Online Change-Point Detection in Categorical Time Series

Michael Höhle[1,2]
[1] Department of Statistics, Ludwig-Maximilians-Universität München, Germany
[2] Munich Center of Health Sciences, Munich, Germany
michael.hoehle@stat.uni-muenchen.de

This is an author-created preprint of the bookchapter which appeared in the book *Statistical Modelling and Regression Structures - Festschrift in Honour of Ludwig Fahrmeir* (2010), T. Kneib and G. Tutz (Eds.), Springer. pp. 377-397. The original publication is available as DOI 10.1007/978-3-7908-2413-1_20 from http://www.springerlink.com

Abstract

This contribution considers the monitoring of change-points in categorical time series. In its simplest form these can be binomial or beta-binomial time series modeled by logistic regression or generalized additive models for location, scale and shape. The aim of the monitoring is to online detect a structural change in the intercept of the expectation model based on a cumulative sum approach known from statistical process control. This is then extended to change-point detection in multicategorical regression models such as multinomial or cumulative logit models. Furthermore, a Markov chain based method is given for the approximate computation of the run-length distribution of the proposed CUSUM detectors. The proposed methods are illustrated using three categorical time series representing meat inspection at a Danish abattoir, monitoring the age of varicella cases at a pediatrist and an analysis of German Bundesliga teams by a Bradley-Terry model.

1 Introduction

In the year 2000 and as part of my Ph.D. project, I had the pleasurable experience of getting hold of a copy of Fahrmeir & Tutz (1994b) in my attempt of modeling a multivariate binomial time series of disease treatments in a pig farm. After some enquiries, I ended up implementing the extended Kalman filter approach described in Fahrmeir & Wagenpfeil (1997) and in Section 8.3 of Fahrmeir & Tutz (1994b). With the present contribution I take the opportunity to return to this problem from another point of view while at the same time honoring the work of Ludwig Fahrmeir. Specifically, the focus in this chapter is on monitoring time series with categorical regression models by statistical process control (SPC) methods.

A general introduction to SPC can be found in Montgomery (2005). Hawkins & Olwell (1998) give an in-depth analysis of the CUSUM chart, which is one commonly used SPC method. Detection based on regression charts with normal response can be found in the statistics and engineering literature (Brown, Durbin & Evans 1975, Kim & Siegmund 1989, Basseville & Nikiforov 1998, Lai 1995, Lai & Shan 1999). Generalized linear models based detectors are described in the literature for especially count data time series (Rossi, Lampugnani & Marchi 1999, Skinner, Montgomery & Runger 2003, Rogerson & Yamada 2004, Höhle & Paul 2008). For categorical time series, however, less development has been seen – with monitoring of a binomial proportion being the exception (Chen 1978, Reynolds & Stoumbos 2000, Steiner, Cook, Farewell & Treasure 2000). Retrospective monitoring of multinomial sequences is discussed in Wolfe &

Chen (1990). Prospective monitoring of multivariate discrete response variable imposes a great challenge.

The present work contains a novel adaptation of the likelihood ratio based cumulative sum (CUSUM) for the categorical regression context. Accompanying this CUSUM is a newly formulated approximate Markov chain approach for calculating its run-length distribution. Three examples are presented as illustration of the proposed categorical CUSUM: Meat inspection data from a Danish abattoir monitored by a beta-binomial regression model, disease surveillance by a multinomial logit model for the age distribution of varicella cases at a sentinel pediatrist, and finally – in honour of Fahrmeir & Tutz (1994a) – an analysis of paired comparison data for six teams playing in the best German national soccer league (1. Bundesliga). Fahrmeir & Tutz (1994a) analyzed the 1966/67–1986/87 seasons of this example using state-space methodology for categorical time series. My contribution continues their analysis up to the 2008/09 season with a special focus on change-point detection.

The structure of this chapter is as follows. Section 2 provides an introduction to modeling categorical time series while Section 3 contains the novel proposals for performing online change-point detection in such models. Application of the proposed methodology is given in Section 4. Section 5 closes the chapter with a discussion.

2 Modeling Categorical Time Series

Modeling categorical data using appropriate regression models is covered in Agresti (2002) or Fahrmeir & Tutz (1994b). The interest of this chapter lies in using such regression approaches for the modeling of time series with categorical response. Kedem & Fokianos (2002) and also Fahrmeir & Tutz (1994b) provide an introduction to this topic. A *categorical time series* is a time series where the response variable at each time point t takes on *one* of $k \geq 2$ possible categories. Let $\boldsymbol{X}_t = (X_{t1}, \ldots, X_{tk})'$ be a length k vector with X_{tj}, $j = 1, \ldots, k$, being one if the j'th category is observed at time t and zero otherwise. Consequently, $\sum_{j=1}^{k} X_{tj} = 1$. Assuming that a total n_t of such variables are observed at time t, define $\boldsymbol{Y}_t = \sum_{l=1}^{n_t} \boldsymbol{X}_{t,l}$ as the response of interest. Furthermore, assume that the distribution of $\boldsymbol{Y}_t$ can adequately be described by a multinomial distribution with time series structure, i.e.

$$\boldsymbol{Y}_t \sim \mathrm{M}_k(n_t, \boldsymbol{\pi}_t), \tag{1}$$

for $t = 1, 2, \ldots$, $\boldsymbol{\pi}_t = (\pi_{t1}, \ldots, \pi_{tk})'$ and $\sum_{j=1}^{k} \pi_{tj} = 1$ for all t. Here $\pi_{tj} = P(Y_t = j | \mathscr{F}_{t-1})$ is the probability for class j at time t and $\mathscr{F}_{t-1}$ denotes the history of the time series up to time $t-1$, i.e. just before but not including time t. When considering a single component $j \in \{1, \ldots, k\}$ of a multinomial distributed $\boldsymbol{Y}_t$, the resulting distribution of Y_{tj} is $\mathsf{Bin}(n_t, \pi_{tj})$. As a consequence, one strategy to describe a multinomial time series is to consider it as a set of independent binomial time series for each component. However, this ignores any correlations between the variables and does not provide a model with total probability 1.

2.1 Binomial and Beta-Binomial Data

The simplest form of categorical data is the case $k = 2$, which describes individuals experiencing an event or items as being faulty. In this case, the resulting distribution of Y_{t1} in (1) is $\mathsf{Bin}(n_t, \pi_{t1})$ while $Y_{t2} = n_t - Y_{t1}$. When modeling binomial data, interest is often in having an additional overdispersion not provided by the multinomial distribution. A parametric tool for such time series is the use of the beta-binomial distribution, i.e. $Y_t \sim \mathsf{BetaBin}(n_t, \pi_t, \sigma_t)$, where $t = 1, 2, \ldots$,

$0 < \pi_t < 1$ and $\sigma_t > 0$, and having probability mass function (PMF)

$$f(y_t|n_t,\pi_t,\sigma_t) = \left(\frac{\Gamma(n_t+1)\Gamma(y_t+1)}{\Gamma(n_t-y_t+1)}\right) \cdot \left(\frac{\Gamma(y_t+\frac{\pi_t}{\sigma_t}) \cdot \Gamma(\frac{1}{\sigma_t}) \cdot \Gamma(n_t+\frac{1-\pi_t}{\sigma_t}-y_t)}{\Gamma(n_t+\frac{1}{\sigma_t}) \cdot \Gamma(\frac{\pi_t}{\sigma_t}) \cdot \Gamma(\frac{1-\pi_t}{\sigma_t})}\right),$$

mean $E(Y_t) = n_t \cdot \pi_t$ and variance

$$\mathrm{Var}(Y_t) = n_t\pi_t(1-\pi_t)\left(1+(n_t-1)\frac{\sigma_t}{\sigma_t+1}\right).$$

In other words, σ_t is the dispersion parameter and for $\sigma_t \to 0$ the beta-binomial converges to the binomial distribution. Beta-binomial models can be formulated and fitted in the context of generalized additive models for location, scale and shape (GAMLSS, Rigby & Stasinopoulos (2005)). Here, the time varying proportion π_t is modeled by a linear predictor η_t on the logit-scale similar to binomial logit-modeling, i.e.

$$\mathrm{logit}(\pi_t) = \log\left(\frac{\pi_t}{1-\pi_t}\right) = \eta_t = \boldsymbol{z}_t'\boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{z}_t$ is a $p \times 1$ vector of covariates and $\boldsymbol{\beta}$ is a $p \times 1$ vector of covariate effects. Additionally, in a GAMLSS the dispersion can be modeled by a separate linear predictor $\log(\sigma_t) = \boldsymbol{w}_t'\boldsymbol{\gamma}$, but for notational and computational simplicity the dispersion is assumed to be time constant and not depending on covariates, i.e. $\sigma_t = \sigma$ for all t.

2.2 Nominal Data

In case the k groups of the response variable lack a natural ordering, i.e. in case of a nominal time series, one uses a multinomial logistic model with one of the categories, say category k, as reference:

$$\log\left(\frac{\pi_{tj}}{\pi_{tk}}\right) = \boldsymbol{z}_t'\boldsymbol{\beta}_j, \quad j = 1,\ldots,k-1.$$

As a result, the category specific probabilities can be computed as

$$\begin{aligned}\pi_{tj} &= \frac{\exp(\boldsymbol{z}_t'\boldsymbol{\beta}_j)}{1+\exp(\boldsymbol{z}_t'\boldsymbol{\beta}_j)}, \quad j = 1,\ldots,k-1, \text{ and}\\ \pi_{tk} &= \frac{1}{1+\sum_{j=1}^{k-1}\exp(\boldsymbol{z}_t'\boldsymbol{\beta}_j)}.\end{aligned}$$

Let $\boldsymbol{y}_{1:N} = (\boldsymbol{y}_1,\ldots,\boldsymbol{y}_N)$ denote the observed time series up to time N given as a $(m \times N)$ matrix, where each $\boldsymbol{y}_t = (y_{t1},\ldots,y_{tk})', t = 1,\ldots,N$ contains information on how the n_t observations fell into the k categories, i.e. $\sum_{j=1}^{k} y_{tj} = n_t$. The likelihood of the above model is given by

$$L(\boldsymbol{\beta};\boldsymbol{y}_{1:N}) = \prod_{t=1}^{N}\prod_{j=1}^{k}\pi_{tj}^{y_{tj}}(\boldsymbol{\beta}),$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1',\ldots,\boldsymbol{\beta}_{k-1}')'$. Statistical inference for the model parameters $\boldsymbol{\beta}$ based on this likelihood is described in detail in Fahrmeir & Tutz (1994b, Section 3.4) or Fokianos & Kedem (2003). Asymptotics for such categorical time series is studied in Kaufman (1987) and Fahrmeir & Kaufmann (1987).

2.3 Ordinal Data

If the k categories of the response variable can be considered as ordered, it is beneficial to exploit this additional information in order to obtain more parsimonious models. Denoting the categories of the ordered response variable by the ordered set $\{1,\dots,k\}$, a *cumulative model* described in, e.g., Fahrmeir & Tutz (1994b, p. 76) for the response at time t looks as follows

$$P(Y_t \leq j) = F(\theta_j + \mathbf{z}_t'\boldsymbol{\beta}), \quad j = 1,\dots,k,$$

with $-\infty = \theta_0 < \theta_1 < \ldots < \theta_k = \infty$ being the set of *threshold parameters*. When using the logistic distribution function $F(x) = \exp(x)/(1+\exp(x))$, the resulting model is called the *proportional odds model*, but in use are also other link functions such as the extreme-minimal-value distribution function. Consequently, the specific category probabilities can be derived as

$$\pi_{tj} = F(\theta_j + \mathbf{z}_t'\boldsymbol{\beta}) - F(\theta_{j-1} + \mathbf{z}_t'\boldsymbol{\beta}), \quad j = 1,\dots,k.$$

2.4 Paired Comparisons

One application of the proportional odds model is the analysis of paired-comparison data used to determine preference or strength of items. Such data are typical in sports like chess or tennis, where world rankings of m players are based on pairwise comparisons having categorical outcomes (e.g. win, loose). Other areas of application are consumer preference, sensory studies and studies of animal behavior (Courcoux & Semenou 1997, Bi 2006, Whiting, Stuart-Fox, O'Connor, Firth, Bennett & Blomberg 2006). The basic Bradley-Terry model (Bradley & Terry 1952) is a logistic regression model quantifying the probability of a positive outcome (i.e. winning) for the first mentioned player in a match of two players. Each player $i \in \{1,\dots,m\}$ has ability or strength $\alpha_i \in \mathbb{R}$, and the probability that a match between the i'th and j'th player results in a win for player i is given by

$$\text{logit}\{P(Y_{ij} = 1)\} = \alpha_i - \alpha_j.$$

In the above, Y_{ij} is a binary random variable with states 1 (i wins) and 2 (i looses). As a consequence, $P(Y_{ij} = 1) = 1/2$ if $\alpha_i = \alpha_j$ and $P(Y_{ij} = 1) > 1/2$ if $\alpha_i > \alpha_j$. To ensure identifiably, one has to impose a constraint such as $\alpha_m = 0$ or $\sum_{i=1}^m \alpha_i = 0$ on the $\alpha's$. Extensions of the Bradley-Terry model consist of letting strength be given by additional covariates such as home court advantages, age or injuries (Agresti 2002). Another common extension is to handle additional tied outcomes or even more complicated ordinal response structure (Tutz 1986).

If the time interval over which the paired-comparisons are performed is long, one might expect the abilities of players to change over time (Fahrmeir & Tutz 1994b, Glickman 1999, Knorr-Held 2000). Following Fahrmeir & Tutz (1994a), a general time-dependent ordinal paired-comparison model including covariates can be formulated as

$$P(Y_{tij} = r) = F(\theta_{tr} + \alpha_{ti} - \alpha_{tj} + \mathbf{z}_{tij}'\boldsymbol{\beta}_t) - F(\theta_{t,r-1} + \alpha_{ti} - \alpha_{tj} + \mathbf{z}_{tij}'\boldsymbol{\beta}_t), \tag{3}$$

with $r = 1,\dots,k$ being the category, $t = 1,2,\dots$ denoting time and $i,j = 1,\dots,m$ being the players compared. For example in the application of Section 4.3, Y_{tij} will denote paired comparisons of six teams within each season of the best German national soccer league (1. Bundesliga). In what follows, I will assume time constant covariate effects $\boldsymbol{\beta}_t = \boldsymbol{\beta}$ for all t and similar time constant thresholds $\boldsymbol{\theta}_t = (\theta_{t0},\dots,\theta_{tk})' = \boldsymbol{\theta} = (\theta_0,\dots,\theta_k)'$ for all t.

After having presented the basic modeling techniques, the focus is now on the online detection of changepoints in such models.

3 Prospective CUSUM Changepoint Detection

The cumulative sum (CUSUM) detector is a method known from statistical process control for online detecting structural changes in time series. An overview of the method can be found in Hawkins & Olwell (1998). Use of the method for count, binomial or multicategorical time series using regression models is still a developing field. Höhle & Paul (2008) treats one such approach for count data and Grigg & Farewell (2004) provide an overview. For multicategorical time series Topalidou & Psarakis (2009) contains a survey of existing monitoring approaches. My interest is in monitoring a time varying vector of proportions $\boldsymbol{\pi}_t$ in a binomial, beta-binomial or multinomial setting having time-varying n_t. Regression models for categorical time series provide a versatile modeling framework for such data allowing for time trends with seasonality and possible covariate effects. Sections 3.1–3.3 contain my proposal for combining CUSUM detection with categorical time series analysis.

Let $f(\boldsymbol{y}_t;\boldsymbol{\theta})$ denote the PMF of the response variable at time t. While new observations arrive, the aim is to detect as quickly as possible if the parameters of f have changed from the in-control value of $\boldsymbol{\theta}_0$ to the out-of-control value $\boldsymbol{\theta}_1$. Following Frisén (2003), define the *likelihood ratio based CUSUM statistic* as

$$C_s = \max_{1\leq t\leq s}\left[\sum_{i=t}^{s}\log\left\{\frac{f(\boldsymbol{y}_i;\boldsymbol{\theta}_1)}{f(\boldsymbol{y}_i;\boldsymbol{\theta}_0)}\right\}\right], \quad s=1,2,\ldots. \tag{4}$$

Given a fixed threshold $h>0$, a change-point is detected at the first time s where $C_s>h$, and hence the resulting stopping time S is defined as

$$S = \min\{s\geq 1 : C_s > h\}. \tag{5}$$

At this time point, enough evidence is found to reject $H_0 : \boldsymbol{\theta}=\boldsymbol{\theta}_0$ in favor of $H_1 : \boldsymbol{\theta}=\boldsymbol{\theta}_1$. Let now $LLR_t = \log f(y_t;\boldsymbol{\theta}_1) - \log f(y_t;\boldsymbol{\theta}_0)$ be shorthand for the loglikelihood ratio at time t in (4). If $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$ are known, (4) can be written in recursive form

$$C_0 = 0 \quad \text{and} \quad C_s = \max\left(0, C_{s-1}+LLR_t\right), \quad \text{for } s\geq 1. \tag{6}$$

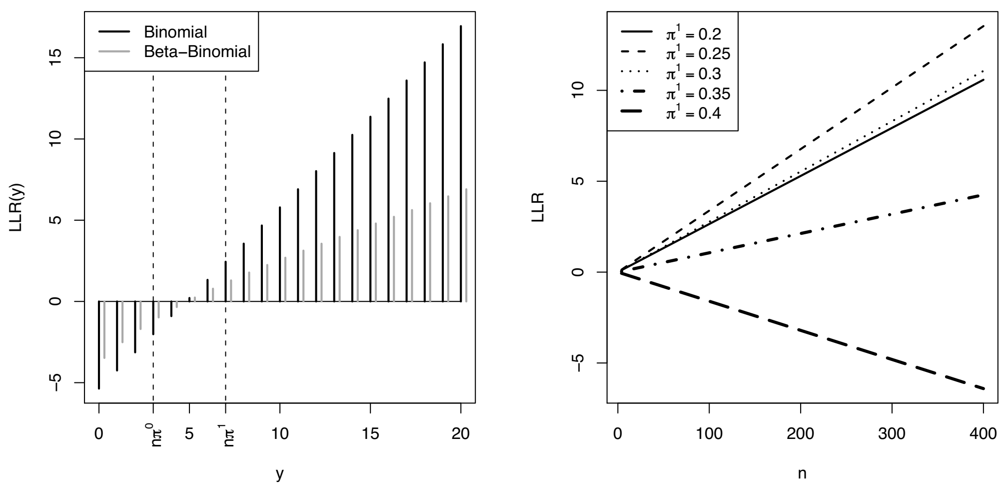
One sees that for time points with $LLR_t>0$, i.e. evidence against in-control, the LLR_t contributions are added up. On the other hand, no credit in the direction of the in-control is given because C_s cannot get below zero.

In practical applications, the in-control and out-of-control parameters are, however, hardly ever known beforehand. A typical procedure in this case is to use historical *phase 1 data* for the estimation of $\boldsymbol{\theta}_0$ with the assumption that these data originate from the in-control state. This estimate is then used as plug-in value in the above CUSUM. Furthermore, the out-of-control parameter $\boldsymbol{\theta}_1$ is specified as a known function of $\boldsymbol{\theta}_0$, e.g. as a known multiplicative increase in the odds. Using categorical regression to model the PMF f as a function of time provides a novel use of statistical process control for monitoring categorical time series. Sections 3.1–3.3 discuss monitoring in case of beta-binomial, multinomial and ordered response. Section 3.4 contains a corresponding method to compute the important run-length distribution of the different CUSUM proposals.

3.1 Binomial and Beta-Binomial CUSUM

Extending the work of Steiner et al. (2000) to a time varying proportion, the aim is to detect a change from odds $\pi_t^0/(1-\pi_t^0)$ to odds $R\cdot\pi_t^0/(1-\pi_t^0)$ for $R>0$, i.e. let

$$\text{logit}(\pi_t^1) = \text{logit}(\pi_t^0) + \log R. \tag{7}$$



(a) (b)

Figure 1: (a) Loglikelihood ratio (LLR) as a function of y for a binomial distribution with $n = 20$ and $\pi^0 = 0.15$ and $\pi^1 = 0.35$. Also shown are the same LLRs for the corresponding beta-binomial distribution with $\sigma = 0.05$. (b) Binomial LLR as a function of n when $y = 0.25 \cdot n$, $\pi^0 = 0.15$ and when comparing against four different π^1.

In other words, let $\text{logit}(\pi_t^1) = \text{logit}(\pi_t^0) + \log R$ correspond to such a change in the intercept of the linear predictor in (2). The change-point detection is thus equivalent to a detection from the in-control proportion π_t^0 to the out-of-control proportion π_t^1 in (6) using the beta-binomial PMF as f.

Figure 1(a) illustrates the LLR as a function of the number of positive responses in a binomial distribution for one specific time point (note that t is dropped from the notation in this example). Starting from $y = 5$ one has $LLR > 0$, i.e. observations with $y \geq 5$ contribute evidence against the null-hypothesis and in favor of the alternative hypothesis. Note also, that the beta-binomial distribution has smaller LLR contributions because the variance of the distribution is larger than for the binomial distribution. Similarly, Figure 1(b) shows that the larger n the larger is the LLR contribution of the observation $y = 0.25 \cdot n$. In other words, the greater n is the more evidence against $H_0 : \pi = 0.15$ there is from an empirical proportion of 0.25. This is of interest in a binomial CUSUM with time varying n_t: the relevance (as measured by its contribution to C_t) of a large proportion of faulty items thus depends on the number of items sampled. However, for $\pi^1 = 0.4$ the value $y = 0.25 \cdot n$ does not provide evidence against H_0 in Figure 1(b). This means that for large out-of-control proportions the observation $y = 0.25 \cdot n$ results in negative LLRs and hence speaks in favor of H_0.

At time t and given the past value of the CUSUM statistic C_{t-1}, the minimum number of cases necessary to reach the threshold h at time t is

$$a_t = \min_{y \in \{0,\ldots,n_t\}} \left\{ \text{LLR}(y; n_t, \pi_t^0, \pi_t^1, \sigma) > h - C_{t-1} \right\}. \qquad (8)$$

Note that the set of y fulfilling the above inequality can be empty, in this case a_t does not exists. If a_t exists, the solution of (8) can be derived explicitly for the binomial case as

$$a_t = \max\left\{0, \left\lceil \frac{h - C_{t-1} - n_t \cdot (\log(1-\pi_t^1) - \log(1-\pi_t^0))}{\log(\pi_t^1) - \log(\pi_t^0) - \log(1-\pi_t^1) + \log(1-\pi_t^0)} \right\rceil \right\}.$$

In the beta-binomial case the solution has to be found numerically, e.g. by trying possible $y \in \{0,\dots,n_t\}$ until the first value fulfills the inequality.

3.2 Multinomial CUSUM

This section looks at generalization of the previous binomial CUSUM to the multinomial distribution $M_k(n_t, \boldsymbol{\pi}_t)$ for $k > 2$, and where $\boldsymbol{\pi}_t$ is modeled by multinomial logistic regression. Let $\boldsymbol{\pi}_t^0$ be the in-control probability vector and $\boldsymbol{\pi}_t^1$ the out-of-control probability vector resulting from the models with parameters $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. A simple approach would be to monitor each of the k components separately using the methodology from Section 3.1. However, this would ignore correlations between the measurements with reduced detection power as consequence. Instead, I consider detection as the task of investigating change-points in the linear predictors of the multicategorical logit model. The proposed approach extends the work of Steiner, Cook & Farewell (1999), who monitored surgical performance of a $k = 4$ outcome using two paired binomial CUSUMs with time-constant means.

Based on a multicategorical logit model, let the in-control probabilities for the non-reference categories be

$$\log\left(\frac{\pi_{tj}^0}{\pi_{tk}^0}\right) = \boldsymbol{z}_t' \boldsymbol{\beta}_j, \quad j = 1,\dots,k-1.$$

As for the binomial CUSUM, the out-of-control probabilities are given by specific changes in the intercept of this model, i.e.

$$\log\left(\frac{\pi_{tj}^1}{\pi_{tk}^1}\right) = \log\left(\frac{\pi_{tj}^0}{\pi_{tk}^0}\right) + \log(R_j), \quad j = 1,\dots,k-1.$$

Figure 2 illustrates the approach for a $\boldsymbol{Y} \sim M_3(20, \boldsymbol{\pi}^0)$ distribution with $\boldsymbol{\pi}^0 = (0.22, 0.17, 0.61)'$ and $\log(\boldsymbol{R}) = (1.30, 1.10)'$. One observes that many states with high LLR are concurrently very unlikely and that for larger n or k, the approximating multivariate Gaussian distribution can be used to determine states with high enough probability to investigate its LLR.

If the number of possible categories k of the multinomial is very high, log-linear models provide an alternative as done by Qiu (2008). However, in his work time-constant problems are dealt with and the prime goal is to detect a shift in the median of any component without a specific formulation of the alternative. However, a suitable extension of the proposed monitoring in this chapter might be to monitor against an entire set of possible out-of-control models with the different $\boldsymbol{R}$'s specifying different directions.

3.3 Ordinal and Bradley-Terry CUSUM

The multinomial CUSUM proposal from the previous section can be used as a change-point detection approach for ordinal time series: Based on the proportional odds model to generate the in-control and out-of-control proportions. In particular, this approach is considered for the time varying Bradley-Terry model (3) from Section 2.4. Let $\boldsymbol{Y}_t = (Y_{tij}\,; i = 1,\dots,m, j = 1,\dots,m, i \neq j)$ consist of all $K = m \times (m-1)$ paired comparisons occurring at time t, i.e. $\boldsymbol{Y}_t \in \{1,\dots,k\}^K$. Given the parameters of a time varying Bradley-Terry model, the probability of a state $\boldsymbol{Y}_t = \boldsymbol{y}_t$ can thus be computed as

$$f(\boldsymbol{y}_t\,; \boldsymbol{\alpha}_t, \boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^{m} \prod_{j=1, i\neq j}^{m} f(y_{tij}\,; \boldsymbol{\alpha}_t, \boldsymbol{\beta}, \boldsymbol{\theta}),$$

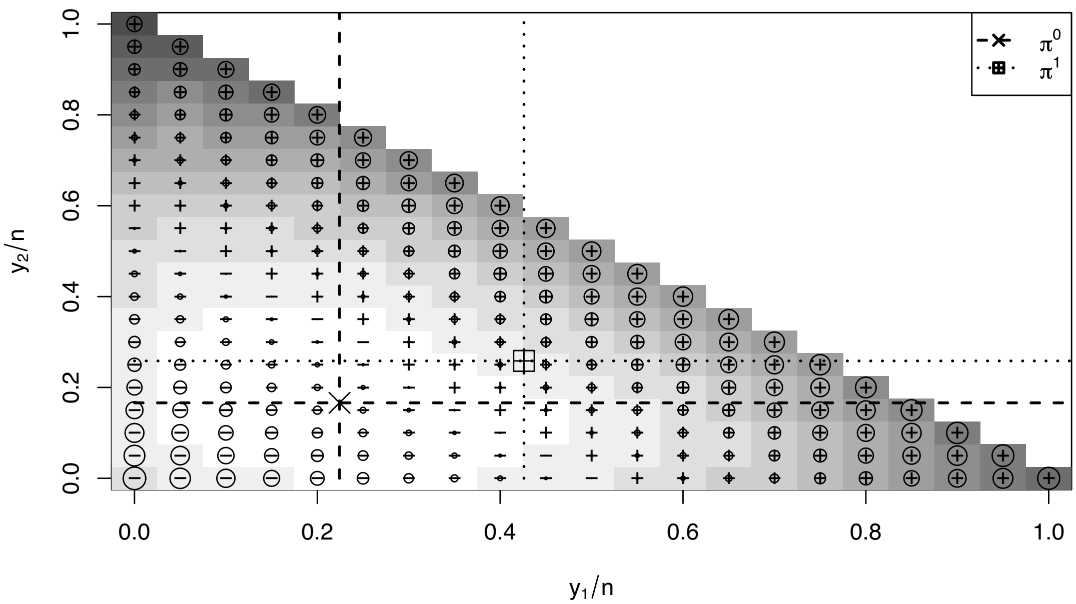


Figure 2: Illustration of the LLR for a $M_3(20, \boldsymbol{\pi})$ multinomial CUSUM with $\boldsymbol{\pi}^0 = (0.22, 0.17, 0.61)'$ and $\boldsymbol{\pi}^1 = (0.43, 0.26, 0.32)'$. Shown are the first two components y_1 and y_2 of each possible state $\boldsymbol{y}$. Circle sizes indicate magnitude and $\pm$ the sign of the LLR. Also shown are the in-control and out-of-control probabilities. Shading indicates the probability of $\boldsymbol{y}$ in a model with $\boldsymbol{\pi} = \boldsymbol{\pi}^0$ – the whiter the cell the higher is the probability of the corresponding state.

where $f(\cdot)$ denotes the PMF given in (3). The interest is now on detecting a structural change in the ability of one or several teams, i.e. $\boldsymbol{\alpha}_t^1 = \boldsymbol{\alpha}_t^0 + \boldsymbol{R}$, where $\boldsymbol{R}$ is a vector of length m with for example one component being different from zero. The LLR in a corresponding CUSUM detector can then be computed as

$$LLR_t = \sum_{i=1}^{m} \sum_{j=1, i\neq j}^{m} \log \frac{f(\boldsymbol{y}_{tij}; \boldsymbol{\alpha}_t^1, \boldsymbol{\beta}, \boldsymbol{\theta})}{f(\boldsymbol{y}_{tij}; \boldsymbol{\alpha}_t^0, \boldsymbol{\beta}, \boldsymbol{\theta})}. \tag{9}$$

3.4 Run-length of Time Varying Categorical CUSUM

The distribution of the stopping time S in (5) for the CUSUMs proposed in sections 3.1–3.3 when data are sampled from either $\boldsymbol{\pi}_t^0$ or $\boldsymbol{\pi}_t^1$ is an important quantity to know when choosing the appropriate threshold h. Specifically, the expected run length $E(S)$ (aka. the average run length (ARL)), the median run length or the probability $P(S \leq s)$ for a specific $s \geq 1$ are often used summaries of the distribution and can be computed once the PMF of S is known. Let $\boldsymbol{\theta}$ be the set of parameters in the multicategorical regression model and let $\boldsymbol{\pi}$ be the resulting proportions under which the distribution of S is to be computed. For example, the above $\boldsymbol{\theta}$ is equal to $\boldsymbol{\theta}_0$ if the in-control ARL is of interest.

Brook & Evans (1972) formulated an approximate approach based on Markov chains to determine the PMF of the stopping time S of a time-constant CUSUM detector. They describe the dynamics of the CUSUM statistic C_t by a Markov chain with a discretized state space of size $M+2$:

$$\begin{array}{ll} \text{State } 0: & C_t = 0 \\ \text{State } i: & C_t \in \left((i-1)\cdot \frac{h}{M}, i \cdot \frac{h}{M}\right],\ i = 1,2,\ldots,M \\ \text{State } M+1: & C_t > h \end{array}$$

Note that state $M+1$ is absorbing, i.e. reaching this state results in H_0 being rejected, and therefore no further actions are taken. The discretization of the continuum of values of the CUSUM statistic into a discrete set of states represents an approximation. The size of M controls the quality of the approximation. Adopting this approach to the present time-varying context, let $\boldsymbol{P}_t$ be the $(M+2)\times(M+2)$ transition matrix of $C_t|C_{t-1}$, i.e.

$$p_{tij} = P(C_t \in \text{State } j | C_{t-1} \in \text{State } i), \quad i,j = 0,1,\ldots,M+1$$

Let $a < b$ and $c < d$ represent the lower and upper limits of class j and i, respectively. To operationalize the Markov chain approach one needs to compute

$$p_{t,i,j} = P(a < C_t < b | c < C_{n-1} < d) = \int_c^d \{F_t(b-s) - F_t(a-s)\}\, d\mu(s), \tag{10}$$

where $\mu(x)$ is the unknown distribution function of C_{t-1} conditional on $c < C_{t-1} < d$ and $F_t(\cdot)$ is the distribution function of the likelihood ratio LLR_t at time t when $\boldsymbol{y}_t$ is distributed according to a multinomial distribution with parameters derived from a categorical regression model with parameters $\boldsymbol{\theta}$. Investigations in Hawkins (1992) for the homogeneous case suggest using the uniform distribution for measure $\mu(x)$. Furthermore, he suggests using Simpson's quadrature rule with midpoint $m = (c+d)/2$ to approximate the integral in (10) instead of the Riemann integral used in Brook & Evans (1972). Altogether, Hawkins (1992) adapted to the present time varying case yields

$$\begin{aligned} P(a < C_t < b | c < C_{n-1} < d) \approx & \frac{1}{6}\{F_t(b-c) + 4F_t(b-m) + F_t(b-d)\} \\ & - \frac{1}{6}\{F_t(a-c) + 4F_t(a-f) + F_t(a-d)\}. \end{aligned}$$

Specifically, $F_t(\cdot)$ can be computed for the categorical CUSUM by computing the likelihood ratio of all valid configurations $\boldsymbol{y}_t \in \{0,1,\ldots,n_t\}^k, \sum_{j=1}^k y_{tj} = n_t$, together with the probability $P(\boldsymbol{Y}_t = \boldsymbol{y}_t)$ of its occurrence under $\boldsymbol{\theta}$. However, if n_t or k is large, this enumeration strategy can quickly become infeasible and one would try to identify relevant states with $P(\boldsymbol{y}) > \varepsilon$ and approximate $F_t(\cdot)$ by only considering these states in the computations. One strategy to perform this identification could be to compare with the approximating normal distribution.

Borrowing ideas from Bissell (1984), the cumulative probability of an alarm at any step up to time s, $s \geq 1$, is

$$P(S \leq s) = \left[\prod_{t=1}^{s} \boldsymbol{P}_t\right]_{0,M+1},$$

i.e. the required probability is equivalent to the probability of going from state zero at time one to the absorbing state at time s as determined by the s-step transition matrix of the Markov chain. The PMF of S can thus be determined by $P(S=s) = P(S \leq s) - P(S \leq s-1)$, where for $s=1$ one defines $P(S=0)=0$. Hence, $E(S)$ can be computed by the usual expression $\sum_{s=1}^{\infty} s \cdot P(S=s)$. In practice, one would usually compute $P(S \leq s)$ only up to some sufficiently large $s = s_{\max}$ such that $P(S \leq s) \geq 1-\varepsilon$ for a small ε. This results in a slightly downward bias in the derived ARL. If the Markov chain is homogeneous, then the ARL can alternatively be computed as the first element of $(\boldsymbol{I}-\boldsymbol{R})^{-1} \cdot \mathbf{1}$, where $\boldsymbol{R}$ is obtained from $\boldsymbol{P}$ by deleting the last row and column, $\boldsymbol{I}$ is the identity matrix and $\mathbf{1}$ a vector of ones.

In practice, covariates or n_t are usually not available for future time points. As the predicted in-control and out-of-control probabilities are conditional on these values, it is more practicable to compute $P(S \leq s)$ for phase 2 data where the covariates already have been observed instead of trying to impute them for future time points.

4 Applications

The following three examples illustrate the use of the proposed CUSUM monitoring for categorical time series by applications from veterinary quality control, human epidemiology and – as continuation of Fahrmeir & Tutz (1994b) – sports statistics.

4.1 Meat Inspection

At Danish abattoirs, auditing is performed for each processed pig in order to provide guarantees of meat quality and hygiene and as part of the official control on products of animal origin intended for human consumption (regulated by the European Council Regulation No 854/2004). Figure 3 shows the time series of the weekly proportion of positive audit reports for a specific pig abattoir in Denmark. Reports for a total of 171 weeks are available with monitoring starting in week 1 of 2006.

Using the data of the first two years as phase 1 data, a beta-binomial model with intercept and two sinusoidal components for $\text{logit}(\pi_t)$ is estimated using the R function `gamlss` (Rigby & Stasinopoulos 2005). These estimated values are then used as plug-in values in the model to predict π_t^0 for phase 2. The out-of-control π_t^1 is then defined by specifying $R=2$ in (7), i.e. a doubling in the odds of a positive audit report is to be detected as quickly as possible. Figure 4 shows the results from this monitoring. After the first change-point is detected the CUSUM statistic is set to zero and monitoring is restarted.

Figure 5 displays the run length distribution when using $h=4$ by comparing the Markov chain approximation using $M=5$ with the results of a simulation based on 10000 runs. Note that the Markov chain method provides results much faster than the simulation approach.

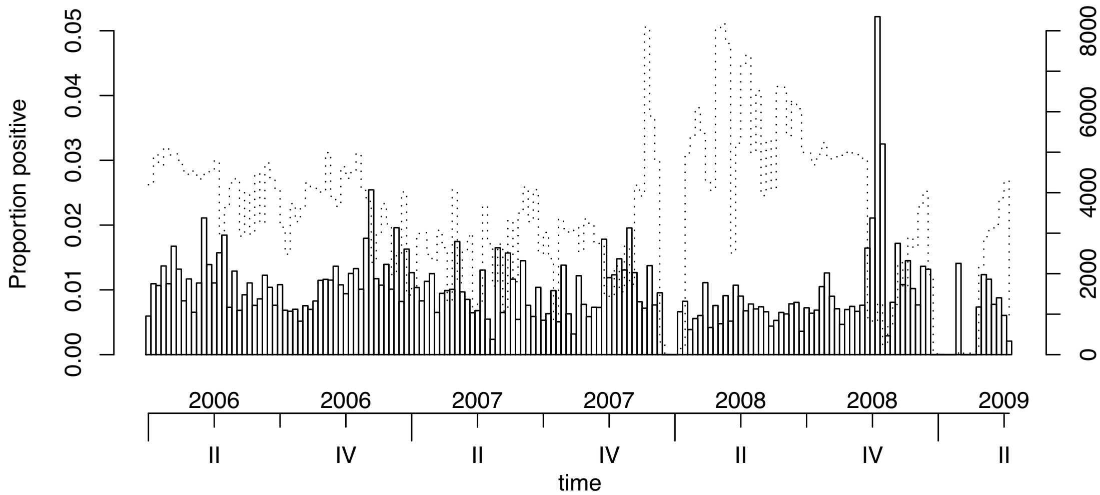


Figure 3: Weekly proportion y_t/n_t of pigs with positive audit reports indicated by bars (scale on the left axis). The dotted line shows the weekly total number of pigs n_t (scale as on right axis). Roman letters denote quarters of the year.

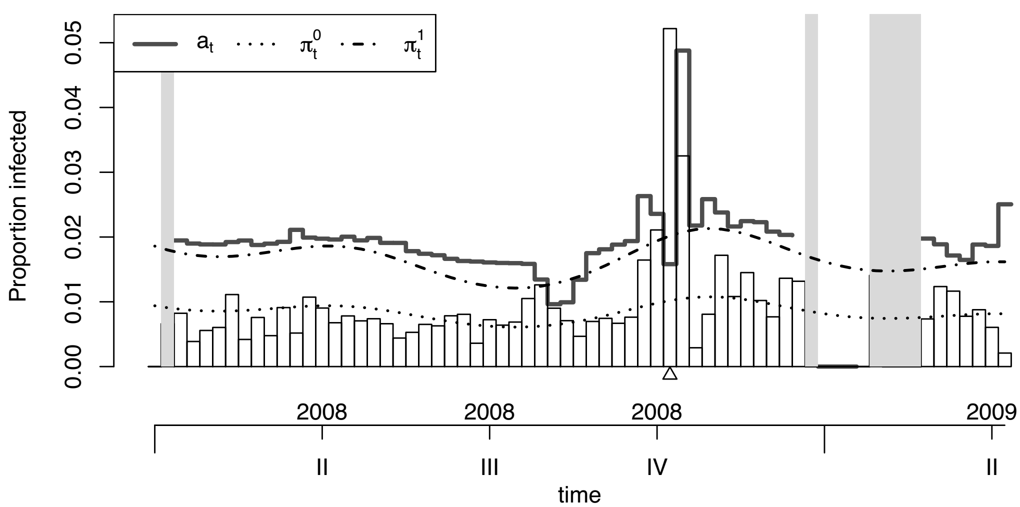


Figure 4: Results of beta-binomial CUSUM monitoring for phase 2. Shaded bars indicate weeks where $n_t < 200$. The triangle indicates the alarm in week 41 of 2008.

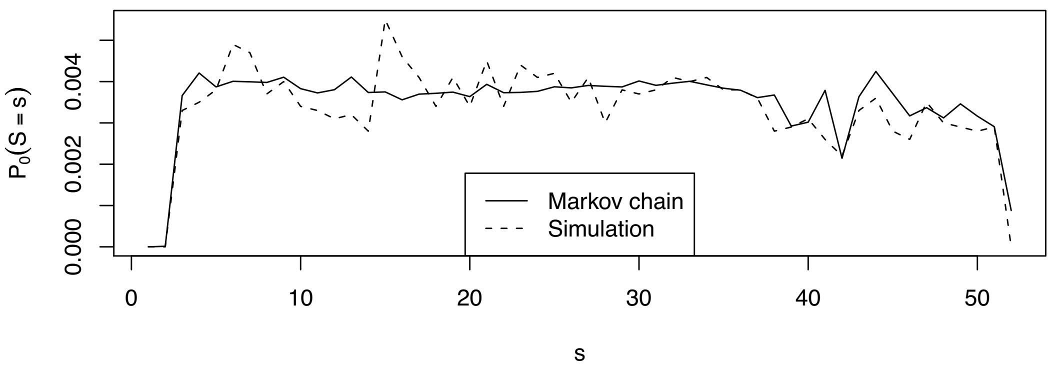


Figure 5: Comparison of the in-control run-length PMF $P_0(S = s)$ between the Markov chain method and a simulation based on 10000 samples.

4.2 Agegroups of Varicella Cases

A varicella sentinel was established in April 2005 by the *Arbeitsgemeinschaft Masern und Varizellen* (Robert Koch Institute 2006) to monitor a possible decline in the number of monthly varicella after the introduction of a vaccination recommendation. One particular point of interest is the monitoring of possible shifts in the age distribution of the cases. This is done by dividing the age of cases into one of five groups: <1, 1-2, 3-4, 5-9, and >9 years. A shift in the age distribution is now defined to be a structural change in the proportions $\boldsymbol{\pi}$ controlling which of the five age groups a case falls into. As proof of concept of the proposed methodology, the time series of a single pediatrist participating in the sentinel is considered. Figure 6 shows the time series of monthly proportions across the five age groups – note that summer vacations result in a seasonal pattern. Using the first 24 months as phase 1 data, a multinomial logistic model using intercept, linear time trend and two seasonal components is fitted by the R function `multinom` (Venables & Ripley 2002). Figure 7 shows the fitted model and the resulting in-control proportions for the five age groups for the subsequent 18 months.

Applying the proposed categorical CUSUM based on the multinomial PMF with the age group 1-2 acting as reference category, one obtains Figure 8. From an epidemiological point of view it is in the 1-2 age group where a decline of cases is expected because primarily this group is vaccinated. Detecting an increase in the remaining four groups is one way to identify such a shift. As a consequence, $\log(\boldsymbol{R}) = (1,1,1,1)'$ is used. Figure 8 shows the resulting C_t together with the two detected change-points.

The threshold $h = 2.911$ is selected such that $P_0(S \leq 18) = 0.058$ as computed by the Markov chain approach with $M = 25$. By simulation of the run-length using 10000 runs, one obtains $P_0(S \leq 18) = 0.060$. To get an understanding of the consequences of currently ignored estimation error for the phase 1 parameters, a parametric bootstrap investigation is performed. Let $\widehat{\boldsymbol{\theta}}_0$ represent the estimated phase 1 parameters. In the b'th bootstrap sample, simulate new data phase 1 data $\boldsymbol{y}_{t,b}, t = 1,\ldots,24$ by sampling from a multinomial model with probabilities derived from $\hat{\boldsymbol{\theta}}_0$. Then use this $\boldsymbol{y}_{t,b}$ to estimate the phase 1 parameters $\widehat{\boldsymbol{\theta}}_{0,b}$ and derive $\boldsymbol{\pi}_b^0$ and $\boldsymbol{\pi}_b^1$ from $\widehat{\boldsymbol{\theta}}_{0,b}$ for phase 2. Now use the Markov chain procedure to compute $P_{0,b}(S \leq 18)$. A 95% percentile bootstrap interval for $P_0(S \leq 18)$ based on 100 bootstrap replications is $(0.013, 0.070)$, which emphasizes the effect of estimation error on the run length properties.

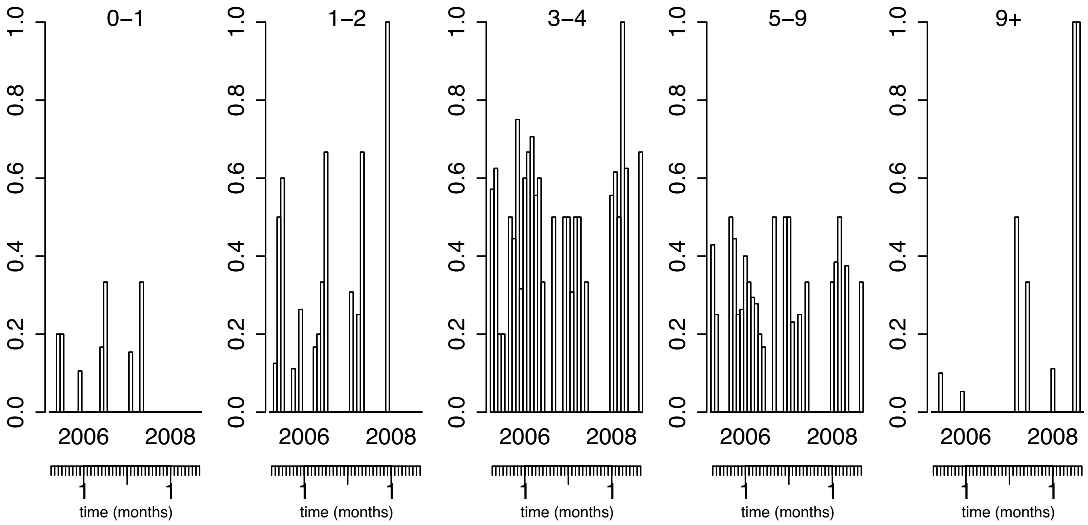


Figure 6: Multinomial time series of monthly cases at a pediatrist participating in the varicella sentinel surveillance. The values of n_t range from 0 to 19.

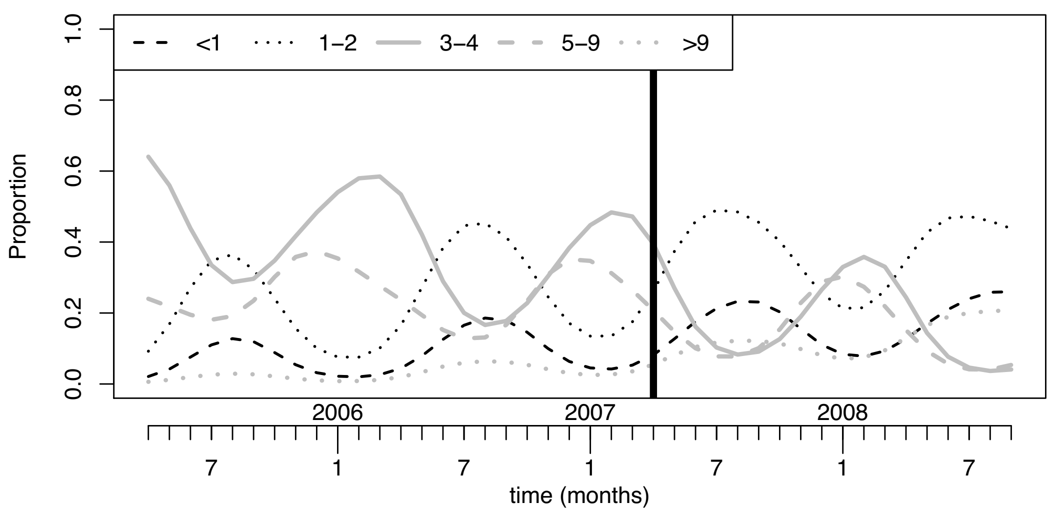


Figure 7: Age group proportions obtained from fitting a multicategorical logit model to the observed phase 1 data (to the left of the vertical bar). Also shown are the resulting predictions for $\boldsymbol{\pi}^0$ during phase 2 (starting from the vertical bar).

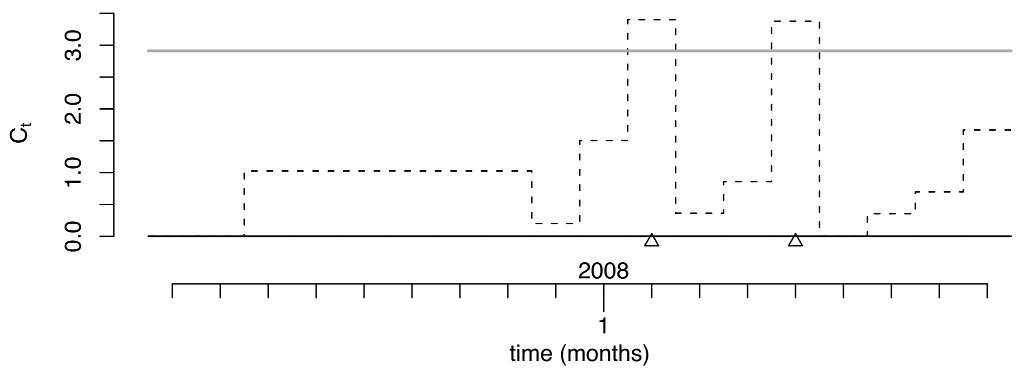


Figure 8: CUSUM statistic C_t for the pediatrist data together with the threshold $h = 2.911$. Triangles indicate detected change-points.

4.3 Strength of Bundesliga Teams

The time series analysed in this section contains the paired comparison data for a subset of six teams playing in the best German national soccer league (1. Bundesliga) as described in Fahrmeir & Tutz (1994a). For each of the 44 seasons from 1966/67–2008/09, all teams play against each other twice – once with the first team having home-court advantage and once with the second team having this advantage. Conceptually, it would have been feasible to perform the comparison based on all teams having played in the primary division since 1965/66, but I conduct the analysis in spirit of Fahrmeir & Tutz (1994a) by using only six teams. Each match has one of three possible outcomes: home team wins, tie and away team wins. In what follows, the ability of each team is assumed constant within the season but varies from season to season, i.e. α_{it} denotes the ability of team i in season $t = 1, \ldots, 44$.

Figure 9 shows the resulting abilities of each team as determined by a Bradley-Terry model fitted using the `vglm` function from package `VGAM` (Yee & Wild 1996, Yee 2008). The team VfB Stuttgart is selected as reference category with $\alpha_{3t} = 0$ for all t. For each team a time trend is modeled by a cubic B-spline with five equidistant interior knots and an intercept, i.e. $\alpha_{it} = f_i(t) = \beta_{i0} + \sum_{k=1}^{8} \beta_{ik} B_k(t)$. This model was found to be the model with equidistant knots minimizing Akaike's information criterion. Seasons where a team did not play in the first division are indicated in Figure 9 by missing abilities for that particular season.

The estimated abilities only reflect strengths based on the six selected teams. Hence, they do not necessarily reflect the overall strength of the team that season, which explains for example the somewhat weak ability of 1. FC Kaiserslautern in the 1990/91 season where they won the cup. Fahrmeir & Tutz (1994a), with their state space approach also noted the drop for FC Bayern München around 1976-1980, which was due to Franz Beckenbauer leaving the club.

From a sports manager perspective, it could be of interest to online monitor the ability of a team for the purpose of performing strategic interventions. Applying the methodology from Section 2.4, consider the case of monitoring the ability of FC Bayern München starting from year 1990. A Bradley-Terry model with an intercept only is fitted to the data before 1990 and a change of $\boldsymbol{R} = (-0.5, 0, 0, 0, 0)'$ is to be detected for the abilities of all teams except the reference team. Figure 10 illustrates both the abilities obtained from fitting the phase 1 data and the resulting predicted out-of-control abilities for phase 2. The aim is to detect when the strength of FC Bayern München drops by 0.5 units compared to the average strength of 0.741 during the 1965/1966 to 1988/1989 seasons. This means that the probability of winning against VfB

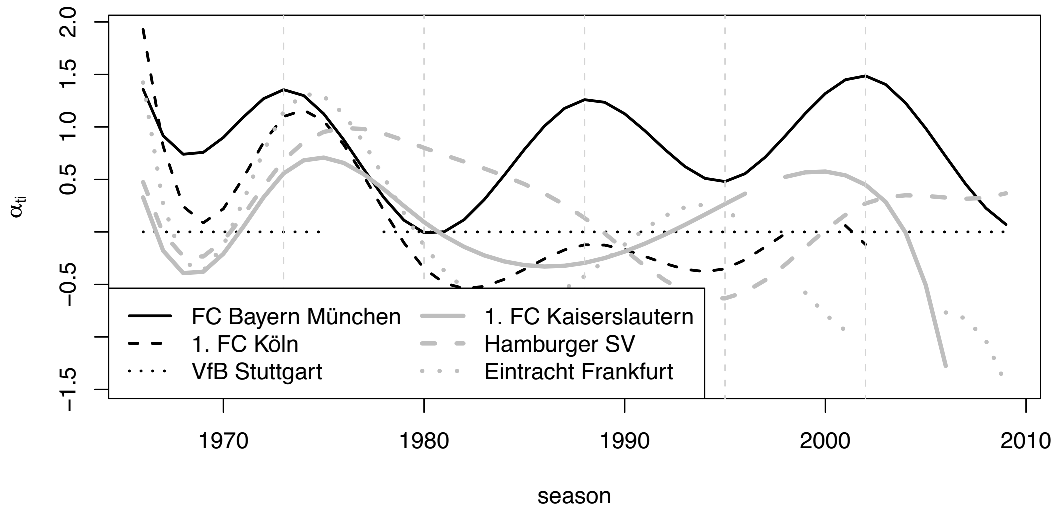


Figure 9: Abilities α_{ti} of each team as fitted by a proportional odds model described in the text. Seasons without comparisons are indicated by not plotting the ability for this season. The thin shaded lines indicate the know locations.

Stuttgart at home court drops from 0.742 to 0.636 ($\widehat{\theta}_1 = 0.315$).

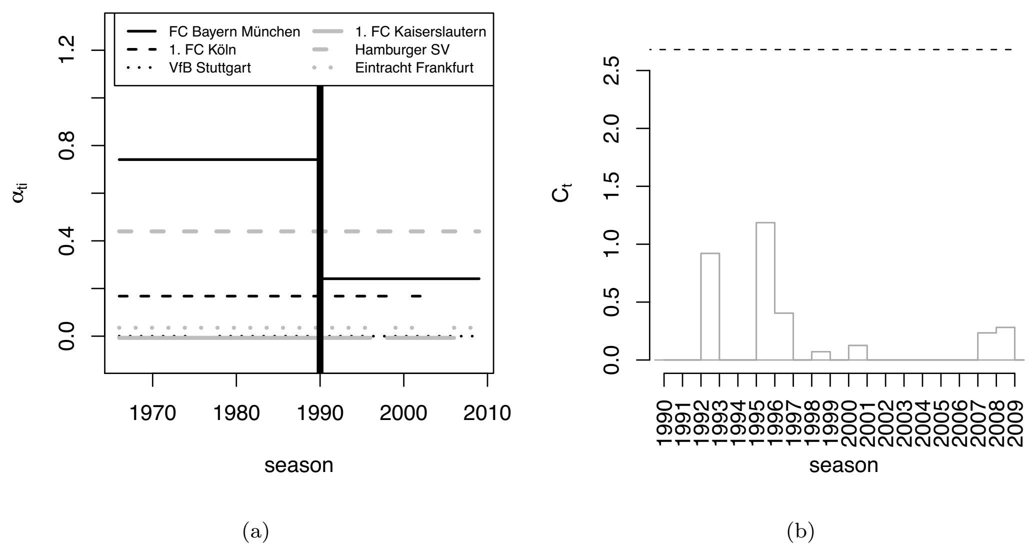


Figure 10: (a) In-control abilities fitted from phase 1 data (before the vertical bar). Also shown are the out-of-control abilities for phase 2 (starting at the vertical bar) obtained by prediction from the phase 1 fitted model. (b) CUSUM statistics when monitoring using the in-control and out-of-control abilities from (a) for phase 2. The upper line shows the threshold h.

Using $h = 2.681$, Figure 10(b) shows the resulting C_t statistic of such CUSUM monitoring. No change-points are detected, but one notices the seasons with weaker performance as compared with Figure 9. Run-length computations are not immediately possible in this case as the determination of the distribution function of the LLR requires enumeration over $k^{m(m-1)} = 3^{30} = 2.06 \cdot 10^{14}$ states. As seen from (9), the LLR is a sum over the 30 possible paired-comparisons, i.e. it is the convolution of 30 independent but not identically distributed three-state variables. However, with the specific value of $\boldsymbol{R}$, where the ability of only one team changes between in-control and out-of-control, only the $2(m-1) = 10$ matches involving FC Bayern München will have a non-zero contribution to the LLR. Hence, it is only necessary to investigate $3^{10} = 59049$ states.

Since the proposed in-control and out-of-control models are time-constant, the in-control ARL can be computed by inversion of the approximate CUSUM transition matrix based on $M = 25$. Using the specified $h = 2.681$ yields an ARL of 100.05. In other words, using $h = 2.681$ means that a structural change from $\boldsymbol{\alpha}^0$ to $\boldsymbol{\alpha}^1$ is detected by pure chance on average every 100.05'th season when the data generating mechanism is $\boldsymbol{\alpha}^0$.

5 Discussion

A likelihood ratio CUSUM method for the online changepoint detection in categorical time series was presented based on categorical regression models, such as the multinomial logit model and the proportional odds model. Altogether, the presented categorical CUSUM together with the proposed run-length computation provides a comprehensive and flexible tool for monitoring categorical data streams of very different nature.

The utilized time series modeling assumed that observations were independent given the time trend and other covariates of the model. This assumption could be relaxed using for example pair-likelihood approaches (Varin & Vidoni 2006) or autoregressive models (Fahrmeir & Kaufmann 1987). It would also be of interest to embed the change-point detection within the non-Gaussian state-space modeling for ordinal time series of Fahrmeir & Tutz (1994a).

The Markov chain approximation for deriving the run length distribution of the proposed CUSUM constitutes a versatile tool for the design of categorical CUSUMs. It also constitutes a much faster alternative to this problem than simulation approaches. Embedding the approach in a numerical search procedure could be useful when performing the reverse ARL computation: Given $\boldsymbol{\pi}^0$, ARL_0, ARL_1 and a direction $\boldsymbol{R}^*, ||\boldsymbol{R}^*|| = 1$, find the corresponding magnitude $c > 0$ such that the desired run-length results are obtained for $\boldsymbol{R} = c \cdot \boldsymbol{R}^*$. Currently, the distribution function of the likelihood ratio is calculated by investigating all possible states – an approach which for large k or n_t can become intractable. Section 4.3 showed that reductions for the number of states to investigate are possible in specific applications. Still, clever approximate strategies are subject to further research – for example by identifying a subset of most probable configurations. Finally, use of the Markov chain approximation is not limited to categorical time series – also the run length of time varying count data CUSUMs can be analyzed. For example, Höhle & Mazick (2009) consider CUSUM detectors for negative binomial time series models with fixed overdispersion parameter which could be analyzed by the proposed Markov chain approach.

Other approaches exist to perform retrospective and prospective monitoring based on regression models. For example the work in Zeileis & Hornik (2007) provides a general framework for retrospective change-point detection based on fluctuation tests, which also finds prospective use. The method is, for example used in Strobl, Wickelmaier & Zeileis (2009) to retrospectively assess parameter instability in Bradley-Terry models in a psychometric context. Instead of monitoring against a specific change, another alternative is to try to detect a general change based on model residuals. For this approach, the deviance statistic is an immediate likelihood ratio based alternative suitable for monitoring within the proposed categorical CUSUM framework.

An implementation of the methods is available as functions `categoricalCUSUM` and `LR-CUSUM.runlength` in the R package `surveillance` (Höhle 2007) available from CRAN.

6 Acknowledgements

I thank Niels Peter Baadsgaard, Danish Pig Production, Denmark, for providing the abattoir monitoring data; Anette Siedler, Robert Koch Institute, Germany, and the Arbeitsgemeinschaft Masern und Varizellen, Germany, for introducing me to the varicella vaccination data. I also thank Nora Fenske, Department of Statistics, Munich, Germany, for her proofreading of the manuscript.

References

Agresti, A. (2002). *Categorical Data Analysis*, 2nd edn, Wiley.

Basseville, M. & Nikiforov, I. (1998). *Detection of Abrupt Changes: Theory and Application*. Online version of the 1994 book published by Prentice-Hall, Inc. Available from http://www.irisa.fr/sisthem/kniga/.

Bi, J. (2006). *Sensory Discrimination Tests and Measurements: Statistical Principles, Procedures and Tables*, Wiley.

Bissell, A. F. (1984). The performance of control charts and cusums under linear trend, *Applied Statistics* **33**(2): 145–151.

Bradley, R. A. & Terry, M. E. (1952). Rank analysis of incomplete block designs: I. method of paired comparisons, *Biometrika* **39**(3/4): 324–345.

Brook, D. & Evans, D. A. (1972). An approach to the probability distribution of cusum run length, *Biometrika* **59**(3): 539–549.

Brown, R., Durbin, J. & Evans, J. (1975). Techniques for testing the constancy of regression relationships over time, *Journal of the Royal Statistical Society, Series B* **37**(2): 149–192.

Chen, R. (1978). A surveillance system for congenital malformations, *Journal of the American Statistical Association* **73**: 323–327.

Courcoux, P. & Semenou, M. (1997). Preference data analysis using a paired comparison model, *Food Quality and Preference* **8**(5–6): 353–358.

Fahrmeir, L. & Kaufmann, H. (1987). Regression models for nonstationary categorical time series, *Journal of time series Analysis* **8**: 147–160.

Fahrmeir, L. & Tutz, G. (1994a). Dynamic stochastic models for time-dependt ordered paired comparison systems, *Journal of the American Statistical Association* **89**(428): 1438–1449.

Fahrmeir, L. & Tutz, G. (1994b). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 1st edn, Springer.

Fahrmeir, L. & Wagenpfeil, S. (1997). Penalized likelihood estimation and iterative kalman smoothing for non-gaussian dynamic regression models, *Computational Statistics & Data Analysis* **24**(3): 295–320.

Fokianos, K. & Kedem, B. (2003). Regression theory for categorical time series, *Statistical Science* **18**(3): 357–376.

Frisén, M. (2003). Statistical surveillance: Optimality and methods, *International Statistical Review* **71**(2): 403–434.

Glickman, M. E. (1999). Estimation in large dynamic paired comparison experiments, *Journal of the Royal Statistical Society, Series C* **48**(3): 377–394.

Grigg, O. & Farewell, V. (2004). An overview of risk-adjusted charts, *Journal of the Royal Statistical Society, Series A* **167**(3): 523–539.

Hawkins, D. M. (1992). Evaluation of average run lengths of cumulative sum charts for an arbitrary data distribution, *Communications in Statistics. Simulation and Computation* **21**(4): 1001–1020.

Hawkins, D. M. & Olwell, D. H. (1998). *Cumulative Sum Charts and Charting for Quality Improvement*, Statistics for Engineering and Physical Science, Springer.

Höhle, M. (2007). surveillance: An R package for the monitoring of infectious diseases, *Computational Statistics* **22**(4): 571–582.

Höhle, M. & Mazick, A. (2009). Aberration detection in R illustrated by Danish mortality monitoring, *in* T. Kass-Hout & X. Zhang (eds), *Biosurveillance: A Health Protection Priority*, CRC Press. To appear.

Höhle, M. & Paul, M. (2008). Count data regression charts for the monitoring of surveillance time series, *Computational Statistics & Data Analysis* **52**(9): 4357–4368.

Kaufman, H. (1987). Regression models for nonstationary categorical time series: Asmyptotic estimation theory, *Annals of Statistics* **15**: 79–98.

Kedem, B. & Fokianos, K. (2002). *Regression Models for Time Series Analysis*, Wiley.

Kim, H.-J. & Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression, *Biometrika* **76**(3): 409–423.

Knorr-Held, L. (2000). Dynamic rating of sports teams, *The Statistician* **49**(2): 261–276.

Lai, T. (1995). Sequential changepoint detection in quality control and dynamical systems, *Journal of the Royal Statistical Society, Series B* **57**: 613–658.

Lai, T. & Shan, J. (1999). Efficient recursive algorithms for detection of abrupt changes in signals and control systems, *IEEE Transactions on Automatic Control* **44**: 952–966.

Montgomery, D. C. D. (2005). *Introduction to Statistical Quality Control*, 5th edn, John Wiley.

Qiu, P. (2008). Distribution-free multivariate process control based on log-linear modeling, *IEE Transations* **40**(7): 664–677.

Reynolds, M. R. & Stoumbos, Z. G. (2000). A general approach to modeling CUSUM charts for a proportion, *IIE* **32**: 515–535.

Rigby, R. A. & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion), *Applied Statistics* **54**: 1–38.

Robert Koch Institute (2006). Epidemiologisches Bulletin 33, Available from http://www.rki.de.

Rogerson, P. & Yamada, I. (2004). Approaches to syndromic surveillance when data consist of small regional counts, *Morbidity and Mortality Weekly Report* **53**: 79–85.

Rossi, G., Lampugnani, L. & Marchi, M. (1999). An approximate CUSUM procedure for surveillance of health events, *Statistics in Medicine* **18**: 2111–2122.

Skinner, K., Montgomery, D. & Runger, G. (2003). Process monitoring for multiple count data using generalized linear model-based control charts, *International Journal of Production Research* **41**(6): 1167–180.

Steiner, S. H., Cook, R. J. & Farewell, V. T. (1999). Monitoring paired binary surgical outcomes using cumulative sum charts, *Statistics in Medicine* **18**: 69–86.

Steiner, S. H., Cook, R. J., Farewell, V. T. & Treasure, T. (2000). Monitoring surgical performance using risk-adjusted cumulative sum charts, *Biostatistics* **1**(4): 441–452.

Strobl, C., Wickelmaier, F. & Zeileis, A. (2009). Accounting for individual differences in Bradley-Terry models by means of recursive partitioning, *Technical Report 54*, Department of Statistics, University of Munich. Available as `http://epub.ub.uni-muenchen.de/10588/`.

Topalidou, E. & Psarakis, S. (2009). Review of multinomial and multiattribute quality control charts, *Quality and Reliability Engineering International* . In press.

Tutz, G. (1986). Bradley-Terry-Luce models with an ordered response, *Journal of Mathematical Psychology* **30**: 306–316.

Varin, C. & Vidoni, P. (2006). Pairwise likelihood inference for ordinal categorical time series, *Computational Statistics & Data Analysis* **51**: 2365–2373.

Venables, W. N. & Ripley, B. D. (2002). *Modern Applied Statistics with S*, 4th edn, Springer.

Whiting, M. J., Stuart-Fox, D. M., O'Connor, D., Firth, D., Bennett, N. & Blomberg, S. P. (2006). Ultraviolet signals ultra-aggression in a lizard, *Animal Behaviour* **72**(353–363).

Wolfe, D. A. & Chen, Y.-S. (1990). The changepoint problem in a multinomial sequence, *Communications in Statistics – Simulation and Computation* **19**(2): 603–618.

Yee, T. W. (2008). The VGAM package, *R News* **8**(2): 28–39.

Yee, T. W. & Wild, C. J. (1996). Vector generalized additive models, *Journal of the Royal Statistical Society, Series B, Methodological* **58**: 481–493.

Zeileis, A. & Hornik, K. (2007). Generalized M-fluctuation tests for parameter instability, *Statistica Neerlandica* **61**(4): 488–508.