Statistic Seminar, 6th talk:

Prediction of New Observations

Martina Albers
Monday, 12. April 2010

1 Introduction

When speaking of a prediction for a linear mixed model, we need to specify what kind of prediction we are interested in. We differentiate between the prediction of a random effect (or mixed effects) and the prediction of future observations. In the second case there many different approaches. In this talk we will concentrate on predictions of future observations, where we will assume that the future observation has a certain distribution. This distribution is defined up to a finite number of parameters which we will have to estimate, from which we will get a so called "prediction interval".

2 Recall: the linear mixed model

We consider the model $y = X\beta + Zb + \epsilon = W\gamma + \epsilon$, where b and ϵ are uncorrelated, and

y	$n \times 1$	vector of observations
β	$p \times 1$	vector of fixed effects
X	$n \times p$	matrix associating observ. with fixed effects
b	$q \times 1$	vector of random effects
Z	$n \times q$	matrix associating observ. with random effects
e	$n \times 1$	vector of residual errors
W, γ		combined design matrix resp. vector of effects

$$(Y\,|B = b) \sim N\left(X\beta + Zb, \sigma_1^2 I_n\right)$$
$$B \sim N\left(0, \Sigma_\theta\right)$$

, where $\Sigma_\theta = \mathrm{cov}(B) = \sigma_1^2 \Lambda_\theta \Lambda_\theta'$ is a $q \times q$ symmetric matrix, $\epsilon \sim N(0, \sigma^2 I_n)$. We rewrite the model using $B = \Lambda_\theta U$, where $U \sim N\left(0, \sigma^2 I_n\right)$. In order to find $\hat{u}$ and $\hat{\beta}_\theta$ such that

$$\begin{pmatrix} \hat{u} \\ \hat{\beta}_\theta \end{pmatrix} = \mathrm{argmin}_{u,\beta} \left\| \begin{pmatrix} y \\ 0 \end{pmatrix} - \begin{pmatrix} Z\Lambda_\theta X & X \\ I & 0 \end{pmatrix} \begin{pmatrix} u \\ \beta \end{pmatrix} \right\|^2$$

we need to solve the following equations for $\hat{u}$ and $\hat{\beta}_\theta$

$$\begin{pmatrix} \Lambda_\theta Z' Z \Lambda_\theta + I_q & \Lambda_\theta' Z' X \\ X' Y \Lambda_\theta & X'X \end{pmatrix} - \begin{pmatrix} \hat{u} \\ \hat{\beta}_\theta \end{pmatrix} = \begin{pmatrix} \Lambda_\theta' Z' y \\ X' y \end{pmatrix}$$

3 Prediction of a future observation

3.1 Prediction intervals

A prediction interval is defined to be an interval in which a future observation will fall with a certain probability given what has been observed. In order to fully understand this definition, one needs to see the difference between a prediction and a confidence interval.

3.1.1 Prediction intervals for a fixed effect model

Lets assume we have a fixed effect model: $y = X\beta + \epsilon$, where $\epsilon \sim N(0, \sigma^2)$. β will be estimated by: $\hat{\beta} = (X'X)^{-1} X'y$.
Use the following notation:

- true: $\mathring{y}_i = x'_i\beta$
- observed: $y_i = x'_i\beta + \epsilon$
- modelled: $\hat{y}_i = x'_i\hat{\beta}$
- predicted: $\hat{y}_{n+1} = x'_{n+1}\hat{\beta}$

For constructing a confidence interval we need to consider

$$\mathrm{cov}(\hat{y}_i - \mathring{y}_i) = \sigma^2 x'_i (X'X)^{-1} x_i$$

For a prediction interval on the other hand we need the variance of the difference between our **predicted** and **observed** values, i.e.

$$\mathrm{cov}(\hat{y}_{n+1} - y_{n+1}) = \sigma^2 \left(x'_{n+1} (X'X)^{-1} x_{n+1} + 1 \right)$$

So (under normal approximation) the confidence interval is:

$$\hat{y} \pm 1.96 \cdot \sigma^2 x'_i (X'X)^{-1} x_i$$

and the prediction interval:

$$\hat{y} \pm 1.96 \cdot \sigma^2 \left(x'_{n+1} (X'X)^{-1} x_{n+1} + 1 \right)$$

3.1.2 Prediction intervals for a mixed effects model

Now we consider the linear mixed model $y = X\beta + Zb + \epsilon = W\gamma + \epsilon$. Defining V to be the variance of y we get: $V = Z\mathrm{cov}(b)Z' + \mathrm{cov}(\epsilon) = {\sigma_1}^2 ZZ' + \sigma^2 I_n$.
By estimating the two factors σ^2 and ${\sigma_1}^2$ we can estimate β and b by:

$$\hat{\beta} = \left(X'\widehat{V}^{-1}X \right)^{-1} X'\widehat{V}^{-1}y$$
$$\tilde{b} = \hat{\sigma_1}^2 Z'\widehat{V}^{-1} \left(y - X\hat{\beta} \right)$$

, where $\widehat{V} = \hat{\sigma_1}^2 Z'Z + \hat{\sigma}^2 I_n$.
If we compute the prediction interval as in the fixed effect model, we get an interval for a so-called **marginal** prediction. The variance-term that is needed is:

$$\mathrm{cov}(\hat{y}_{n+1} - y_{n+1}) = x'_{n+1}\mathrm{cov}(\hat{\beta} - \beta)x_{n+1} + z'_{n+1}\mathrm{cov}(b)z_{n+1} + \mathrm{cov}(\epsilon)$$

There is also another type of prediction, the **conditional** one. In this case we average over the random effects, so we get a prediction specific to the observed random effect. To determine the prediction interval, we take also the estimation of b, $\tilde{b}$, in consideration and we get:

$$\mathrm{cov}\left(\hat{y}_{n+1} - y_{n+1} \middle| b = \tilde{b}\right) = \sigma^2 \left(x_{n+1} \left(X'X\right)^{-1} x'_{n+1} + 1\right)$$

4 Principles of prediction

As we have seen, there is a remarkable difference between marginal and conditional predictions. How to decide which one is of interest? It is rather obvious that there are no general rules for how to proceed. It depends on the experiment itself and of course on our interest.
A prediction is typically associated with a combination of explanatory variables (e.v.). The e.v. can be averaged over, ignored or evaluated at a specific value. This is why we partition the e.v. into three sets: (1) classifying set, (2) averaging set, (3) rest. The first set contains the variables for which we require predicted values. The second on the other hand contains variables over which we'll average over. At last, the third set is to be ignored.

4.1 Conditional vs. Marginal prediction

Consider random factor terms. We have seen above that there are three possibilities how to deal with them. Averaging over this set of random effects corresponds to a conditional prediction, which yields a prediction specific to the random effects observed.
On the other hand, if we omit the random factor term, we get a marginal prediction. This means that be predict at the population average, or in other words: we substitute the assumed population mean for an unknown effect. While considering fixed factors, omitting a fixed term yields no natural interpretation. Thus we average over all present levels to give a conditional average.

4.2 Prediction process in 4 steps

These are the four main conceptual steps involved in the prediction process:
1. Determine the classifying set
2. Determine the averaging set
3. Determine terms that are needed to compute parameters and estimations
4. Choose the weighting for taking means over margins in the averaging set