Predictions in mixed-effects models

Štefan Varga

Department of Mathematics
Faculty of Chemical and Food Technology STU,
Radlinského 9, 812 34 Bratislava
E-mail : stefan.varga@stuba.sk

Abstract. We consider models of indirect measurements with mixed effects and their applications in chemical and food technology. Predictions of response variables are obtained using the maximum likelihood method.

The classical regression model of indirect measurements is usually studied in the form

$$y = a_1 f_1(x) + a_2 f_2(x) + ... + a_m f_m(x) \qquad (1)$$

where $f_i(x)$ $(i = 1, 2, ..., m)$ are known functions of the input variable x (predictor), y is an output variable (response) and the vector of unknown parameters $\theta = (a_1, a_2, ..., a_m)^T$ is not random but fixed. On the other hand mixed-effects models are regression models with a random vector of unknown parameters. They are used to analyze grouped data, repeated measures data or data impacted some factor variables. In the model of direct measurements, using analysis of variance, we can examine an influence of the factor variables on the values (expected values) of the measured variable. In the case of the model of indirect measurements an influence of the factor variables on the dependence of the observed variable Y on the predictor x is studied. Therefore it is reasonable to consider the unknown parameters $a_1, a_2, ..., a_m$ as sums of fixed and random parts.

The simplest mixed-effects model of indirect measurements could be

$$y_{ij} = (a_1 + a_{i1}) f_1(x_j) + (a_2 + a_{i2}) f_2(x_j) + ... + (a_m + a_{im}) f_m(x_j) + e_{ij} \qquad (2)$$

where y_{ij} is a value of the observed variable Y, the level of the factor is i $(i = 1, 2, ..., k)$ and the value of the predictor is x_j $(j = 1, 2, ..., n)$. The values $a_1, a_2, ..., a_m$ are fixed components of the unknown regression coefficients and $a_{i1}, a_{i2}, ..., a_{im}$ are the random effects in the coefficients associated with the i_{th} level of the factor. It is assumed that the vectors $\theta_i = (a_{i1}, a_{i2}, ..., a_{im})^T$ are independent $(i = 1, 2, ..., k)$ and identically distributed with k-dimensional normal distribution $N_k(0, \sigma^2 H)$ and that the e_{ij} are errors of the measurements y_{ij} independent and identically distributed with $N(0, \sigma^2)$ distribution.

The aim is to estimate both the fixed parts and the random parts of the unknown regression coefficients, to test statistical significance of the vector of random components of the unknown regression coefficients and to predict observed variable Y for different level of the factor (factors) or for different subjects. We will present solutions of these problems using the example from the field of food technology.

Literature:

[1] Härdle, W. : Smoothing Techniques with Implementation in S. New York, Springer - Verlag 1981.

[2] Krishnaiah, P.R., Sen, P.K. :Handbook of statistics 4, Nonparametric methods. Amsterdam, North Holland 1984.

[3] Venables, W.N., Ripley, B.D. : Modern Applied Statistics with S-plus. New York, Springer - Verlag 1994.

[4] Varga, Š. : Uncertainty in regression models. Proceedings of the Scientific Colloquium PRASTAN, Kočovce 2001, 154 – 160.