Detection of structural changes in generalized linear models

Jaromír Antoch
Department of Statistics
Charles University of Prague
jaromir.antoch@karlin.mff.cuni.cz

Daniela Jarušková
Department of Mathematics
Czech Technical University
jarus@math.fsv.cvut.cz

Keywords: GLIM, stuctural changes, change point analysis.

Abstract

In many statistical applications it is typical to model the dependence of a response Y on a vector of covariates $\mathbf{X}$ using the classical linear model, i.e.:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p + e. \tag{1}$$

Unfortunately, the model (1) does not cover many real situations as, e.g., the logistic regression where the response takes only two values $\{0, 1\}$. Therefore, the generalized linear model (GLIM) were introduced and studied since more than thirty years.

The basic setup of GLIM consists of modelling a transformation of the regression function $\mu(\mathbf{x}) = E(Y \mid \mathbf{X} = \mathbf{x})$ as linear, that is

$$g(\mu(\mathbf{x})) = \beta_0 + \sum_{j=1}^{p} \beta_j X_j = \mathbf{x}'\beta, \tag{2}$$

where $\mathbf{x}$ is the vector of values of covariates, β is a vector of parameters and $g(\cdot)$ is a proper link function.

In our contribution we will assume that the observations $(y_i, \mathbf{x_i})$ were obtained in time moments $t_1 < t_2 < \ldots < t_n$. Such a situation typically occurs in the longitudinal studies in psychometry, medicine or biology.

It can happen that at an *unknown time point* t_m the way how the response depends on covariates may change. This unknown time point t_m is usually called the *change point* and this type of problem is called the *change point analysis.* For broad overviews of results see, e.g., Csörgő and Horváth (1997) or Antoch, Hušková and Jarušková (2002).

While there exist some results concerning the change point detection in classical linear model (1), very little attention has been payed to the GLIM setup. Therefore, in our contribution we will concentrate on that. More precisely, instead of (2) we will consider the model

$$g(\mu(\mathbf{x})) = \begin{cases} \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} = \mathbf{x}_i'\beta, & i = 1, \ldots, m, \\ \beta_0^\star + \sum_{j=1}^{p} \beta_j^\star x_{ij} = \mathbf{x}_i'\beta^\star, & i = m+1, \ldots, n, \end{cases} \tag{3}$$

where $\beta \neq \beta^\star$ and m are unknown parameters. This model describes the situation when the first m observations follow the GLIM with the parameter β and the remaining $n - m$ observations the same model with the parameter $\beta^\star$.

During the lecture we will concentrate on the first goal of the statistical inference, i.e., how to decide whether a change has occurred. This problem is typically solved by testing the null hypothesis H_0 against the alternative H_1, where:

$$H_0: \quad m = n \qquad \textit{and} \qquad H_1 : m < n. \tag{4}$$

The appropriate test statistics and their limit distribution under the null hypothesis will be derived, that may be used to find approximate critical values.

The approach will be illustrated on several practical examples.

References

[1] Antoch J., Hušková M. and Jarušková D. (2002). *Off-line quality control.* In: Multivariate Total Quality Control: Foundations and Recent Advances, N. C. Lauro et al. eds., Springer-Verlag, Heidelberg, 1 − 86.

[2] Csörgő M. and Horváth L. (1997). *Limit Theorems in Change-Point Analysis.* J. Wiley, New York.

[3] Fahrmeir L. and Tutz G. (1994). *Multivariate statistical modelling based on generalized linear models.* Springer, New York.

[4] McCullagh P. and Nelder J. (1989). *Generalized Linear Models.* Chapman & Hall, London, 2nd ed.