  (Intercept)
A  -16.628222
B    0.369516
C   26.974671
D  -21.801446
E   53.579825
F  -42.494344
attr(,"class")
[1] "ranef.mer"
```

Applying `str` to the result of `ranef`

```
> str(ranef(fm01ML))

List of 1
 $ Batch:'data.frame':        6 obs. of  1 variable:
  ..$ (Intercept): num [1:6] -16.628 0.37 26.975 -21.801 53.58 ...
 - attr(*, "class")= chr "ranef.mer"
```

shows that the value is a list of data frames. In this case the list is of length 1 because there is only one random-effects term, `(1|Batch)`, in the model and, hence, only one grouping factor, `Batch`, for the random effects. There is only one column in this data frame because the random-effects term, `(1|Batch)`, is a simple, scalar term.

To make this more explicit, random-effects terms in the model formula are those that contain the vertical bar (`"|"`) character. The `Batch` variable is the grouping factor for the random effects generated by this term. An expression for the grouping factor, usually just the name of a variable, occurs to the right of the vertical bar. If the expression on the left of the vertical bar is `1`, as it is here, we describe the term as a *simple, scalar, random-effects term*. The designation "scalar" means there will be exactly one random effect generated

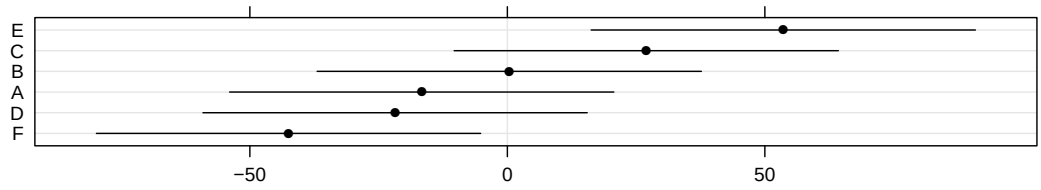

Fig. 1.11 95% prediction intervals on the random effects in `fm01ML`, shown as a dotplot.

for each level of the grouping factor. A simple, scalar term generates a block of indicator columns — the indicators for the grouping factor — in $\mathbf{Z}$. Because there is only one random-effects term in this model and because that term is a simple, scalar term, the model matrix $\mathbf{Z}$ for this model is the indicator matrix for the levels of `Batch`.

In the next chapter we fit models with multiple simple, scalar terms and, in subsequent chapters, we extend random-effects terms beyond simple, scalar terms. When we have only simple, scalar terms in the model, each term has a unique grouping factor and the elements of the list returned by `ranef` can be considered as associated with terms or with grouping factors. In more complex models a particular grouping factor may occur in more than one term, in which case the elements of the list are associated with the grouping factors, not the terms.

Given the data, $\mathbf{y}$, and the parameter estimates, we can evaluate a measure of the dispersion of $(\mathcal{B}|\mathcal{Y} = \mathbf{y})$. In the case of a linear mixed model, this is the conditional standard deviation, from which we can obtain a prediction interval. The `ranef` extractor takes an optional argument, `postVar = TRUE`, which adds these dispersion measures as an attribute of the result. (The name stands for "posterior variance", which is a misnomer that had become established as an argument name before I realized that it wasn't the correct term.)

We can plot these prediction intervals using

```
> dotplot(ranef(fm01ML, postVar = TRUE))
```

(Fig.˜1.11), which provides linear spacing of the levels on the y axis, or using

```
> qqmath(ranef(fm01ML, postVar=TRUE))
```

(Fig.˜1.12), where the intervals are plotted with spacing determined by quantiles of the standard normal distribution.

The dotplot is preferred when there are only a few levels of the grouping factor, as in this case. When there are hundreds or thousands of random effects the `qqmath` form is preferred because it focuses attention on the "important few" at the extremes and de-emphasizes the "trivial many" that are close to zero.

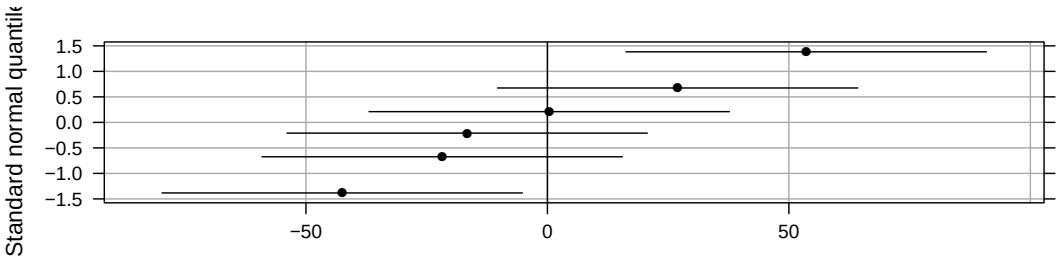

Fig. 1.12 95% prediction intervals on the random effects in `fm01ML` versus quantiles of the standard normal distribution.

1.7 Chapter Summary

A considerable amount of material has been presented in this chapter, especially considering the word "simple" in its title (it's the model that is simple, not the material). A summary may be in order.

A mixed-effects model incorporates fixed-effects parameters and random effects, which are unobserved random variables, $\mathscr{B}$. In a linear mixed model, both the unconditional distribution of $\mathscr{B}$ and the conditional distribution, $(\mathscr{Y}|\mathscr{B}=\mathbf{b})$, are multivariate Gaussian distributions. Furthermore, this conditional distribution is a spherical Gaussian with mean, μ, determined by the linear predictor, $\mathbf{Zb}+\mathbf{X}\beta$. That is,

$$(\mathscr{Y}|\mathscr{B}=\mathbf{b}) \sim \mathscr{N}(\mathbf{Zb}+\mathbf{X}\beta, \sigma^2\mathbf{I}_n).$$

The unconditional distribution of $\mathscr{B}$ has mean $\mathbf{0}$ and a parameterized $q \times q$ variance-covariance matrix, Σ_θ.

In the models we considered in this chapter, Σ_θ, is a simple multiple of the identity matrix, $\mathbf{I}_6$. This matrix is always a multiple of the identity in models with just one random-effects term that is a simple, scalar term. The reason for introducing all the machinery that we did is to allow for more general model specifications.

The maximum likelihood estimates of the parameters are obtained by minimizing the deviance. For linear mixed models we can minimize the profiled deviance, which is a function of θ only, thereby considerably simplifying the optimization problem.

To assess the precision of the parameter estimates, we profile the deviance function with respect to each parameter and apply a signed square root transformation to the likelihood ratio test statistic, producing a profile zeta function for each parameter. These functions provide likelihood-based confidence intervals for the parameters. Profile zeta plots allow us to visually assess the precision of individual parameters. Density plots derived from the profile zeta function provide another way of examining the distribution of the estimators of the parameters.

Prediction intervals from the conditional distribution of the random effects, given the observed data, allow us to assess the precision of the random effects.

Notation

Random Variables

$\mathscr{Y}$ The responses (n-dimensional Gaussian)
$\mathscr{B}$ The random effects on the original scale (q-dimensional Gaussian with mean $\mathbf{0}$)
$\mathscr{U}$ The orthogonal random effects (q-dimensional spherical Gaussian)

Values of these random variables are denoted by the corresponding boldface, lower-case letters: $\mathbf{y}$, $\mathbf{b}$ and $\mathbf{u}$. We observe $\mathbf{y}$. We do not observe $\mathbf{b}$ or $\mathbf{u}$.

Parameters in the Probability Model

β The p-dimension fixed-effects parameter vector.
θ The variance-component parameter vector. Its (unnamed) dimension is typically very small. Dimensions of 1, 2 or 3 are common in practice.
σ The (scalar) common scale parameter, $\sigma > 0$. It is called the common scale parameter because it is incorporated in the variance-covariance matrices of both $\mathscr{Y}$ and $\mathscr{U}$.

The parameter θ determines the $q \times q$ lower triangular matrix Λ_θ, called the *relative covariance factor*, which, in turn, determines the $q \times q$ sparse, symmetric semidefinite variance-covariance matrix $\Sigma_\theta = \sigma^2 \Lambda_\theta \Lambda_\theta^\mathsf{T}$ that defines the distribution of $\mathscr{B}$.

Model Matrices

$\mathbf{X}$ Fixed-effects model matrix of size $n \times p$.
$\mathbf{Z}$ Random-effects model matrix of size $n \times q$.

Derived Matrices

$\mathbf{L}_\theta$ The sparse, lower triangular Cholesky factor of $\Lambda_\theta^\mathsf{T} \mathbf{Z}^\mathsf{T} \mathbf{Z} \Lambda_\theta + \mathbf{I}_q$

In Chap.˜5 this definition will be modified to allow for a *fill-reducing permutation* of the rows and columns of $\Lambda_\theta^\mathsf{T} \mathbf{Z}^\mathsf{T} \mathbf{Z} \Lambda_\theta$.

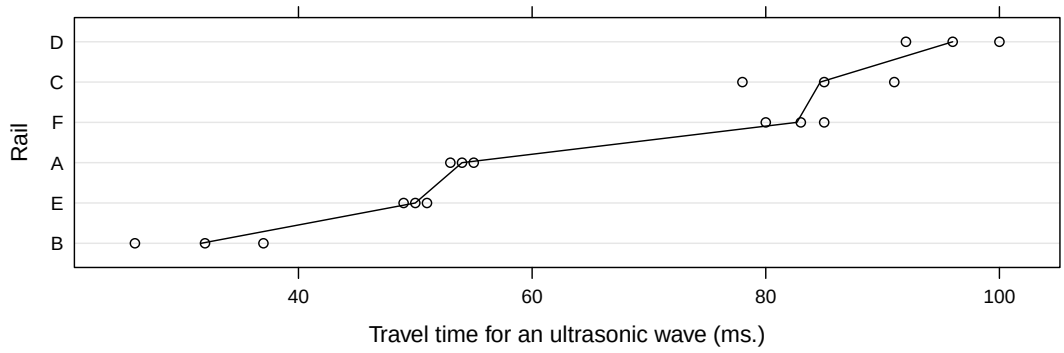

Fig. 1.13 Travel time for an ultrasonic wave test on 6 rails

Vectors

In addition to the parameter vectors already mentioned, we define

$\mathbf{y}$ the n-dimensional observed response vector

γ the n-dimension linear predictor,

$$\gamma = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} = \mathbf{Z}\Lambda_\theta\mathbf{u} + \mathbf{X}\beta$$

μ the n-dimensional conditional mean of $\mathscr{Y}$ given $\mathscr{B} = \mathbf{b}$ (or, equivalently, given $\mathscr{U} = \mathbf{u}$)

$$\mu = \mathrm{E}[\mathscr{Y}|\mathscr{B} = \mathbf{b}] = \mathrm{E}[\mathscr{Y}|\mathscr{U} = \mathbf{u}]$$

$\tilde{u}_\theta$ the q-dimensional conditional mode (the value at which the conditional density is maximized) of $\mathscr{U}$ given $\mathscr{Y} = \mathbf{y}$.

Exercises

These exercises and several others in this book use data sets from the `MEMSS` package for R. You will need to ensure that this package is installed before you can access the data sets.

To load a particular data set, either attach the package

```
> library(MEMSS)
```

or load just the one data set

```
> data(Rail, package = "MEMSS")
```

1.1. Check the documentation, the structure (`str`) and a summary of the `Rail` data (Fig.˜1.13) from the `MEMSS` package. Note that if you used `data` to access this data set (i.e. you did not attach the whole `MEMSS` package) then you must use

```
> help(Rail, package = "MEMSS")
```

to display the documentation for it.

1.2. Fit a model with `travel` as the response and a simple, scalar random-effects term for the variable `Rail`. Use the REML criterion, which is the default. Create a dotplot of the conditional modes of the random effects.

1.3. Refit the model using maximum likelihood. Check the parameter estimates and, in the case of the fixed-effects parameter, its standard error. In what ways have the parameter estimates changed? Which parameter estimates have not changed?

1.4. Profile the fitted model and construct 95% profile-based confidence intervals on the parameters. Is the confidence interval on σ_1 close to being symmetric about the estimate? Is the corresponding interval on $\log(\sigma_1)$ close to being symmetric about its estimate?

1.5. Create the profile zeta plot for this model. For which parameters are there good normal approximations?

1.6. Plot the prediction intervals on the random effects from this model. Do any of these prediction intervals contain zero? Consider the relative magnitudes of $\widehat{\sigma_1}$ and $\widehat{\sigma}$ in this model compared to those in model `fm01` for the `Dyestuff` data. Should these ratios of σ_1/σ lead you to expect a different pattern of prediction intervals in this plot than those in Fig.˜1.11?