## Maximum Likelihood (ML) Estimation

Let $\boldsymbol{\vartheta}$ be a parameter vector for $\boldsymbol{\Delta}$ based on an unconstrained, nonredundant parametrization. The parameters of model (P&B 2.1) are then $\boldsymbol{\beta}$, $\sigma^2$, and $\boldsymbol{\vartheta}$. The likelihood function (given the entire response vector $\boldsymbol{y}$) can be written as

$$L(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \sigma^2|\boldsymbol{y}) = \prod_{i=1}^{M} \int f(\boldsymbol{y}_i|\boldsymbol{b}_i, \boldsymbol{\beta}, \sigma^2) f(\boldsymbol{b}_i|\boldsymbol{\vartheta}, \sigma^2)\, d\boldsymbol{b}_i \tag{P\&B 2.3}$$

$$= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left( \frac{-\sum_{i=1}^{M} \left\| \tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\hat{\boldsymbol{b}}_i \right\|^2}{2\sigma^2} \right) \prod_{i=1}^{M} \frac{\operatorname{abs}|\boldsymbol{\Delta}|}{\sqrt{\left|\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{Z}}_i\right|}}, \tag{P\&B 2.10}$$

where $|\boldsymbol{A}|$ is the determinant of a matrix $\boldsymbol{A}$,

$$\tilde{\boldsymbol{y}}_i = \begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{0} \end{bmatrix}, \qquad \tilde{\boldsymbol{X}}_i = \begin{bmatrix} \boldsymbol{X}_i \\ \boldsymbol{0} \end{bmatrix}, \qquad \tilde{\boldsymbol{Z}}_i = \begin{bmatrix} \boldsymbol{Z}_i \\ \boldsymbol{\Delta} \end{bmatrix} \tag{P\&B 2.7}$$

are augmented data vectors and model matrices (this is called a *pseudo-data* approach), and $\hat{\boldsymbol{b}}_i$ is the mode of the conditional density of $\boldsymbol{b}_i$ given the data, which minimizes $\|\tilde{\boldsymbol{y}}_i - \tilde{\boldsymbol{X}}_i\boldsymbol{\beta} - \tilde{\boldsymbol{Z}}_i\boldsymbol{b}_i\|$. (See P&B, p. 62–64 for the derivation of (P&B 2.10).) However, the $\hat{\boldsymbol{b}}_i$'s depend on $\boldsymbol{\beta}$ and $\boldsymbol{\Delta}$.

The *profiled likelihood* is based on conditional estimates $\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$ and $\hat{\sigma}^2(\boldsymbol{\vartheta})$ (which maximize $L(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \sigma^2|\boldsymbol{y})$ for a given $\boldsymbol{\vartheta}$). The profiled likelihood is a function of $\boldsymbol{\vartheta}$ only and therefore easier to maximize. It is found to be

$$L(\boldsymbol{\vartheta}|\boldsymbol{y}) = L(\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta}), \boldsymbol{\vartheta}, \hat{\sigma}^2(\boldsymbol{\vartheta})|\boldsymbol{y})$$

$$= \frac{\exp(-N/2)}{[2\pi\hat{\sigma}^2(\boldsymbol{\vartheta})]^{N/2}} \prod_{i=1}^{M} \frac{\operatorname{abs}|\boldsymbol{\Delta}|}{\sqrt{\left|\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{Z}}_i\right|}}, \tag{P\&B 2.13}$$

where $\hat{\sigma}^2(\boldsymbol{\vartheta})$ is $1/N$ times the residual sum of squares of a least squares problem:

$$\hat{\sigma}^2(\boldsymbol{\vartheta}) = \frac{1}{N} \min_{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_M,\boldsymbol{\beta}} \left\| \begin{bmatrix} \tilde{\boldsymbol{y}}_1 \\ \tilde{\boldsymbol{y}}_2 \\ \vdots \\ \tilde{\boldsymbol{y}}_M \end{bmatrix} - \begin{bmatrix} \tilde{\boldsymbol{Z}}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} & \tilde{\boldsymbol{X}}_1 \\ \boldsymbol{0} & \tilde{\boldsymbol{Z}}_2 & & \boldsymbol{0} & \tilde{\boldsymbol{X}}_2 \\ \vdots & & \ddots & & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & & \tilde{\boldsymbol{Z}}_M & \tilde{\boldsymbol{X}}_M \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots \\ \boldsymbol{b}_M \\ \boldsymbol{\beta} \end{bmatrix} \right\|^2$$

(See P&B, p. 64–65 for the derivation.) This residual sum of squares can be calculated efficiently by an orthogonal-triangular decomposition (also called QR decomposition) of the $\tilde{\boldsymbol{Z}}_i$'s (or, equivalently, of the above model matrix). From the same minimization problem and using the same decomposition, one can obtain $\hat{\boldsymbol{\beta}}(\boldsymbol{\vartheta})$ and $\hat{\boldsymbol{b}}_i(\boldsymbol{\vartheta})$. (See P&B, p. 66–71.)

The profiled log-likelihood is

$$\ell(\boldsymbol{\vartheta}|\boldsymbol{y}) = \log L(\boldsymbol{\vartheta}|\boldsymbol{y}) = \log \left[ \frac{\exp(-N/2)}{[2\pi\hat{\sigma}^2(\boldsymbol{\vartheta})]^{N/2}} \prod_{i=1}^{M} \frac{\operatorname{abs}|\boldsymbol{\Delta}|}{\sqrt{\left|\tilde{\boldsymbol{Z}}_i^{\mathrm{T}} \tilde{\boldsymbol{Z}}_i\right|}} \right]$$
$$= \frac{N}{2}\left[\log N - \log(2\pi) - 1\right] - \frac{N}{2}\log\left[N\hat{\sigma}^2(\boldsymbol{\vartheta})\right] + \sum_{i=1}^{M} \log \frac{\operatorname{abs}|\boldsymbol{\Delta}|}{\sqrt{\left|\tilde{\boldsymbol{Z}}_i^{\mathrm{T}} \tilde{\boldsymbol{Z}}_i\right|}}$$

and has to be maximized with respect to $\boldsymbol{\vartheta}$.

From the likelihood $L$, the *Akaike information criterion (AIC)* and the *Bayesian information criterion (BIC)* are calculated as follows:

$$\mathrm{AIC} = -2\log L + 2n_{\mathrm{par}}$$
$$\mathrm{BIC} = -2\log L + n_{\mathrm{par}}\log N$$

$n_{\mathrm{par}}$ denotes the number of parameters in the model, $N$ the number of observations. These information criteria summarize both lack of fit and model complexity. They can be used for model selection; see Section 2.3.

**Example 2.5.** Rail data (continued, P&B p. 71ff). The profiled log-likelihood for the model for the rail data given in Examples 2.1 and 2.2 is

$$\ell(\vartheta|\boldsymbol{y}) = \frac{N}{2}\left[\log N - \log(2\pi) - 1\right] - \frac{N}{2}\log\left[N\hat{\sigma}^2(\vartheta)\right] + \sum_{i=1}^{M} \log \frac{\Delta}{\sqrt{\tilde{\boldsymbol{Z}}_i^{\mathrm{T}} \tilde{\boldsymbol{Z}}_i}},$$

which is slightly simpler than the general expression since $\tilde{\boldsymbol{Z}}_i$ only has one column ($q = 1$) and since the relative precision factor $\Delta$ is a (positive) scalar in this case. We use $\vartheta = \log\Delta$ as the unconstrained parameter for $\Delta$.

The first (constant) term can be neglected for the optimization.

The second term, $(N/2)\log[N\hat{\sigma}^2(\vartheta)] = N\log[\sqrt{N\hat{\sigma}^2(\vartheta)}]$, is (up to the factor $N$) the logarithm of the norm of the residual vector in the least-squares model given in the formula for $\hat{\sigma}^2(\boldsymbol{\vartheta})$. This corresponds to a penalized least-squares fit for model (P&B 2.1) with penalty term $\sum_{i=1}^{M}(\Delta b_i)^2$. When $\Delta$ is close to 0 ($\vartheta \to -\infty$), which happens for large $\sigma_b^2$, the second term is approximately $N$ times the logarithm of the norm of the residual vector of the corresponding unpenalized model,

$$\boldsymbol{y} = \begin{bmatrix} \boldsymbol{Z} & \boldsymbol{X} \end{bmatrix} \begin{bmatrix} \boldsymbol{b} \\ \beta \end{bmatrix} + \boldsymbol{\varepsilon}.$$

If the penalty coefficient $\Delta$ is large ($\vartheta \to \infty$), which happens for small $\sigma_b^2$, then the "estimates" of the random effects are forced towards 0, and the second term is approximately $N$ times the logarithm of the norm of the residual vector of the model

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{\varepsilon}$$

without random effects.

The third term is close to 0 if $\Delta$ is large enough (since $\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{Z}}_i = \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{Z}_i + \Delta^2$ and the ratio is therefore close to 1 for every $i$). It is always negative since the ratios are less than 1. If $\Delta$ is small ($\vartheta \to -\infty$), $\tilde{\boldsymbol{Z}}_i^{\mathrm{T}}\tilde{\boldsymbol{Z}}_i \approx \boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{Z}_i$, and the third term is therefore roughly

$$M\vartheta - \sum_{i=1}^{M} \log\sqrt{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{Z}_i},$$

a linear function in $\vartheta = \log\Delta$. The second and third term (as well as the entire profiled log-likelihood) are plotted as functions of $\Delta$ and $\vartheta$ in Figures 2.1 and 2.2 of P&B, respectively.

If a maximum of the profiled log-likelihood exists for a finite $\vartheta$, a possible starting point for finding this $\vartheta$ is $\vartheta_0 = \sum_{i=1}^{M} \log\sqrt{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{Z}_i}/M$. Here, $\vartheta_0 = 0.549$, while the final maximum likelihood estimate is

$$\hat{\vartheta} = \log\sqrt{\frac{\hat{\sigma}^2}{\hat{\sigma}_b^2}} = \log\sqrt{\frac{4.021^2}{22.62^2}} = -1.728,$$

where $\hat{\sigma}$ and $\hat{\sigma}_b$ can be taken from the R output of:

```
rail.lme.ml <- lme(travel ~ 1, random = ~ 1 | Rail,
  data = rail.df, method='ML')
summary(rail.lme.ml)
```

This output also contains the maximized log-likelihood of $-64.28$ as well as the values of AIC $= -2(-64.28) + 2 \cdot 3 = 134.56$ and BIC $= -2(-64.28) + 3 \cdot \log(18) = 137.23$. $\triangle$

Once the variance parameters have been estimated, the fixed-effects parameter $\boldsymbol{\beta}$ can in principle also be estimated as

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^{\mathrm{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{y},$$

where $\hat{\boldsymbol{V}} = \boldsymbol{V}(\hat{\boldsymbol{\vartheta}})$, with estimated variance-covariance matrix $(\boldsymbol{X}^{\mathrm{T}}\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}$.

The random effects $\boldsymbol{b}_i$ are not parameters of the model. Therefore, they are not estimated, but it is possible to obtain *predictions*. The predictor of $\boldsymbol{b}$ based on ML is

$$\hat{\boldsymbol{b}} = \left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z} + \hat{\sigma}^2\hat{\boldsymbol{G}}^{-1}\right)^{-1}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}),$$

where

$$\hat{\boldsymbol{G}} = \left.\begin{bmatrix} \hat{\boldsymbol{\Psi}} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{\Psi}} & & \boldsymbol{0} \\ \vdots & & \ddots & \\ \boldsymbol{0} & \boldsymbol{0} & & \hat{\boldsymbol{\Psi}} \end{bmatrix}\right\} M \text{ times}$$

The least squares estimate for $\boldsymbol{b}$ in a fixed-effects model $\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{Z}\boldsymbol{b} + \boldsymbol{\varepsilon}$ would be $(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})$. In the above predictor for the random effect, the additional term $\hat{\sigma}^2\hat{\boldsymbol{G}}^{-1}$ causes $\hat{\boldsymbol{b}}$ to be closer to $\boldsymbol{0}$ than a fixed-effects estimate ("shrunken").

In R (in both the `nlme` and the `lme4` package), the predictions of the random effects can be obtained using the `ranef` function.

### Restricted Maximum Likelihood (REML) Estimation

Since maximum likelihood tends to underestimate variance parameters (such as $\sigma^2$, $\sigma_b^2$), restricted (or residual) maximum likelihood is often preferred.

One possible definition of the restricted likelihood is

$$L_{\mathrm{R}}(\boldsymbol{\vartheta}, \sigma^2|\boldsymbol{y}) = \int L(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \sigma^2|\boldsymbol{y})\, d\boldsymbol{\beta}.$$

Again, it is easier to use a profiled (log) restricted likelihood, i. e., to express $\hat{\sigma}^2_{\mathrm{R}}$ as a function of $\boldsymbol{\vartheta}$ and then optimize with respect to $\boldsymbol{\vartheta}$, which gives $\hat{\boldsymbol{\vartheta}}_{\mathrm{R}}$ and $\hat{\sigma}^2_{\mathrm{R}} = \hat{\sigma}^2_{\mathrm{R}}(\hat{\boldsymbol{\vartheta}}_{\mathrm{R}})$.

REML does not directly lead to an estimate for $\boldsymbol{\beta}$ since $L_{\mathrm{R}}(\boldsymbol{\vartheta}, \sigma^2|\boldsymbol{y})$ does not depend on $\boldsymbol{\beta}$. Nevertheless, after calculating $\hat{\boldsymbol{\vartheta}}_{\mathrm{R}}$, one can obtain a "best guess" at $\boldsymbol{\beta}$ using formulas from ML theory.

The restricted likelihood depends on the fixed-effects model matrices $\boldsymbol{X}_i$ in the sense that a reparametrization (e. g., using different contrasts) has an influence on the restricted likelihood. Therefore, models with different fixed-effects model matrices should not be compared based on the restricted likelihoods – the likelihoods of the models fitted by ML should be used instead (P&B p. 75f).

The AIC and the BIC for REML are defined as follows (P&B p. 84):

$$\begin{aligned}\mathrm{AIC} &= -2\log L_{\mathrm{R}} + 2n_{\mathrm{par}}\\ \mathrm{BIC} &= -2\log L_{\mathrm{R}} + n_{\mathrm{par}}\log(N-p)\end{aligned}$$

**Example 2.6.** Rail data (continued, P&B p. 10). REML is the default fitting method of `lme`. The log-restricted-likelihood (from the output of `summary(rail.lme)`, see p. 4) of this model is $-61.09$, from which AIC $= -2(-61.09) + 2 \cdot 3 = 128.18$ and BIC $= -2(-61.09) + 3\log(17) = 130.68$. △

**Example 2.7.** Machines data (continued). In Example 2.4, we saw that the models fitted by REML depended on the contrasts used for the fixed effects: Evidently, the fixed-effects parameter estimates are different for Helmert contrasts compared to a parametrization with a reference category. But the restricted log-likelihood is different, too – even though the two models are equivalent. △

## 2.2 Multi-Level LME Models

For two-level models, i. e., models with two nested levels of random effects, the notation from Section 2.1 can be extended as follows:

$$\begin{aligned}\boldsymbol{y}_{ij} &= \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{i,j}\boldsymbol{b}_i + \boldsymbol{Z}_{ij}\boldsymbol{b}_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad i = 1,\ldots,M, \quad j = 1,\ldots,M_i,\\ \boldsymbol{b}_i &\sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Psi}_1),\\ \boldsymbol{b}_{ij} &\sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Psi}_2),\\ \boldsymbol{\varepsilon}_{ij} &\sim \mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_{n_{ij}}).\end{aligned} \qquad \text{(P\&B 2.2)}$$

$\boldsymbol{y}_{ij}$ is the $n_{ij}$-dimensional response vector for the $j$th second-level (sub-)group within the $i$th first-level group. $\boldsymbol{b}_i$ is the $q_1$-dimensional vector of first-level random effects in group $i$, $\boldsymbol{b}_{ij}$ the $q_2$-dimensional vector of second-level random effects in subgroup $j$ of group $i$. $\boldsymbol{X}_{ij}$ is the fixed-effects regressor matrix ($n_{ij} \times p$), $\boldsymbol{Z}_{i,j}$ the first-level random-effects regressor matrix ($n_{ij} \times q_1$), and $\boldsymbol{Z}_{ij}$ the second-level random-effects regressor matrix ($n_{ij} \times q_2$) for subgroup $j$ of group $i$. $\boldsymbol{\varepsilon}_{ij}$ is the $n_{ij}$-dimensional within-subgroup error vector. All $\boldsymbol{b}_i$, $\boldsymbol{b}_{ij}$, and $\boldsymbol{\varepsilon}_{ij}$ are assumed to be independent.

Per first-level group, this two-level model can be written as

$$\begin{aligned}\boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{y}_{i1} \\ \vdots \\ \boldsymbol{y}_{iM_i} \end{bmatrix} &= \underbrace{\begin{bmatrix} \boldsymbol{X}_{i1} \\ \vdots \\ \boldsymbol{X}_{iM_i} \end{bmatrix}}_{\boldsymbol{X}_i} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{Z}_{i,1} \\ \vdots \\ \boldsymbol{Z}_{i,M_i} \end{bmatrix} \boldsymbol{b}_i + \begin{bmatrix} \boldsymbol{Z}_{i1} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{Z}_{iM_i} \end{bmatrix} \begin{bmatrix} \boldsymbol{b}_{i1} \\ \vdots \\ \boldsymbol{b}_{iM_i} \end{bmatrix} + \underbrace{\begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \\ \vdots \\ \boldsymbol{\varepsilon}_{iM_i} \end{bmatrix}}_{\boldsymbol{\varepsilon}_i} \\ &= \boldsymbol{X}_i\boldsymbol{\beta} + \underbrace{\begin{bmatrix} \boldsymbol{Z}_{i,1} & \boldsymbol{Z}_{i1} & & \boldsymbol{0} \\ \vdots & & \ddots & \\ \boldsymbol{Z}_{i,M_i} & \boldsymbol{0} & & \boldsymbol{Z}_{iM_i} \end{bmatrix}}_{\boldsymbol{Z}_i} \underbrace{\begin{bmatrix} \boldsymbol{b}_i \\ \boldsymbol{b}_{i1} \\ \vdots \\ \boldsymbol{b}_{iM_i} \end{bmatrix}}_{\tilde{\boldsymbol{b}}_i} + \boldsymbol{\varepsilon}_i,\end{aligned}$$

with

$$\tilde{\boldsymbol{b}}_i \sim \mathrm{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Psi}_1 & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Psi}_2 & & \boldsymbol{0} \\ \vdots & & \ddots & \\ \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{\Psi}_2 \end{bmatrix}\right), \quad \boldsymbol{\varepsilon}_i \sim \mathrm{N}\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{\sum_{j=1}^{M_i} n_{ij}}\right).$$

If all $M_i$ are the same, then the dimension $q := q_1 + M_i q_2$ of the random effects vector $\tilde{\boldsymbol{b}}_i$ does not depend on $i$, and this two-level model can be interpreted as a single-level model with constraints on the variance-covariance matrix of the random effects (see Section 2.5) and with a specific form of the random-effects regressor matrix.

An extension of the notation for a three-level model is given in P&B on p. 60f. Model fitting for multiple levels of random effects is described in P&B on p. 77f.

**Example 2.8.** Pixel intensity on CT scans in dogs (P&B p. 40ff). In this experiment on the pixel intensity in computerized tomography (CT) scans, ten dogs received an injection of a dye contrast. The mean pixel intensities from CT scans of the right and left lymph nodes in the axillary region were recorded for each dog on several occasions up to 21 days after the injection.

It is expected that the results of the experiment may be different for each dog and also for both sides of the same dog, but not systematically, i. e., the values may be higher on the left side for one dog and on the right side for an other dog. Therefore, we use random effects at two levels, where `Side` is nested within `dog`.

Figure 2.5 shows that the intensity values typically increase at the beginning and decrease again later. For this reason, a model with a quadratic effect of `day` is used. Since the

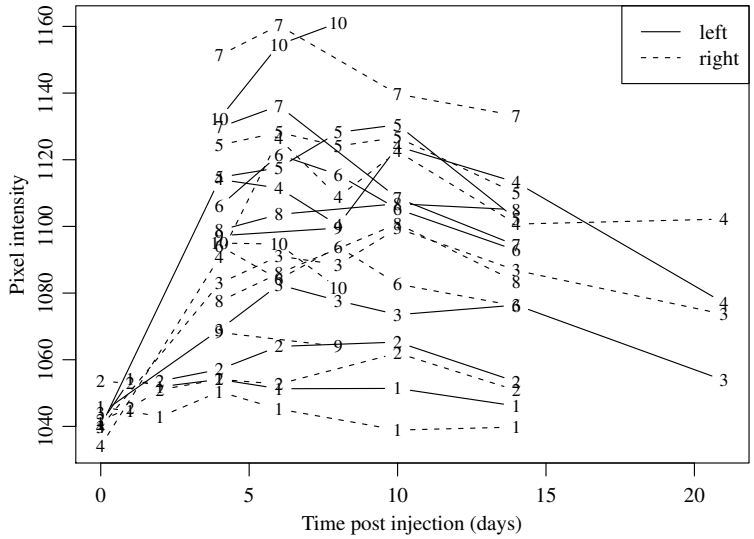


Figure 2.5: The pixel intensity data. The numbers indicate the dogs.

maximum does not seem to be reached at the same time for all dogs, we include random intercepts and slopes at the `Dog` level (level 1). The curves for the left and right side of the same dog roughly differ by a vertical shift only. At the `Side` level (level 2), we therefore only use a random intercept. The resulting two-level mixed-effects model can be expressed as

$$\begin{aligned} &y_{ijk} = \beta_1 + \beta_2 d_{ik} + \beta_3 d_{ik}^2 + b_{i,1} + b_{i,2} d_{ik} + b_{ij} + \varepsilon_{ijk}, \\ &i = 1, \ldots, 10, \quad j = 1, 2, \quad k = 1, \ldots, n_{ij}. \end{aligned} \tag{P\&B 1.9}$$

$y_{ijk}$ is the mean pixel intensity at the $k$th measurement (at time $d_{ik}$) on the $j$th side of the $i$th dog.

With $\boldsymbol{y}_{ij}$ for all observations on one side of one dog as in the general model formula (P&B 2.2), the model is

$$\begin{aligned} &\boldsymbol{y}_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{i,j}\boldsymbol{b}_i + \boldsymbol{Z}_{ij} b_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, M_i, \\ &\boldsymbol{b}_i \sim \mathrm{N}(\boldsymbol{0}, \boldsymbol{\Psi}_1), \quad b_{ij} \sim \mathrm{N}(0, \sigma_2^2), \quad \boldsymbol{\varepsilon}_{ij} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_{n_{ij}}). \end{aligned}$$

Measurements for both sides were taken at the same times, and $n_{i1} = n_{i2}$ for all dogs $i$. Therefore, the design matrices do not depend on $j$, but only on $i$. E. g., for dog 8, scans were made on days 4, 6, 10, and 14, and therefore

$$\boldsymbol{X}_{81} = \boldsymbol{X}_{82} = \begin{bmatrix} 1 & 4 & 16 \\ 1 & 6 & 36 \\ 1 & 10 & 100 \\ 1 & 14 & 196 \end{bmatrix}, \quad \boldsymbol{Z}_{8,1} = \boldsymbol{Z}_{8,2} = \begin{bmatrix} 1 & 4 \\ 1 & 6 \\ 1 & 10 \\ 1 & 14 \end{bmatrix}, \quad \boldsymbol{Z}_{81} = \boldsymbol{Z}_{82} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

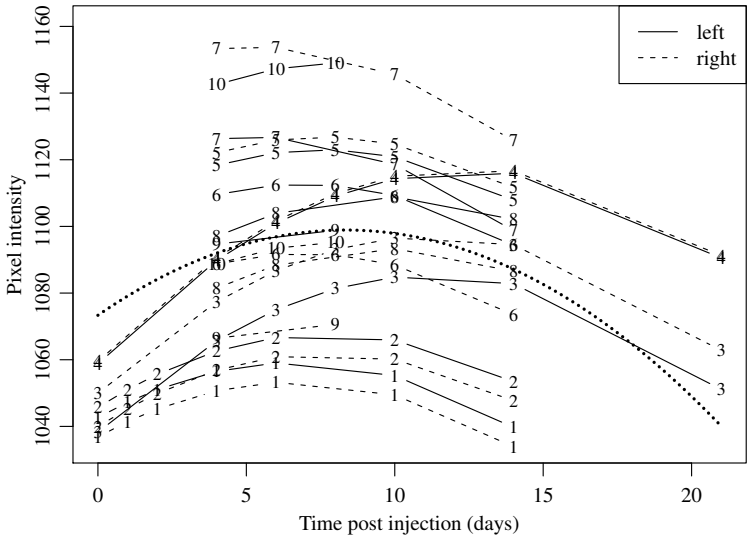


Figure 2.6: Fitted curves for each side of each dog, as well as the fixed-effects part of model `pixel.lme.1` (dotted curve).

This model can be fitted using the following code:

```
pixel.lme.1 <- lme(pixel ~ day + I(day^2),
  random = list(Dog = ~ day, Side = ~ 1), data=pixel.df)
```

From `summary(pixel.lme.1)`, we obtain $1073.34 + 6.130\, d_{ik} - 0.3674\, d_{ik}^2$ as the fixed-effects part of the model. This is shown as the dotted curve in Figure 2.6, together with the fitted values for all observations. The variance/covariance parameter estimates can be obtained from `summary` (giving standard deviations) and/or `VarCorr(pixel.lme.1)` (giving variances). They are

$$\hat{\boldsymbol{\Psi}}_1 = \begin{bmatrix} 804.85 & -29.03 \\ -29.03 & 3.40 \end{bmatrix}, \quad \hat{\sigma}_2^2 = 283.06, \quad \hat{\sigma}^2 = 80.81.$$

With the `lmer` function, the same model could be specified as `pixel ~ day + I(day^2) + (day | Dog) + (1 | Side:Dog)`. △

**Example 2.9.** Machines data (continued, P&B, p. 23f). We can use a two-level model to fit random intercepts both at the `Worker` level ($b_i$, level 1) and at the `Machine` level ($b_{ij}$, level 2):

$$\begin{aligned} y_{ijk} &= \beta_j + b_i + b_{ij} + \varepsilon_{ijk}, \quad i = 1, \ldots, 6, \quad j = 1, \ldots, 3, \quad k = 1, \ldots, 3, \\ b_i &\sim \mathrm{N}(0, \sigma_1^2), \quad b_{ij} \sim \mathrm{N}(0, \sigma_2^2), \quad \varepsilon_{ijk} \sim \mathrm{N}(0, \sigma^2), \end{aligned}$$

or

$$\begin{aligned} \boldsymbol{y}_{ij} &= \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{i,j} b_i + \boldsymbol{Z}_{ij} b_{ij} + \boldsymbol{\varepsilon}_{ij}, \quad i = 1, \ldots, 6, \quad j = 1, \ldots, 3, \\ b_i &\sim \mathrm{N}(0, \sigma_1^2), \quad b_{ij} \sim \mathrm{N}(0, \sigma_2^2), \quad \boldsymbol{\varepsilon}_{ij} \sim \mathrm{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_3), \end{aligned}$$

where $\boldsymbol{X}_{ij}$ $(3 \times 3)$ has ones in the $j$th column and zeros elsewhere, and

$$\boldsymbol{Z}_{i,j} = \boldsymbol{Z}_{ij} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

This model (but with a reference-group parametrization for the fixed effects, as in Example 2.4) can be fitted using the following R code:

```
machines.lme.2 <- lme(score ~ Machine, random = ~ 1 | Worker/Machine,
  data=machines.df)
```

An equivalent specification of the random part, similar to the one used in Example 2.8, would be `random = list(Worker = ~ 1, Machine = ~ 1)`. And with `lmer`, the formula could be specified as `score ~ Machine + (1 | Worker/Machine)` or `score ~ Machine + (1 | Worker) + (1 | Machine:Worker)`. △

## 2.3 Hypothesis Tests and Confidence Intervals

### Likelihood Ratio Tests for Random Effects

A model is called *nested* within another model if it can be written as a special case of the other, more general model. As for many other model families, a comparison of two nested models fitted by maximum likelihood can be done by a *likelihood ratio test* (LRT; P&B p. 83ff).

Let $L_2$ be the likelihood of the more general model $M_2$, $L_1$ that of the more specific model $M_1$. Then $L_2 \geq L_1$, and the LRT statistic,

$$2 \log \frac{L_2}{L_1} = 2(\log L_2 - \log L_1),$$

is nonnegative. Let $k_i$ be the number of unknown parameters of model $M_i$. Under the null hypothesis that model $M_1$ is adequate, the distribution of the LRT statistic is asymptotically $\chi^2$ with $k_2 - k_1$ degrees of freedom.

Likelihood ratio tests can also be used for the comparison of two REML fits if they are based on the same fixed-effects design matrix. (Remember that the restricted likelihood depends on this design matrix.)

In R, likelihood ratio tests can be obtained by applying the function `anova` to two (or more) models.

**Example 2.10.** Growth curves (continued). In Example 2.3, we fitted two models to the growth curves data for eleven girls. `ortho.fem.lme.1` had a random intercept per girl, `ortho.fem.lme.2` both a random intercept and a random slope (for a linear age effect) per girl; both models had an intercept and a slope in the fixed-effects part. We compare these two models by a likelihood ratio test: