Computational Statistics and Data Analysis 57 (2013) 684–698

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Change point models for cognitive tests using semi-parametric maximum likelihood

Ardo van den Hout [a,*], Graciela Muniz-Terrera [b], Fiona E. Matthews [c]

[a] Department of Statistical Science, University College London, 1–19 Torrington Place, London WC1E 7HB, UK
[b] MRC Unit for Lifelong Health and Ageing, London, UK
[c] MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK

ARTICLE INFO

Article history:
Received 15 September 2011
Received in revised form 23 July 2012
Accepted 26 July 2012
Available online 3 August 2012

Keywords:
Beta-binomial distribution
Latent class model
Mini-mental state examination
Random-effects model

ABSTRACT

Random-effects change point models are formulated for longitudinal data obtained from cognitive tests. The conditional distribution of the response variable in a change point model is often assumed to be normal even if the response variable is discrete and shows ceiling effects. For the sum score of a cognitive test, the binomial and the beta-binomial distributions are presented as alternatives to the normal distribution. Smooth shapes for the change point models are imposed. Estimation is by marginal maximum likelihood where a parametric population distribution for the random change point is combined with a non-parametric mixing distribution for other random effects. An extension to latent class modelling is possible in case some individuals do not experience a change in cognitive ability. The approach is illustrated using data from a longitudinal study of Swedish octogenarians and nonagenarians that began in 1991. Change point models are applied to investigate cognitive change in the years before death.

1. Introduction

The scale of a cognitive test is often discrete. A typical example is the Mini-Mental State Examination (MMSE Folstein et al., 1975) which has integer scoring. The MMSE is a questionnaire for screening dementia and has items on, for instance, language and memory. Scores for each of the questions are added up to obtain a final integer sum score ranging from 0 to 30.

This paper discusses and extends methodology for random-effects change point models for longitudinal data on cognitive tests. A change point model assumes a stochastic process over time that shows a one-off change in direction, see, e.g., Dominicus et al. (2008). Change points are sometimes called *turning points* (McArdle and Wang, 2008) or *break points* (Stasinopoulos and Rigby, 1992; Muggeo, 2008). Models with more than one change point are typically applied to time series data, see, e.g., Bauwens and Rombouts (2012).

Cognitive test data are often analysed using the normal distribution, see, e.g., Laukka et al. (2006). This may be problematic for many reasons. We illustrate this with the MMSE. If the normal distribution is used, then prediction of MMSE scores is not restricted to the original test scale and this can lead to interpretation problems when predicted scores are outside the scale 0–30. Ceiling effects further undermine the use of the normal distribution as these effects cause a dependency between residuals and fitted values which violate model assumptions. In the MMSE, a majority of observed sum scores in the range 28–30 is indicative of a ceiling effect.

The wider framework of our statistical modelling is that of random-effects growth models with a non-linear link between the response and the predictor, where the predictor is non-linear in the parameters. We propose change point

* Corresponding author. Tel.: +44 0 20 31083243; fax: +44 0 20 31083105.
E-mail address: ardo.vandenhout@ucl.ac.uk (A. van den Hout).

0167-9473/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2012.07.024

regression models with discrete probability distributions—appreciating the essential discrete nature of cognitive test data. Dependencies within the repeated measurements of an individual are dealt with by using random effects. In addition, we formulate a latent class model, which allows a priori for two latent groups in the data: one group where the cognitive process changes over time, and one group where the process is stable. For the first group a change point model is formulated. For both groups random-effects are included in the predictors. Special attention is given to residual diagnostics for model validation.

A common choice for the distribution of the random effects in regression models for longitudinal data is the multivariate normal (Rabe-Hesketh and Skrondal, 2009). As an alternative, the models in this paper assume a non-parametric distribution for the regression coefficients combined with a parametric distribution for the change point. Non-parametric maximum likelihood estimation of random effects in models with linear predictors has been discussed in Aitkin (1999), Molenberghs and Verbeke (2005), and Muthén and Asparouhov (2009). By adopting the non-parametric approach, the assumption of normality for the random effects is avoided, and optimizing the likelihood is computationally less demanding. The specification of the distribution of the random effects does not always have an impact on the estimation of the parameters of interest (Aitkin, 1999), but there are examples where the normality assumption leads to bias (Muthén and Asparouhov, 2009). The main advantage of the non-parametric approach is that it works well when the effects are normally distributed *and* when they are not. We extend the non-parametric approach to models with non-linear predictors. The choice of the parametric distribution for the change point is a truncated normal, which is specific to our application.

A general way to define a class of change point models is to assume a polynomial regression model of degree d_1 before the change point, and a polynomial regression model of degree d_2 after, see, e.g., Rudoy et al. (2010). The *broken-stick model* is a member of this class: there are two linear parts, one before and one after the change point, and continuity is imposed such that the linear parts intersect at the change point. The broken-stick model can also be described as a piecewise linear model with one free knot. It has been used in many applications, e.g. in AIDS research (Kiuchi et al., 1995), in social statistics (Cohen, 2008), and in medical statistics, (Hall et al., 2003; Muniz-Terrera et al., 2011).

Van den Hout et al. (2011) introduced a model where the two linear parts are bridged by a third-degree polynomial which induces a smooth transition between the parts. Similarities between this model and *bent-cable* regression as presented in Chiu et al. (2006) will be investigated. The class of models introduced by Bacon and Watts (1971) will also be considered. The current paper can be seen as a follow-up to Van den Hout et al. (2011) in the sense that we improved upon the choice of the change point predictor and its selection, and improve the modelling with respect to the distributional assumptions for the conditional response and the random effects.

In the application, change point models will be used to investigate features of cognitive change in the older population in the years before death. The modelling is tailored to the terminal decline hypothesis which states that individuals experience a change in the rate of decline of cognitive function before death (Riegel and Riegel, 1972). Where there is a decline, we are interested in the timing of the rate change, and in its shape. Longitudinal MMSE data are available from the Swedish OCTO-Twin study (McClearn et al., 1997). In this longitudinal study of aging (1991–2009), MMSE scores are recorded over time. Because almost all death times are available (94%) in this study, we assume that the effect of ignoring the data of the survivors is negligible and we analyse the data of those who died using years-to-death as the time scale.

Section 2 introduces the various change point models and choices for the conditional distribution. In Section 3, semi-parametric likelihood inference is discussed. Section 4 extends methodology to a latent class model that distinguishes a stable class versus a change class for cognitive function over time. In Section 5, data from the OCTO study are analysed. Section 6 concludes the paper.

2. Models

Given response variable Y, predictor η, link function $l()$, and time t as explanatory variable, the conditional mean of Y is given by $\mathbb{E}[Y|t] = l(\eta)$ with $\eta = h(t, \boldsymbol{\beta}, \tau)$, where $h()$ is the function that defines the predictor using coefficient vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)$ and change point τ.

The predictors in this section are non-linear in the change point parameter τ. Although the same notation for the regression coefficients is used for the various change point predictors, the interpretation of the coefficients varies across the models.

Extensions can be defined in a straightforward manner by including additional explanatory variables $\boldsymbol{x}$ to capture observed heterogeneity. In that case, $\eta = h(t, \boldsymbol{x}, \boldsymbol{\beta}, \tau)$.

The structure of the models in this section is similar to that of generalised non-linear random-effects models. The difference is that using the beta-binomial distribution for the response defines a model outside the natural exponential family, see Agresti (2002).

2.1. Predictors

The broken-stick model is given by

$$\eta_{BS} = h_{BS}(t, \boldsymbol{\beta}, \tau) = \begin{cases} \beta_0 + \beta_1 t & t < \tau \\ \beta_0 + \beta_1 \tau + \beta_2 (t - \tau) & t \geq \tau. \end{cases} \tag{1}$$

In this model the change is not smooth. As a function of t, there is no derivative of h_{BS} at t equal to τ.