Estimating the change point of binary profiles in phase II

Alireza Sharafi

Industrial Engineering Department,
Amirkabir University of Technology,
P.O. Box 15875-4413, Tehran, Iran
E-mail: a.sharafi@aut.ac.ir

Majid Aminnayeri

Industrial Engineering Department,
Amirkabir University of Technology,
Tehran, P.O. Box 15875-4413, Iran
E-mail: mjnayeri@aut.ac.ir

Amirhossein Amiri*

Industrial Engineering Department,
Shahed University,
Tehran, P.O. Box 18151/159, Iran
Fax: (+9821)51212021
E-mail: amiri@shahed.ac.ir

Abstract: Identifying the real time change in a process, when an out-of-control signal is given by a control chart is crucial and leads to cost and time savings significantly. There are processes or products whose quality is described by a relationship between a response variable and one or more explanatory variable. This relationship is known as profile in the literature of statistical process control and is modelled by different regression types including logistic regression. This type of profile is used where the response variable follows a binomial distribution. In this paper, a maximum likelihood estimator (MLE) based on regression parameters is developed to find the real time of a step change in logistic regression profiles. Simulation studies are provided to evaluate the performance of the proposed change point estimator under both single step change and linear trend change.

Keywords: binary profiles; change point; maximum likelihood estimator; MLE; exponentially weighted moving average; EWMA; control chart; statistical process control; SPC.

Reference to this paper should be made as follows: Sharafi, A., Aminnayeri, M. and Amiri, A. (2014) 'Estimating the change point of binary profiles in phase II', *Int. J. Productivity and Quality Management*, Vol. 14, No. 3, pp.336–351.

Biographical notes: Alireza Sharafi holds an MS in Industrial Engineering from Amirkabir University of Technology in 2011. He received his BSc in Industrial Engineering from Qom University in Iran in 2009. His current research interests include statistical methods for quality improvement, profile monitoring and design of experiments.

Majid Aminnayeri received his BSc in Mathematics from Shiraz University in 1972 and his MSc in Statistics from the same university in 1974. He received his PhD double major in Industrial Engineering and Statistic from Iowa State University in 1981. He served as an academic member in Iowa State University and Kerman University. He is currently an Associate Professor in Amirkabir University of Technology. The scope of his work is mostly in statistical quality control and scheduling as well as general modelling.

Amirhossein Amiri is an Assistant Professor at Shahed University. He holds a BS, MS, and PhD in Industrial Engineering from Khajeh Nasir University of Technology, Iran University of Science and Technology, and Tarbiat Modares University, respectively. He is a member of the Iranian Statistical Association. His research interests are statistical quality control, profile monitoring, and Six Sigma.

1 Introduction

The world is witnessing a extraordinary change in the business environment and technological advances. Successful organisations are challenged to stay ahead of its competitors by producing high quality products at low cost. This can only be achieved through continuous improvement of organisation's processes. To answers to these challenges, industries have adopted various quality improvement methods. Many companies considered total quality management (TQM) as the best way to improve profit, market share and competitiveness (Singh and Shrivastava, 2012; Talib and Rahman, 2012). Refer to Sousa and Voss (2002) and Kull and Wacker (2010) for reviews of quality management (QM) methods and cultural aspects.

Statistical process control (SPC) including its powerful tools is a method used in QM systems to monitor the process and detect assignable causes. Akram et al. (2012) covered preliminaries on SPC and automatic process control (APC). SPC mainly involves plotting and interpretation of statistical control charts. Control charts are the most commonly used statistical process monitoring tools which signal when the process is being affected by a special cause. Jarrett and Pan (2009) suggest multivariate methods for the construction of quality control charts for the control and improvement of output of manufacturing processes.

Statistical process control (SPC) is used to monitor and reduce the variation of a process. Control charts are the most powerful tools in SPC which are used to monitor quality characteristics. These charts can detect any changes or shifts in a process; however, a shift usually occurs much earlier before it is detected. When a control chart signals a special cause, quality engineers should identify and remove the source(s) of variation and return the process to in-control state. However, knowing when a process has changed would help quality engineers to limit the time window within which they should search for assignable causes. Consequently, the assignable causes can be

identified sooner and corrective action can be implemented more quickly. Identifying the real time of the process change is known as change point estimation problem.

Change point problems are mainly classified according to change types including step, drift and monotonic shifts. To find the real time of a change, many authors have suggested several methods such as maximum likelihood estimator (MLE), cumulative sum (CUSUM), exponentially weighted moving average (EWMA) and intelligent methods (artificial network, clustering and decision tree). For example, Pignatiello and Samuel (1998, 2001) and Perry and Pignatiello (2005, 2006) proposed a MLE in different control charts to find the real time of a change under step and drift shifts, respectively. Pignatiello and Samuel (2001) proposed MLE to determine the change point of a normal process when a signal is issued by an EWMA or a CUSUM control chart. Shao and Hou (2011) proposed the combination of an EWMA control chart and MLE to detect any changes and to estimate the change point of a gamma process. They concluded that the performance of the MLE is better than the built-in estimators over the range of shift considered. For a comprehensive review on change point estimation methods, refer to Amiri and Allahyari (2012).

In the area of SPC, usually the quality of a process or a product was represented by the distribution of one (or more) quality characteristic and monitored by univariate (or multivariate) control chart. Two recent researches on multivariate control charts can be found in Doroudyan and Amiri (2013) and Amiri et al. (2013). However, Kang and Albin (2000) presented a popular topic with widespread application in SPC namely, profiles monitoring, in which a relationship between a response variable and one or more independent variables, known as profiles is monitored over time. A collection of data points of these variables can be observed at each sampling stage, which can be represented by a curve (i.e., a profile). The objective in profile monitoring is to detect any changes in the desired relationship between response and explanatory variables. Scientists classify profiles based on the type of relationships into the linear (simple, multiple and polynomial), non-linear and geometric profiles. Profiles are studied in phases I and II. Mestek et al. (1994), Mahmoud and Woodall (2004) and Mahmoud et al. (2007) have studied phase I monitoring of Simple linear profiles. Kang and Albin (2000), Kim et al. (2003), Zou et al. (2006), Saghaei et al. (2009) and Zhang et al. (2009) have investigated phase II monitoring of simple linear profiles.

Identifying the change point in profiles has been studied by some researchers. Mahmoud et al. (2007) used a likelihood ratio-based method to identify the real time of a step change in phase I monitoring of a simple linear profiles. Zou et al. (2006) proposed a method based on likelihood ratio statistics to find the change point in parameters of a simple linear profile in phase II. Kazemzadeh et al. (2008) used the same method to estimate the change point in polynomial profiles under a step shift in phase I. Note that the change point problem in profile data is under a different sampling framework from that of the other models. In the profile applications, multiple datasets are collected over time in a functional data sampling framework and a possible change is occurred after any profile sample.

Most researches in profile monitoring, assume the response variable is continuous (usually normal) and characterise profiles with linear or non-linear models. However, in many industrial applications the response variable is discrete such as binary (in which a product is classified into conforming and non-conforming) or countable (for example, the number of defects in a product). Yeh et al. (2009) studied binary profiles in phase I. They proposed different T^2 control charts for monitoring logistic regression profiles.

Shang et al. (2011) proposed a control scheme based on EWMA-GLM to monitor the relationship between the binary response and random explanatory variables in Phase II. Their approach assumes that explanatory variables are random variable and takes different values in each profile sample. To the best of our knowledge, only Sharafi et al. (2012) suggested a method to identify the real time of a step change in Phase II monitoring of binary profiles. They used a T^2 control chart to monitor the binary profiles and assumed that shift in the mean of the response variable is occurred in different levels of the explanatory variable not in the logistic regression parameters. Sharafi et al. (2013) proposed an MLE estimator to identify the real time of a step change in phase II monitoring of Poisson regression profiles, based on Poisson regression parameters. In this paper we propose a new method to find the real time of a change in Phase II monitoring of logistic regression profiles. In the proposed method, we first use an EWMA-based control chart to detect any changes in the parameters of the logistic regression profile. Then, we develop an MLE change point estimator based on the parameters of logistic regression profiles. The rest of the paper is organised as follows: Section 2 illustrates logistic regression model and explains the steps of estimating the model parameters. Section 3 presents the change point model and assumptions of the problem. The performance of the proposed model is investigated in Section 4. Conclusions and some future researches are provided in the final section.

2 Logistic regression model

Many categorical response variables have only two categories: for example, whether one takes public transportation today (yes, no), or whether one has had a physical exam in the past year (yes, no). Denote a binary response variable by y and $E(y) = \pi$. The value of π can vary as the value of x changes, so we replace π by $\pi(x)$. The relationships between $\pi(x)$ and x are usually non-linear and the S-shaped curves are often realistic shapes for this relationship which are called the logistic regression models. Therefore, the logistic model is a mathematical modelling approach that can be used to describe the relationship of several X's to a dichotomous dependent variable, such as π. The fact that the logistic function $\pi(x)$ ranges between 0 and 1 is the primary reason the logistic model is so popular.

In a logistic regression model, there are n independent experimental sets, with p predictor variables for each set, which is shown by $\mathbf{x}_i = (x_{i1}, x_{i2},\ldots,x_{ip})^T$ as well as corresponding Bernoulli response variables namely z_i for $i = 1, 2,\ldots,n$. For n independent observations, the number of successes has the binomial distribution with parameter $\boldsymbol{\pi}$ ($\boldsymbol{\pi} = (\pi_1, \pi_2,\ldots,\pi_n)^T$). Thus, the logistic model may be written as $p(x)$, where the bold X is a shortcut notation for the collection of variables $\mathbf{x}_1$ through $\mathbf{x}_n$. The probability of success in each set is denoted by π_i and each π_i is a function of $\mathbf{x}_i$. In the logistic regression model this function is characterised by the link function $g(\pi_i)$, defined as

$$g(\pi_i) = \frac{\log(\pi_i)}{1-\log(\pi_i)} = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2,\ldots,\beta_p)^T$ is the vector of logistic regression parameters. It is usual to set $x_{i1} \equiv 1$ such that β_1 to be the intercept of the model. Hence, equation (1) is simplified as

$$\pi_i = \frac{\exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)}{1+\exp\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)} = \frac{\exp\left(\eta_i\right)}{1+\exp\left(\eta_i\right)}. \tag{2}$$

In this equation $\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} = \sum_{k=1}^{p}\beta_k x_{ik}$. We also assume that data are grouped so that for the ith setting of the predictor variables, there are m_i observations, $i = 1, 2,\ldots,n$. $M = \sum_{i=1}^{n} m_i$ is the total number of observations. The response variable is $y_i = \sum_{k=1}^{m_i} z_{ik}$, where z_{ik} is the kth observation in i^{th} predictor variable settings, y_i follows a binomial distribution with parameters m_i, and π_i.

To estimate the parameters of this profile, popular methods such as least square error (LSE) cannot be applied. Because response variable follows a Bernoulli distribution, and this method leads to unbiased estimators which do not have minimum variance. Albert and Anderson (1984) proposed an MLE method to estimate the model parameters. They used the following likelihood function:

$$L(\boldsymbol{\pi}, \mathbf{y}) = \prod_{i=1}^{n}\binom{m_i}{y_i}[\pi_i]^{y_i}\,[1-\pi_i]^{m_i-y_i} \tag{3}$$

where $\boldsymbol{\pi} = (\pi_1, \pi_2,\ldots,\pi_n)^T$ and $\mathbf{y} = (y_1, y_2,\ldots,y_n)^T$. Taking the logarithm of equation (3) and using $\eta_i = \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} = \sum_{k=1}^{p}\beta_k x_{ik} = \log\frac{\pi_i}{1-\pi_i}$, one can reexpress the log-likelihood as

$$l(\boldsymbol{\beta}, \mathbf{y}) = \sum_{i=1}^{n}\log\binom{m_i}{y_i} + \sum_{i=1}^{n}\sum_{k=1}^{p} y_i\beta_x x_{ik} - \sum_{i=1}^{n} m_i \log\left[1+\exp\left(\sum_{k=1}^{p}\beta_k x_{ik}\right)\right] \tag{4}$$

an iterative algorithm is used to estimate the parameters in logistic regression profiles by using the MLE method. Taking derivative of equation (4) with respect to $\boldsymbol{\beta}$, and using iterative weighted least square estimation method suggested by McCullagh and Nelder (1989), the logistic regression parameters can be estimated as follows:

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}\right)^{-1}\mathbf{X}^T\hat{\mathbf{W}}\mathbf{q} \tag{5}$$

In equation (5)

$$\mathbf{X} = \left(\mathbf{x}_1, \mathbf{x}_2,\ldots,\mathbf{x}_n\right)^T,$$

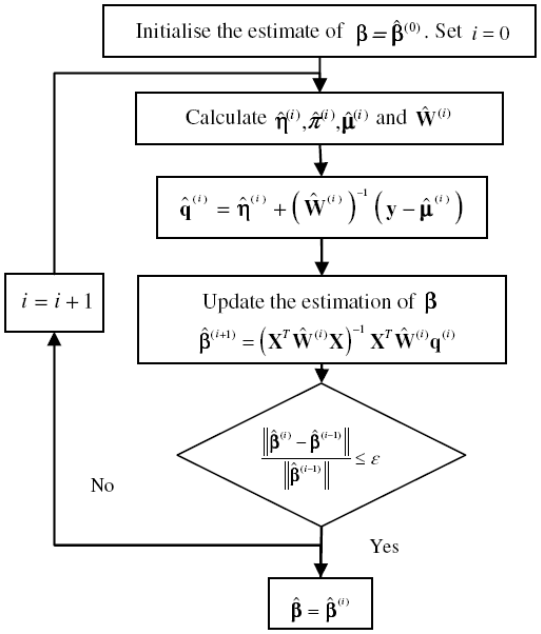
$$\hat{\mathbf{W}} = diag[m_1\hat{\pi}_1(1-\hat{\pi}_1), m_2\hat{\pi}_2(1-\hat{\pi}_2),\ldots,m_n\hat{\pi}_n(1-\hat{\pi}_n)]$$

and

$$\hat{\boldsymbol{\mu}} = \left(m_1\hat{\pi}_1, m_2\hat{\pi}_2,\ldots,m_n\hat{\pi}_n\right)^T.$$

The procedure iterations are described in Figure 1.

Figure 1 The procedure of estimation of logistic regression parameters



McCullagh and Nelder (1989) proved that as either n becomes large or under fixed n as m_i becomes large, $\hat{\boldsymbol{\beta}}$ is distributed asymptotically as a p-dimensional normal distribution $N_p(\boldsymbol{\beta}, (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1})$. This procedure will be used in the maximum likelihood change point estimator described in Section 3.

3 Methodology

3.1 Control scheme for monitoring the binary profile

With respect to the monitoring of categorical data, conventional charts such as p and np control charts have been developed. However, regarding the profile monitoring schemes for categorical data, few studies have been conducted recently. Yeh et al. (2009) modelled the relationship between the binary response and explanatory variables by using the logistic regression model. They introduced five Hotelling T^2 control charts to monitor binary profiles in phase I. The plotting statistic for profile j is defined as

$$T_j^2 = \left(\hat{\boldsymbol{\beta}}_j - \bar{\hat{\boldsymbol{\beta}}}\right)^T \boldsymbol{S}^{-1}\left(\hat{\boldsymbol{\beta}}_j - \bar{\hat{\boldsymbol{\beta}}}\right) \tag{6}$$

where $\hat{\boldsymbol{\beta}}_j$ is estimator of the logistic regression parameters in j^{th} profile, $\mathbf{S} = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$ is the variance covariance matrix of $\hat{\boldsymbol{\beta}}_j$ and $\mathbf{X}^T$ and $\mathbf{W}$ are the corresponding matrices defined at the previous section for the jth profile sample. Any of these T^2 charts presents a different way to estimate $\bar{\hat{\boldsymbol{\beta}}}$ and $\mathbf{S}$. They showed that T_I^2 control chart in which the covariance matrix is estimated by averaging the covariance estimates of each sample, is

the best method in detecting both step shifts and drift in the regression parameters. The T_I^2 control chart is referred to as Shewhart-GLM chart in Shang et al. (2011).

Shang et al. (2011) proposed a control scheme based on EWMA-GLM to represent the relationship between the binary response and random explanatory variables in phase II. For the jth random profile sample collected over time, their charting statistics is defined as follows:

$$lr_j = \left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0\right)^T \sum\nolimits_{\hat{\boldsymbol{\beta}}_j}^{-1} \left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0\right) + \frac{n(2-\lambda)}{\lambda}\left(\mathbf{e}_j - \boldsymbol{\mu}_0\right)^T \sum\nolimits^{-1} \left(\mathbf{e}_j - \boldsymbol{\mu}_0\right) \tag{7}$$

where λ is a weighting parameter, Σ is the variance-covariance matrix of $\mathbf{e}_j$ and

$$\sum\nolimits_{\boldsymbol{\beta}_j}^{-1} \frac{\lambda}{(2-\lambda)}\left(\hat{\mathbf{X}}_j^T \hat{\mathbf{W}}_j \hat{\mathbf{X}}_j\right)^{-1}, \quad \hat{\mathbf{X}}_j = \left(\mathbf{X}_1^T, \ldots, \mathbf{X}_j^T\right), \quad \hat{\mathbf{W}}_j = diag\left\{\hat{\mathbf{V}}_1 \ldots \hat{\mathbf{V}}_j\right\}$$
$$\hat{\mathbf{V}}_k = diag\{\hat{w}_{k1}, \ldots, \hat{w}_{kn}\}, \quad \hat{w}_{ki} = \lambda(1-\lambda)^{j-k} m_i \pi_{ki}\left(1-\pi_{ki}\right)$$
$$\mathbf{e}_j = \lambda \bar{\mathbf{x}}_j + (1-\lambda)\mathbf{e}_{j-1}, \quad j = 1, 2, \ldots$$

where $\mathbf{e}_0 = \boldsymbol{\mu}_0$ is the starting vector and $\bar{\mathbf{x}}_j = \sum_{i=1}^{n} \mathbf{x}_{ji}/n$. In addition, when the process is in-control, they assumed that $\mathbf{x}_{ji}$ follows $N_p(\boldsymbol{\mu}_0, \Sigma)$. In equation (7), $\hat{\boldsymbol{\beta}}_j$ can be obtained by implementing the procedure in Section 2, replacing $\mathbf{W}$ with $\hat{\mathbf{W}}_j$. Their approach assumes that explanatory variables are random and take different values from sample to sample. They showed that the EWMA-GLM scheme performs always roughly better than the Shewhart-GLM scheme in detecting the shifts in the regression parameters. Hence, we used this control chart to detect the out-of-control state in phase II. Here we assume that the explanatory variables are fixed from profile to profile and equal to $\boldsymbol{\mu}_0$ under both in-control and out-of-control situations; hence the second term of equation (7) is equal to zero and the charting statistic for profile j is defined as

$$lr_j = \left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0\right) \sum\nolimits_{\hat{\boldsymbol{\beta}}_j}^{-1} \left(\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_0\right) \tag{8}$$

The chart signals when $lr_j > L_M$, where L_M is the upper control limit determined such that a specific in-control average run length (ARL_0) is obtained.

3.2 MLE change point estimator

Here, it is assumed that the underlying process initially operates in a state of statistical control, with observations coming from a binomial distribution with the parameters m_i and π_i. So, the mass probability function of the observations is $f(y_{ij}) = \binom{m_i}{y_{ij}} \pi_i^{y_{ij}} (1-\pi_i)^{m_i - y_{ij}}$. By replacing π_i with logistic link function given in equation (2), the mass probability function of the observations is simplified as $f(y_{ij}) = \binom{m_i}{y_{ij}} (\exp(\mathbf{x}_i\boldsymbol{\beta}))^{y_{ij}} \left(\frac{1}{1+\exp(\mathbf{x}_i\boldsymbol{\beta})}\right)^{m_i}$, where y_{ij} is the response variable for the ith predictor variable in the jth profile. It is also assumed that the vector of logistic regression

parameters ($\boldsymbol{\beta}_0$) are known and are estimated from a historical dataset in phase I analysis. After elapsing an unknown amount of time, the vector of $\boldsymbol{\beta}$ changes from its in control state of $\boldsymbol{\beta}_0$ to an unknown out-of-control state of $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \boldsymbol{\Delta}$, and it remains at the new level until the source of the assignable cause is identified and eliminated, where $\boldsymbol{\Delta} = (\delta_1\sigma_1, \delta_2\sigma_2)^T$ and δ_1, δ_2 are constant. In other words, for profiles $j = 1, 2,\ldots,\tau$, the vector of regression parameters $\boldsymbol{\beta}$ is equal to its known in-control value $\boldsymbol{\beta}_0$. For profiles $j = \tau + 1, \tau + 2,\ldots,T$, the vector of regression parameters $\boldsymbol{\beta}$ becomes equal to some unknown vector of regression parameters $\boldsymbol{\beta}_1$, where T is the last profile sampled in which the control chart signalled an out-of-control state. Here we describe the level of shifts in $\boldsymbol{\beta}$ based on the non-centrality parameter (*ncp*) which is defined as $ncp = \boldsymbol{\Delta}^T\mathbf{S}^{-1}\boldsymbol{\Delta}$. τ is the unknown parameter representing the last profile taken from an in-control process. To estimate this unknown parameter, change point, an MLE approach is used. We denote the MLE estimator of the change point as $\hat{\tau}$. The likelihood function is as follows:

$$\begin{aligned} L(\tau, \boldsymbol{\beta}_1 \mid \mathbf{y}) = & \prod_{j=1}^{\tau}\prod_{j=1}^{n}\binom{m_i}{y_{ij}} \times \prod_{j=1}^{\tau}\prod_{i=1}^{n}\left[\exp(\mathbf{x}_i\boldsymbol{\beta}_0)\right]^{y_{ij}}\left[\frac{1}{1+\exp(\mathbf{x}_i\boldsymbol{\beta}_0)}\right]^{m_i} \\ & \times \prod_{j=\pi+1}^{L}\prod_{i=1}^{n}\left[\exp(\mathbf{x}_i\boldsymbol{\beta}_1)\right]^{y_{ij}}\left[\frac{1}{1+\exp(\mathbf{x}_i\boldsymbol{\beta}_1)}\right]^{m_i} \end{aligned} \tag{9}$$

The MLE of τ is the value of τ that maximises the likelihood function in equation (9). Taking the logarithm of equation (9), we have

$$\begin{aligned} Ln\, L(\tau, \boldsymbol{\beta}_1 \mid \mathbf{y}) = & \sum_{j=1}^{\tau}\sum_{i=1}^{n}\binom{m_i}{y_{ij}} + \sum_{j=1}^{\tau}\sum_{i=1}^{n} y_{ij}\mathbf{x}_i\boldsymbol{\beta}_0 - m_i\sum_{j=1}^{\tau}\sum_{i=1}^{n} Ln\left[\exp(1+\mathbf{x}_i\boldsymbol{\beta}_0)\right] \\ & \times \sum_{j=\pi+1}^{L}\sum_{i=1}^{n} y_{ij}\mathbf{x}_i\boldsymbol{\beta}_1 - m_i\sum_{j=\tau+1}^{L}\sum_{i=1}^{n} Ln\left[1+\exp(\mathbf{x}_i\boldsymbol{\beta}_1)\right] \end{aligned} \tag{10}$$

To determine the unknown parameter τ in equation (10), first we should estimate the unknown vector of $\boldsymbol{\beta}_1$ which maximise the log-likelihood function in equation (10), defined as $\hat{\boldsymbol{\beta}}_1$. The partial derivative of equation (10) with respect to $\boldsymbol{\beta}_1$ is given by

$$\frac{\partial Ln\, L(\tau, \boldsymbol{\beta}_1 \mid \mathbf{y})}{\partial \boldsymbol{\beta}_1} = \sum_{j=\tau+1}^{L}\sum_{i=1}^{n} y_{ij}\mathbf{x}_i - m_i\sum_{j=\tau+1}^{L}\sum_{i=1}^{n}\frac{\mathbf{x}_i\exp(\mathbf{x}_i\boldsymbol{\beta}_1)}{1+\exp(\mathbf{x}_i\boldsymbol{\beta}_1)} \tag{11}$$

Since the vector $\boldsymbol{\beta}$ appears with the $\mathbf{X}$ in the terms of equation (10), an expression of $\mathbf{x}_i\boldsymbol{\beta}_1$ can also help to estimate the change point. For a fixed value for τ, the MLE for $\mathbf{x}_i\boldsymbol{\beta}_1$ is

$$\mathbf{x}_i\hat{\boldsymbol{\beta}}_1 = Ln\left[\frac{\sum_{j=\tau+1}^{T}\mathbf{y}_j}{n(T-\tau) - \sum_{j=\tau+1}^{T}\mathbf{y}_j}\right] \tag{12}$$

By obtaining $\mathbf{x}_i\hat{\boldsymbol{\beta}}_1$, replacing it in equation (10) and calculating the logarithm of the likelihood function for all possible change-point values, the MLE of the change point τ is

the value which maximises the expression in equation (10). The estimate of the change point is as follows:

$$\hat{\tau} = \arg\max \left\{ \sum_{j=1}^{\tau} \sum_{i=1}^{n} y_{ij} \mathbf{x}_i \boldsymbol{\beta}_0 - m_i \sum_{j=1}^{\tau} \sum_{i=1}^{n} Ln\left[1 + \exp\left(\mathbf{x}_i \boldsymbol{\beta}_0\right)\right] \right.$$
$$\left. + \sum_{j=\tau+1}^{L} \sum_{i=1}^{n} y_{ij} \mathbf{x}_i \hat{\boldsymbol{\beta}}_1 - m_i \sum_{j=\tau+1}^{L} \sum_{i=1}^{n} Ln\left[1 + \exp\left(\mathbf{x}_i \hat{\boldsymbol{\beta}}_1\right)\right] \right\}. \tag{13}$$

Whenever the EWMA-GLM chart signals an out-of-control state, the real time of a change can be estimated via equation (13). Sharafi et al. (2012) suggested a similar method to identify the real time of a step change in phase II monitoring of binary profiles. They used a Shewhart-GLM control chart to monitor these profiles and assumed that shifts are occurred in different levels of explanatory variable (π_i) not in the logistic regression parameters ($\boldsymbol{\beta}$). As an alternative, we consider the EWMA-GLM chart and set another type of change that is more valid.

4 Performance of the MLE estimator

In this section, the performance of the proposed estimator is examined using Monte Carlo simulation. For this simulation study, we assume that the number of predictor variables in binary profile is one (p = 2). Thus, the link function is simplified as $g(\pi_i) = \beta_1 + \beta_2 x_i$ where β_1 and β_2 are the intercept and the slope of the regression function respectively and are shown by the vector $\boldsymbol{\beta} = (\beta_1, \beta_2)^T$. Also we set the design matrix $\mathbf{X}$ as:

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \log(0.1) & \log(0.2) & \dots & \log(0.9) \end{pmatrix}^T,$$

It is assumed that the number of experiments in each predictor variable is constant and equal to 50, (m_i = 50 for i = 1, 2,…,9) and the in-control $\boldsymbol{\beta}$ is $\boldsymbol{\beta}_0 = (1, 2)^T$ which comes from historical dataset in phase I. The upper control limit of the EWMA-GLM control schemes is obtained by simulation to roughly achieve a specified in-control ARL of 200. We study two distinct change types from all of the non-decreasing change types. These change types are

1 a single step change

2 a linear trend

and the results are described in details in the following subsections:

4.1 A single step change

In this section we compare the performance of the proposed estimator when a single step change is present. This type of change occurs when the process mean suddenly changes due to an assignable cause and remains fixed in the process until a corrective action is taken.

Suppose an out-of-control process whose parameter vector $\boldsymbol{\beta}$ shifts from $\boldsymbol{\beta}_0$ to $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta} = (\delta_1\sigma_1, \delta_2\sigma_2)^T$ and δ_1, δ_2 are constant. A Monte Carlo simulation study is accomplished to examine the performance of the estimator. The change point is considered at profile 50 (τ = 50). For profiles j = 1, 2,…,50, the observations are generated by binomial distribution with parameters 50 and π_i computed by the vector of the regression parameters ($\boldsymbol{\beta}_0$). Starting at profile 51, observations are simulated from the out-of-control process with the vector of regression parameters $\boldsymbol{\beta}_1$ until the EWMAGLM chart signals an out-of-control. At this time, the change point estimator in equation (13) is used and the real time of the process change is determined. This procedure, which is described in the following flowchart, is repeated 10,000 times for each magnitude of the change under study (Figure 2).

Figure 2 The flowchart of the simulation procedure

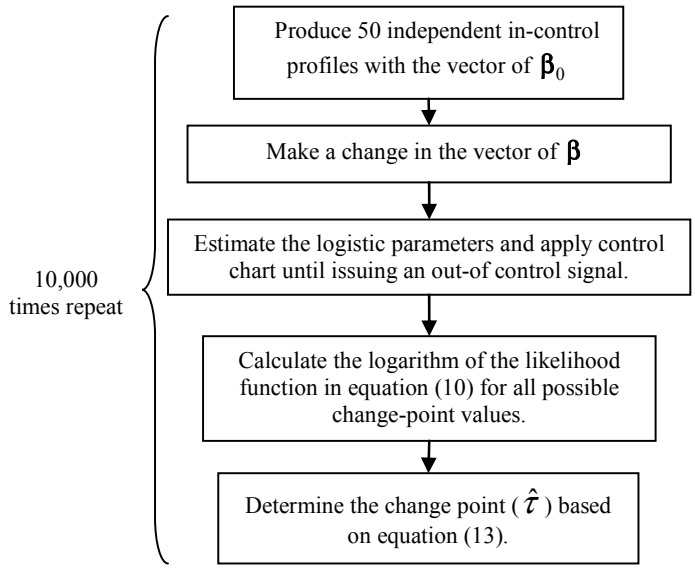


Table 1 Expected number of samples taken until the signal is issued and the mean and standard deviation of the change point estimator under different step shifts with 10,000 simulations runs when $\tau = 50$

ncp	(δ_1, δ_2)	$E(T)$	$\bar{\hat{\tau}}$	$se(\hat{\tau})$
0.30	(0.1, 0.4)	127.20	57.09	7.80
0.45	(0.3, 0.6)	117.98	53.31	7.34
0.59	(0.6, 0.2)	93.79	53.41	6.67
0.72	(0.5, 0)	84.82	52.16	6.08
1.27	(0.5, 1)	79.28	50.25	4.32
2.88	(1, 0)	57.42	50.06	3.39
4.42	(2, 2)	56.29	49.70	3.93
5.10	(1, 2)	56.52	49.58	3.76
6.52	(0, 1.5)	53.76	49.43	2.95
11.54	(2, 0)	51.41	49.14	1.63
14.86	(3, 1)	51.26	48.96	1.42

Table 2 Estimated precision performance of the change point estimator over a range of *ncp* with 10,000 simulations runs when $\tau = 50$

ncp	(δ_1, δ_2)	$\hat{p}(\lvert\hat{\tau}-\tau\rvert = 0)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert \le 1)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert \le 2)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert \le 3)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert \le 4)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert \le 5)$
0.30	(0.1, 0.4)	0.15	0.25	0.33	0.43	0.50	0.56
0.45	(0.3, 0.6)	0.17	0.28	0.37	0.47	0.55	0.61
0.59	(0.6, 0.2)	0.20	0.31	0.42	0.51	0.59	0.68
0.72	(0.5, 0)	0.23	0.38	0.49	0.58	0.66	0.72
1.27	(0.5, 1)	0.29	0.49	0.61	0.70	0.78	0.84
2.88	(1, 0)	0.41	0.63	0.75	0.83	0.88	0.92
4.42	(2, 2)	0.44	0.68	0.76	0.80	0.84	0.89
5.10	(1, 2)	0.48	0.68	0.76	0.81	0.85	0.89
6.52	(0, 1.5)	0.52	0.72	0.83	0.88	0.92	0.94
11.54	(2, 0)	0.66	0.79	0.87	0.92	0.96	0.99
14.86	(3, 1)	0.68	0.83	0.91	0.95	0.97	1.00

The results are summarised in Tables 1 and 2. Table 1 shows the expected length of each simulation run $E(T)$ which is the expected value of the number of samples taken until the first alarm is given by the control chart, i.e., $E(T) = ARL + 50$. Table 1 also shows the average change point estimate and the standard deviation of the change point estimator under different magnitude of the shifts considered. Because the actual change is at time 50, the average change point estimate, $\bar{\hat{\tau}}$, should be close to 50.

To efficiently demonstrate the experimental results, we use the case of *ncp* =1.27 in the Table 1 as an illustration. From this table, we can notice that the expected number of samples taken until the signal is 79.28. It means that the EWMA-GLM control chart on average will signal the change in the process mean on subgroup 79.28, when the actual change has occurred after the 50th subgroup. For this case, the average of change point estimate is 50.25 which is quite close to the actual change point of $\tau = 50$. Moreover, the standard deviation of the change point estimator is 4.32. Hence, our proposed change point estimator works satisfactory for approximately all magnitudes of the shifts. Furthermore, as the magnitude of the step change increases, the performance of the estimator improves significantly. In Table 1, for a number of cases especially when the shift size is large the estimate of change point is below the real one; for example last five consecutive rows of the table. In these cases, there are insufficient numbers of samples beyond the true change point ($\tau = 50$) to support model estimation. As a result, the proposed $\hat{\tau}$ estimator underestimates the real time of the change τ.

Table 2 shows the results of proportion of 10,000 simulation runs that the estimator lies within a specified tolerance of the real change point value. For example if *ncp* =1.27 , the estimated probability that $\hat{\tau}$ lies within the interval of 1 or less from the real change point is 0.49. Also in this case, in 29% of the simulation runs the estimator correctly identifies the real time of the change.

Table 3 Accuracy performance for the change point estimator when a linear trend disturbance is present

ncp	(δ_1, δ_2)	$E(T)$	$\bar{\hat{\tau}}$	$se(\hat{\tau})$
0.005	(0.03, 0.06)	75.02	58.91	8.76
0.011	(0.1, 0.1)	67.39	55.79	8.14
0.028	(0.1, 0)	61.67	53.74	5.44
0.051	(0.2, 0.1)	59.51	52.66	4.48
0.115	(0, 0.2)	57.73	51.48	2.99
0.148	(0.3, 0.1)	56.29	51.17	2.55
0.203	(0.2, 0.4)	56.64	50.80	2.09
0.276	(0.5, 0.5)	55.45	50.59	1.89
0.461	(0.4, 0)	54.09	50.44	1.77
0.721	(0, 0.5)	53.89	50.23	1.54
0.876	(0.7, 0.2)	53.41	50.08	1.44
1.41	(0.7, 0)	52.93	49.83	8.76
1.74	(0.3, 1)	52.89	49.12	1.00

Table 4 Estimated precision performances over a range of *ncp* values when a linear trend disturbance is present

ncp	(δ_1, δ_2)	$\hat{p}(\lvert\hat{\tau}-\tau\rvert=0)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert\le 1)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert\le 2)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert\le 3)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert\le 4)$	$\hat{p}(\lvert\hat{\tau}-\tau\rvert\le 5)$
0.30	(0.1, 0.4)	0.15	0.25	0.33	0.43	0.50	0.56
0.45	(0.3, 0.6)	0.17	0.28	0.37	0.47	0.55	0.61
0.59	(0.6, 0.2)	0.20	0.31	0.42	0.51	0.59	0.68
0.72	(0.5, 0)	0.23	0.38	0.49	0.58	0.66	0.72
1.27	(0.5, 1)	0.29	0.49	0.61	0.70	0.78	0.84
2.88	(1, 0)	0.41	0.63	0.75	0.83	0.88	0.92
4.42	(2, 2)	0.44	0.68	0.76	0.80	0.84	0.89
5.10	(1, 2)	0.48	0.68	0.76	0.81	0.85	0.89
6.52	(0, 1.5)	0.52	0.72	0.83	0.88	0.92	0.94
11.54	(2, 0)	0.66	0.79	0.87	0.92	0.96	0.99
14.86	(3, 1)	0.68	0.83	0.91	0.95	0.97	1.00

4.2 A linear trend change disturbance

In this section, we study the linear trend change case where we assume a constant drift in the process mean. This change type occurs when process parameters change with a constant rate b. For example, tool wear or heat effect can be the likely cause of this change. Here, we use Monte Carlo simulation to study the accuracy and precision of the proposed step change estimator when a linear trend change occurs in the vector of regression parameters $\boldsymbol{\beta}$.

In the simulation study, the process change point is considered at $\tau = 50$. We used again the EWMAGLM control chart to detect the out-of-control signal. During the simulation of profiles $j = 1, 2,\ldots,50$, the vector of the regression parameters is equal to its known in-control vector of $\boldsymbol{\beta}_0$. Therefore, for these profiles, observations are randomly generated from a binomial distribution with parameters 50 and π_i. After profile 50, observations are simulated from the out-of-control process with $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_0 + b(j - 50)$ until the EWMA-GLM control chart signals an out-of-control. At this time, the change point estimator in equation (13) is used and the real time of the process change is determined. This procedure is repeated 10,000 times for different b values.

The simulation results are presented in Tables 3 and 4. Table 3 shows the accuracy performance of the change point estimator and Table 4 shows the results of proportion of 10,000 simulation runs that the estimator lies within a specified tolerance of the real change point value. Adequate results are obtained in this example which shows the acceptable performance of the proposed change point estimator under the linear drift.

Generally, our simulation studies showed that the proposed step change estimator not only performs well in estimating the real time of a step change, but also the real time of a linear trend disturbance. Hence, this estimator can be used for finding the real time of a change under both step and liner changes.

5 Conclusions

The accurate and precise estimate of the change point after an out-of-control signal issued by a control chart is an extremely crucial for quality practitioners and leads to saving cost, time and effort for finding the assignable causes and as a result benefit for senior managers. Confining the search only to the time of the signal is likely to be ineffective since the actual change may have taken place a substantial amount of time before the signal. In this paper, we considered a process with a logistic profile quality characteristic in which the response variable follows a binomial distribution. An MLE change point estimator based on the regression parameters was proposed to identify the real time of a change in logistic profiles. In addition, a EWMA-based control chart was applied to detect the out-of-control signal in the process. The performance of the proposed change point estimator was evaluated under both step shift and drift. Although the change point estimator was proposed for finding the real time of a step shift, the simulation results showed the adequate performance of the change point estimator under different rate of the drift as well as different magnitudes of the step shift. The research findings strongly supported our proposed approach. However, one limitation needs to be addressed. This study assumed that the type of the change is single step shift. However, based on the literature, there are a variety of different change types including multiple step changes, drift, monotonic and sporadic change. Developing this method to the other types of the

change could be investigated by researchers. Furthermore, the other distributions of the exponential family such as Poisson and Gamma can be future researches in this area.

References

Akram, M.A., Saif, A.A. and Rahim, M.A. (2012) 'Quality monitoring and process adjustment by integrating SPC and APC: a review', *Int. J. Industrial and Systems Engineering*, Vol. 11, No. 4, pp.375–405.

Albert, A. and Anderson, J.A. (1984) 'On the existence of maximum likelihood estimates in logistic regression models', *Biometrika*, Vol. 71, No. 1, pp.1–10.

Amiri, A. and Allahyari, S. (2012) 'Change point estimation methods for control charts post-signal diagnostics: a literature review', *Quality and Reliability Engineering International*, Vol. 28, No. 7, pp.637–685.

Amiri, A., Mogouie, H. and Doroudyan, M.H. (2013) 'Multi-objective economic-statistical design of MEWMA control chart', *Int. J. Productivity and Quality Management*, Vol. 11, No. 2, pp.131–149.

Doroudyan, M.H. and Amiri, A. (2013) 'Monitoring correlated variable and attribute quality characteristics based on NORTA inverse technique', *Int. J. Productivity and Quality Management* (to appear).

Jarrett, J.E. and Pan, X. (2009) 'Multivariate process control charts and their use in monitoring output quality: a perspective', *Int. J. Industrial and Systems Engineering*, Vol. 4, No. 5, pp.471–482.

Kang, L. and Albin, S.L. (2000) 'Online monitoring when the process yields a linear profile', *Journal of Quality Technology*, Vol. 32, No. 4, pp.418–426.

Kazemzadeh, R.B., Noorossana, R. and Amiri, A. (2008) 'Phase I monitoring of polynomial profiles', *Communications in Statistics – Theory and Methods*, Vol. 37, No. 10, pp.1671–1686.

Kim, K., Mahmoud, M.A. and Woodall, W.H. (2003) 'On the monitoring of linear profiles', *Journal of Quality Technology*, Vol. 35, No. 3, pp.317–328.

Kull, T.J. and Wacker, J.G. (2010) 'Quality management effectiveness in Asia: the influence of culture', *Journal of Operations Management*, Vol. 28, No. 3, pp.223–239.

Mahmoud, M.A. and Woodall, W.H. (2004) 'Phase I analysis of linear profiles with calibration applications', *Technometrics*, Vol. 46, No. 4, pp.380–391.

Mahmoud, M.A., Parker, P.A., Woodall, W.H. and Hawkins, D.M. (2007) 'A change point method for linear profile data', *Quality and Reliability Engineering International*, Vol. 23, No. 2, pp.247–268.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 1st ed., Chapman & Hall, London, UK.

Mestek, O., Pavlik, J. and Suchnek, M. (1994) 'Multivariate control charts: control charts for calibration curves', *Journal of Analytical Chemistry*, Vol. 350, No. 6, pp.344–351.

Perry, M.B. and Pignatiello, J.J. (2005) 'Estimation of the change point of a poisson rate parameter with a linear trend disturbance', *Quality and Reliability Engineering International*, Vol. 22, No. 4, pp.371–384.

Perry, M.B. and Pignatiello, J.J. (2006) 'Estimation of the change point of a normal process mean with a linear trend disturbance', *Quality Technology and Quantitative Management*, Vol. 3, No. 3, pp.101–115.

Pignatiello, J.J. and Samuel, T.R. (1998) 'Identifying the time of a step change with X control charts', *Journal of Quality Engineering*, Vol. 10, No. 3, pp.521–527.

Pignatiello, J.J. and Samuel, T.R. (2001) 'Estimation of the change point of a normal process mean in SPC applications', *Journal of Quality Technology*, Vol. 33, No. 1, pp.82–95.

Saghaei, A., Mehrjoo, M. and Amiri, A. (2009) 'A CUSUM based method for monitoring simple linear profiles', *The Int. J. Advanced Manufacturing Technology*, Vol. 45, No. 11, pp.1252–1260.

Shang, Y., Tsung, F. and Zou, C. (2011) 'Phase II profile monitoring with binary data and random predictors', *Journal of Quality Technology*, Vol. 43, No. 3, pp.196–208.

Shao, E.Y. and Hou, C.D. (2011) 'A combined MLE and EWMA chart approach to estimating the change point of a gamma process with individual observations', *Int. J. Innovative Computing, Information and Control*, Vol. 7, No. 5, pp.2109–2122.

Sharafi, A., Aminnayeri, M. and Amiri, A. (2012) 'Identifying the time of step change in binary profiles', *The Int. J. Advanced Manufacturing Technology*, Vol. 63, Nos. 1–4, pp.209–214.

Sharafi, A., Aminnayeri, M. and Amiri, A. (2013) 'An MLE approach for estimating the time of step changes in Poisson regression profiles', *Scientia Iranica*, Vol. 20, No. 3, pp.855–860.

Singh, A.K. and Shrivastava, R.L. (2012) 'Total quality management: literature review and an agenda for future research', *Int. J. Productivity and Quality Management*, Vol. 10, No. 3, pp.335–373.

Sousa, R. and Voss, C.A. (2002) 'Quality management re-visited: a reflective review and agenda for future research', *Journal of Operations Management*, Vol. 20, No. 1, pp.91–109.

Talib, F. and Rahman, Z. (2012) 'Total quality management practices in manufacturing and service industries: a comparative study', *Int. J. Advanced Operations Management*, Vol. 4, No. 3, pp.155–176.

Yeh, B., Huwang, L. and Li, Y.M. (2009) 'Profile monitoring for a binary response', *IIE Transaction*, Vol. 41, No. 11, pp.931–941.

Zhang, J., Li, Z. and Wang, Z. (2009) 'Control chart based on likelihood ratio for monitoring linear profiles', *Computational Statistics and Data Analysis*, Vol. 53, No. 4, pp.1440–1448.

Zou, C., Zhang, Y. and Wang, Z. (2006) 'Control chart based on change point model for monitoring linear profiles', *IIE Transactions*, Vol. 38, No. 12, pp.1093–1103.