

监控广义线性Profile数据的变样本容量 动态控制图

邹俊杰, 齐德全*

长春理工大学数学与统计学院, 吉林 长春

收稿日期: 2024年3月19日; 录用日期: 2024年4月15日; 发布日期: 2024年4月22日

摘要

对于离散响应变量的profile数据, 比如二元的、多项的、次序的和Poisson响应的profile, 进行有效的监控越来越受到统计过程控制领域研究人员的关注。这些profile都可看作广义线性模型的特殊情况。本文提出了一种基于加权似然比检验的第二阶段的变样本容量动态控制图用来监控广义线性profile数据。不仅能增强对中小漂移的检测敏感度, 还可以提高检测的速度, 降低抽样成本, 提高企业统计决策的精确度和可靠性。通过平均运行长度来衡量控制图的性能表现, 模拟结果表明, 与静态控制图相比, 所提出的控制图对于模型参数的变化有很好的监控能力。

关键词

广义线性模型, Profile数据, 加权似然比检验, 统计过程控制

Dynamic Control Charts with Variable Sample Size for Monitoring Generalized Linear Profile Data

Junjie Zou, Dequan Qi*

School of Mathematics and Statistics, Changchun University of Science and Technology, Changchun Jilin

Received: Mar. 19th, 2024; accepted: Apr. 15th, 2024; published: Apr. 22nd, 2024

Abstract

In recent years, effective profile monitoring for discrete response variables, such as binary, multinomial, ordinal or Poisson variables, has increasingly attracted interest of researchers in the

area of statistical process control. Such quality characteristics are often modeled as special cases of generalized linear models. A new control chart is developed based on the weighted likelihood ratio test with adaptive sample sizes, which is used to monitoring generalized linear profile data. It can not only enhance the sensitivity of detecting small and medium-sized shift, but it can also improve the speed of detection, reduce sampling costs, and improve the accuracy and reliability of enterprise statistical decision-making. The performance of the control chart is measured by the average running length, and the simulation results show that compared with the static control chart, the proposed control chart has good monitoring ability for changes in model parameters.

Keywords

Generalized Linear Model, Profile Data, Weighted Likelihood Ratio Test, Statistical Process Control

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着现代工业技术的发展,精密制造业的生产过程日趋复杂,有关产品质量的问题已不能被某个或某几个指标的均值或方差(或分布)来完全代表或解释,而是需要用 profile 数据对过程进行更加精细和完整地刻画。Profile 数据指的是在抽样点上,我们所得到的观测值可以看作是一些回归曲线。

统计过程控制中对 profile 数据(连续响应变量或离散响应变量)的监控方法已非常丰富,参见文献[1][2]或专著[3]。关于离散响应变量的 profile, Yeh *et al.* [4]开创性地研究了用于监控二元响应变量的 profile 的第二阶段(Phase II)的控制图。之后,学者们分别提出了用于监控响应变量是 Poisson 的、次序的和多项的 profile 的第二阶段的控制图[5][6][7]。这里,响应变量是二元的、Poisson 的、次序的和多项的 profile 都可以看作是广义线性模型的特例。在广义线性模型的统一框架下, Amiri *et al.* [8]和 Shadman *et al.* [9]建立了第一阶段(Phase I)的控制图, Qi *et al.* [10]建立了第二阶段的控制图。

以上的文献都是以固定的抽样间隔抽取固定数量的样本,而一个非常直观的想法是当一个控制图的检验统计量距离控制线较近,过程失控的可能性比较大时,我们就会以一个较短的抽样间隔抽取较多数量的样本,来达到对当前过程信息更好地了解目的;反之,当一个控制图的检验统计量距离控制线较远,过程失控的可能性比较小时,我们就可以以一个较长的抽样间隔抽取较少数量的样本,这就是动态控制图的思想[11]。尚云艳[12]建立了基于似然比的变样本容量动态控制图监控正态数据。Kazemzadeh *et al.* [13]建立了变样本容量的动态控制图监控简单线性 profile 数据。Mohammadzadeh *et al.* [14]建立了变抽样区间的动态控制图监控二元响应变量的 profile 数据。

综上,关于离散响应变量的 profile,本文考虑在广义线性模型的统一框架下,给出变样本容量的第二阶段控制图监控模型的系数是否发生漂移。通过加权似然比检验建立监控统计量,设置报警限和警戒限,当监控统计量小于警戒限时抽取较少的样本,当监控统计量大于警戒限时抽取较多的样本,当监控统计量大于报警限时,发出过程失控的警报。通过蒙特卡洛模拟,所提出的控制图具有较好的表现。

2. 变点模型

假设在生产过程中收集到的 profile 数据之间是独立的,在第 i 时刻抽取 n_i 个样本,随机样本简记为 (X_{ij}, y_{ij}) , $i=1, 2, \dots, t$, $j=1, \dots, n_i$, 其中 X_{ij} 是解释变量, y_{ij} 是响应变量, n_i 为随时间变化的样本量。

进一步假设数据满足以下广义线性模型:

1) 响应变量 $Y_i = (y_{i1}, \dots, y_{in_i})^T$ 是独立同分布, 服从于指数分布族的随机变量, 其中 y_{ij} 是在时刻 i 的第 j 个样本的响应变量, 响应变量 y_{ij} 的概率密度函数为

$$f(y_{ij}; \theta_{ij}) = \exp[y_{ij}b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})], \quad i=1, 2, \dots, t, \quad j=1, \dots, n_i$$

其中 $b(\theta_{ij}), c(\theta_{ij}), d(y_{ij})$ 是已知函数, θ_{ij} 是自然参数。

2) 解释变量

$$X_i = \begin{pmatrix} X_{i1}^T \\ \vdots \\ X_{in_i}^T \end{pmatrix} = \begin{pmatrix} x_{i1,1} & \cdots & x_{i1,p} \\ \vdots & & \vdots \\ x_{in_i,1} & \cdots & x_{in_i,p} \end{pmatrix}$$

为一个设定的 $n_i \times p$ (且 $p < n_i$) 维矩阵, 其中 $X_{ij}^T = (x_{ij,1}, \dots, x_{ij,p})$, $i=1, 2, \dots, t, \quad j=1, \dots, n_i$ 。

3) 解释变量 X_{ij}^T 与回归系数 β 的一个单调链接函数如下:

$$g(\mu_{ij}) = \eta_{ij} = X_{ij}^T \beta$$

其中 $\mu_{ij} = E(Y_{ij})$, 且回归系数向量 $\beta = (\beta_1, \dots, \beta_p)^T$, $p < n_i, \quad i=1, 2, \dots, t, \quad j=1, 2, \dots, n_i$ 。

假设在一个未知的 τ 时刻, 生产过程失控, 即系数 β 由可控时的 β_{IC} 变为失控时的 β_{OC} 。于是, 在每一时刻 $t=1, 2, \dots$ 监控以下变点模型:

$$\begin{cases} H_0 : \beta = \beta_{IC} \\ H_1 : \beta = \beta_{OC} \end{cases}$$

3. 在线监控方法

首先求加权似然比监控统计量, 在第 t 时刻的对数似然函数为

$$l'_i(\beta) = \sum_{j=1}^{n_i} [y_{ij}b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})], \quad i=1, 2, \dots, t$$

从而加权对数似然函数为

$$wl'_t(\beta) = \sum_{i=0}^t w_i l'_i(\beta) = \sum_{i=0}^t w_i \sum_{j=1}^{n_i} [y_{ij}b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})]$$

其中 $w_0 = (1-\lambda)^t, \quad w_i = \lambda(1-\lambda)^{t-i}, \quad i=1, 2, \dots, t, \quad \lambda \in (0, 1)$ 是光滑参数。

令 $\mathfrak{I}^{(m-1)} = \sum_{i=1}^t w_i \mathfrak{I}_i^{(m-1)}$ 为信息矩阵, $U^{(m-1)} = \sum_{i=0}^t w_i U_i^{(m-1)}$ 为得分矩阵, 由牛顿迭代法[10], 根据公式

$$\beta^{(m)} = \beta^{(m-1)} + [\mathfrak{I}^{(m-1)}]^{-1} U^{(m-1)}$$

可得系数 β 的加权极大似然估计 $\hat{\beta}_t = \arg \max_{\beta} wl'_t(\beta)$ 。

于是, 构造如下变样本容量的加权似然比检验统计量(VSS-WLRT)

$$W'_t = 2 \sum_{i=0}^t w_i \left\{ \sum_{j=1}^{n_i} [y_{ij}b(\hat{\theta}_{ij}) + c(\hat{\theta}_{ij}) + d(y_{ij})/\sqrt{n_i}] - \sum_{j=1}^{n_i} [y_{ij}b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})/\sqrt{n_i}] \right\}$$

然后, 根据公式 $n_1 p_1 + n_2 p_2 = n_0$ 设置报警限 h 和警戒限 g , 其中 $p_1 = P[W'_t \in (g, h)], \quad p_2 = P[W'_t \in (0, g)],$ n_0 为固定样本容量的控制图的抽取样本量, $n_1 < n_0 < n_2$ 且 $p_1 + p_2 = 1$ 。

在时刻 t , 当统计量 $g < W'_t < h$ 时, 在下一时刻抽取较大的容量为 n_2 的样本; 当统计量 $W'_t < g$ 时, 在

下一时刻抽取较小的容量为 n_1 的样本; 当统计量 $W'_i > h$ 时, 此时控制图发出系数 β 发生漂移的警报。变样本容量动态控制图监控过程的流程图如图 1 所示。

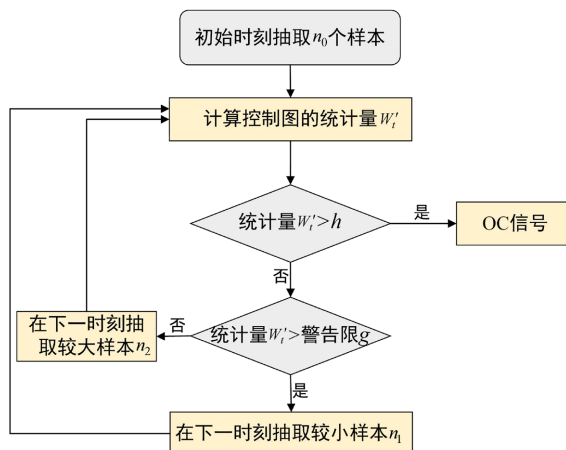


Figure 1. Flow chart of dynamic control chart monitoring with variable sample size
图 1. 变样本容量动态控制图监控的流程图

4. 统计模拟与示例

通过蒙特卡洛模拟将本文所提出的方法(简记为 VSS-WLRT)与 Qi *et al.* [10]的 WLRT 方法进行比较。参考 Qi *et al.* [10]响应变量是 Poisson 分布的统计模拟, 当过程参数可控时给定 $\beta_i = \beta_{IC} = (1, 1)^T$, 过程参数失控时 $\beta_i = \beta_{IC} + \Delta$, 其中 $\Delta = (\delta_1 \sigma_1, \delta_2 \sigma_2)^T$, δ_1, δ_2 为漂移量且不同时为 0, $\sigma_1 = 0.35181$, $\sigma_2 = 0.50947$ 。参考文献[13], 在进行动态控制图与静态控制图比较时, 应设定具有相同的平均采样率, 本文设定 $p_1 = p_2 = 0.5$, 此时样本容量设置为 $(n_0, n_1, n_2) = (5, 2, 8)$ 时, 控制图记为 VSS-WLRT₁; 样本容量设置为 $(n_0, n_1, n_2) = (5, 3, 7)$ 时, 控制图记为 VSS-WLRT₂。通过二分法调整控制限使得不同控制图可控时的平均运行长度 ARL_0 接近 370, 然后比较失控时的平均运行长度 ARL_1 。失控状态下的 ARL_1 越小, 控制图报警越早, 监控效果越好。在理论上, WLRT 和 VSS-WLRT 控制图的监控效果都受到光滑参数的影响, 较大的光滑参数对大漂移较有效, 较小的光滑参数对小漂移较有效。因此, 在光滑参数 λ 分别取 0.05, 0.1 和 0.2 的情况下, 进行 5000 次重复模拟得到各控制图的 ARL_1 , 见表 1、表 2 和表 3。

Table 1. ARL_1 the control chart of the WLRT, VSS-WLRT₁, VSS-WLRT₂ when $\lambda = 0.05$

表 1. $\lambda = 0.05$ 时 WLRT、VSS-WLRT₁、VSS-WLRT₂ 的失控 ARL_1

(δ_1, δ_2)	WLRT	VSS-WLRT ₁	VSS-WLRT ₂
(0.2, 0)	143.5650	87.8400	74.7350
(0, 0.2)	132.8110	67.8060	58.3724
(0.2, 0.2)	121.5890	62.0630	52.4640
(0.5, 0)	72.0090	34.3200	33.2624
(0.32, 0.32)	31.1010	18.9800	11.3245
(0.44, 0.44)	19.6900	8.2140	7.1660
(0.59, 0.59)	7.5540	3.2420	3.8320
(1.0, 1.0)	3.0070	1.6410	1.3820

由表 1~表 3 可以看出, 相同漂移的情况下, VSS-WLRT₁ 与 VSS-WLRT₂ 两个动态控制图的 ARL₁ 均比静态 WLRT 控制图的更小, 从而监控效果更好。例如, 在 $\lambda = 0.05$, 漂移量 $(\delta_1, \delta_2) = (0.2, 0)$ 时, WLRT 控制图的 ARL₁ 为 143.5650, VSS-WLRT₁ 的 ARL₁ 为 87.8400, VSS-WLRT₂ 的 ARL₁ 为 74.7350; 漂移量 $(\delta_1, \delta_2) = (0.59, 0.59)$ 时, WLRT 控制图的 ARL₁ 为 7.5540, 而 VSS-WLRT₁ 的 ARL₁ 为 3.2420, VSS-WLRT₂ 的 ARL₁ 为 3.8320。因此表明, 在不同的漂移量下, 变样本容量的动态控制图的报警速度均比静态控制图更快。

Table 2. ARL₁ the control chart of the WLRT, VSS-WLRT₁, VSS-WLRT₂ when $\lambda = 0.1$

表 2. $\lambda = 0.1$ 时 WLRT、VSS-WLRT₁、VSS-WLRT₂ 的失控 ARL₁

(δ_1, δ_2)	WLRT	VSS-WLRT ₁	VSS-WLRT ₂
(0.2, 0)	240.9420	145.2714	123.9220
(0, 0.2)	194.6180	109.7243	101.4659
(0.2, 0.2)	163.9220	98.7471	84.2321
(0.5, 0)	45.4610	27.3200	23.2624
(0.32, 0.32)	21.6944	8.2350	6.4324
(0.44, 0.44)	12.0610	4.2357	4.5435
(0.59, 0.59)	4.8470	2.7233	2.3134
(1.0, 1.0)	1.7410	1.4172	1.1321

由表 1~表 3 还可以看出, 三个控制图的监控效果受到光滑参数的影响。 $\lambda = 0.2$, 对大漂移比较有效; $\lambda = 0.05$, 对小漂移比较有效。例如, 当漂移量 $(\delta_1, \delta_2) = (0.59, 0.59)$ 时, $\lambda = 0.2$ 的 VSS-WLRT₁ 控制图的 ARL₁ = 1.4172, $\lambda = 0.05$ 的 VSS-WLRT₁ 控制图的 ARL₁ = 3.2420; 当漂移量 $(\delta_1, \delta_2) = (0.2, 0)$ 时, $\lambda = 0.05$ 的 VSS-WLRT₁ 控制图的 ARL₁ = 87.8400, $\lambda = 0.2$ 的 VSS-WLRT₁ 控制图的 ARL₁ = 185.3000, 这一结果与理论一致。整体上看, 与 WLRT 相比, VSS-WLRT 控制图对中小漂移的检测敏感度有所提高。

Table 3. ARL₁ the control chart of the WLRT, VSS-WLRT₁, VSS-WLRT₂ when $\lambda = 0.2$

表 3. $\lambda = 0.2$ 时 WLRT、VSS-WLRT₁、VSS-WLRT₂ 的失控 ARL₁

(δ_1, δ_2)	WLRT	VSS-WLRT ₁	VSS-WLRT ₂
(0.2, 0)	280.2265	185.3000	172.3120
(0, 0.2)	269.9455	164.7223	154.3224
(0.2, 0.2)	144.0915	128.4534	115.4320
(0.5, 0)	33.9685	21.9940	17.3210
(0.32, 0.32)	18.3190	8.1250	6.3571
(0.44, 0.44)	3.0240	2.7233	2.3280
(0.59, 0.59)	2.0985	1.4172	1.3520
(1.0, 1.0)	1.4900	1.0720	1.0300

为了更好地理解所提出方法的应用, 本文在 $\lambda = 0.05$, 漂移量 $(\delta_1, \delta_2) = (0.5, 0)$ 时, 进行一次过程监

控, 给出如图 2 的示例(WLRT 控制图中 n 取 10)。由图 2 可以看出, 固定样本量的 WLRT 控制图在 $t=37$ 时发出失控警报, VSS-WLRT₁ 控制图在 $t=23$ 时发出失控警报, VSS-WLRT₂ 在 $t=16$ 时发出失控警报。于是动态控制图所需的平均样本量更少, 能达到降低成本的目的。

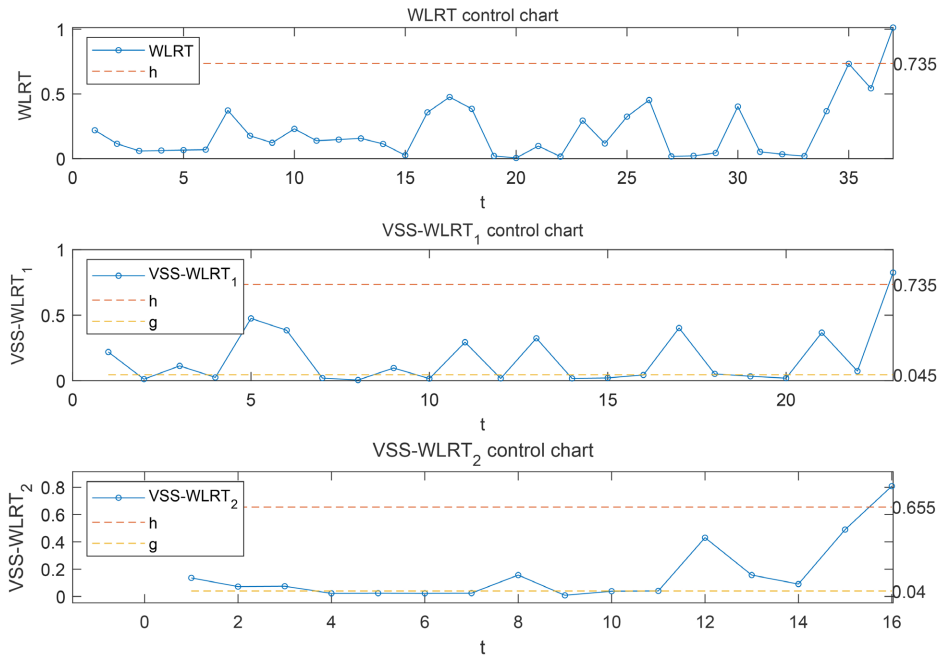


Figure 2. WLRT, VSS-WLRT₁, VSS-WLRT₂ control chart
图 2. WLRT、VSS-WLRT₁、VSS-WLRT₂ 控制图

为了验证所提出方法对其它模型的适用性, 本文扩展了 Goeman 和 Cessie [15] 所讨论的多项式逻辑回归模型作为一个控制图性能比较的示例。在这里, 我们只考虑解释变量是分类变量的情况, 考虑一个有四个类别的响应变量 Y 。令 π_1, \dots, π_4 表示响应变量各自类别的概率, 且 $\pi_1 + \dots + \pi_4 = 1$ 。假设有三个取值为 $-1, 0, 1$ 的协变量 x_1, x_2, x_3 满足:

$$\begin{cases} \log\left(\frac{\pi_2}{\pi_1}\right) = 2x_1 + \delta x_1^2, \\ \log\left(\frac{\pi_3}{\pi_1}\right) = 2x_2, \\ \log\left(\frac{\pi_4}{\pi_1}\right) = 2x_3. \end{cases}$$

取 $\lambda=0.1$, $ARL_0=370$, $\tau=20$, 可控时 $\delta=1$, 失控时 $\delta=1.6$ 。在每个时间点 t , 从 3 个协变量值的 $3^3=27$ 种可能组合中随机抽取, 获得 25 个观察数据集。VSS-WLRT₁ 控制图样本容量设置为 $(n_0, n_1, n_2)=(25, 20, 30)$, VSS-WLRT₂ 控制图样本容量设置为 $(n_0, n_1, n_2)=(25, 15, 35)$, WLRT 控制图样本容量 n 设置为 25。进行一次过程监控, 绘制控制图如图 3 所示。

由图 3 可以看出, 固定样本量的 WLRT 控制图在 $t=47$ 时发出失控警报, VSS-WLRT₁ 控制图在 $t=28$ 时发出失控警报, VSS-WLRT₂ 在 $t=25$ 时发出失控警报。变化样本量的动态控制图监测过程失控速度更快, 且所需样本量更少, 说明所提出方法对监控多项式逻辑回归模型也是适用的。

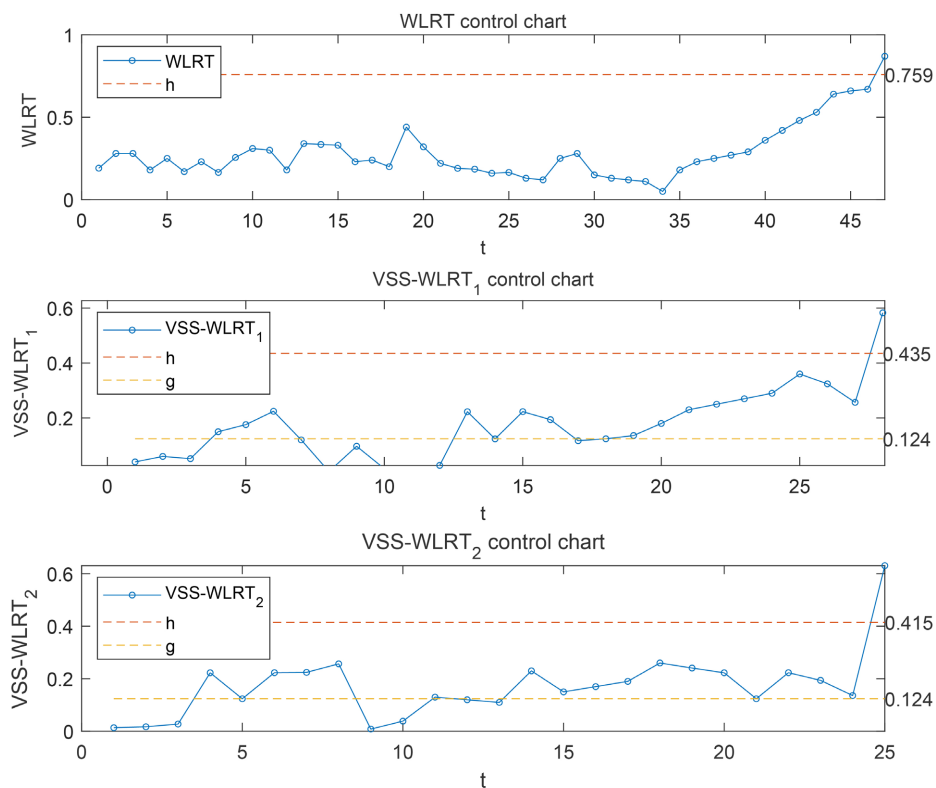


Figure 3. WLRT, VSS-WLRT₁, VSS-WLRT₂ control chart
图 3. WLRT、VSS-WLRT₁、VSS-WLRT₂控制图

5. 结论

采用可变样本容量的方法建立动态的控制图, 可以更有效地监测到生产过程失控, 降低统计过程控制中的抽样成本。本文针对响应变量是离散型的 profile 数据, 给出了在广义线性模型的统一框架下基于加权似然比检验的动态控制图。通过二分法统计模拟确定报警限和警戒限, 当监控统计量小于警戒限时抽取较少的样本, 当监控统计量大于警戒限时抽取较多的样本, 达到降低抽样成本的目的。通过蒙特卡洛模拟得出, 所提出的控制图失控时平均运行长度较小, 具有较好的表现。广义线性模型是线性模型的推广, 所以本文所提出的方法很容易应用于线性模型。本文的动态控制图是基于变样本容量提出, 在之后的研究中还可以考虑变抽样区间对结果的影响。

基金项目

吉林省教育厅项目(JJKH20210809KJ)、国家自然科学基金面上项目(12271271)。

参考文献

- [1] Woodall, W.H. (2007) Current Research on Profile Monitoring. *Producã*, **17**, 420-425. <https://doi.org/10.1590/S0103-65132007000300002>
- [2] Woodall, W.H. and Montgomery, D.C. (2014) Some Current Directions in the Theory and Application of Statistical Process Monitoring. *Journal of Quality Technology*, **46**, 79-94. <https://doi.org/10.1080/00224065.2014.11917955>
- [3] Noorossana, R., Saghaei, A. and Amiri, A. (2011) *Statistical Analysis of Profile Monitoring*. John Wiley & Sons, Inc., Hoboken, NJ. <https://doi.org/10.1002/9781118071984>
- [4] Yeh, A.B., Huwang, L. and Li, Y.M. (2009) Profile Monitoring for a Binary Response. *IIE Transactions*, **41**, 931-941.

- <https://doi.org/10.1080/07408170902735400>
- [5] Amiri, A., Koosha, M. and Azhdari, A. (2011) Profile Monitoring for Poisson Responses. 2011 *IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, Singapore, 6-9 December 2011, 1481-1484. <https://doi.org/10.1109/IEEM.2011.6118163>
- [6] Noorossana, R., Aminnayeri, M. and Izadbakhsh, H. (2013) Statistical Monitoring of Polytomous Logistic Profiles in Phase II. *Scientia Iranica E*, **20**, 958-966.
- [7] Noorossana, R., Saghaei, A., Izadbakhsh, H. and Aghababaei, O. (2013) Monitoring Multinomial Logit Profiles via Log-Linear Models. *International Journal of Industrial Engineering & Production Research*, **24**, 137-142.
- [8] Amiri, A., Koosha, M., Azhdari, A. and Wang, G. (2015) Phase I Monitoring of Generalized Linear Model-Based Regression Profiles. *Journal of Statistical Computation and Simulation*, **85**, 2839-2859. <https://doi.org/10.1080/00949655.2014.942864>
- [9] Shadman, A., Mahlooji, H., Yeh, A.B. and Zou, C. (2015) A Change Point Method for Monitoring Generalized Linear Profiles in Phase I. *Quality and Reliability Engineering International*, **31**, 1367-1381. <https://doi.org/10.1002/qre.1671>
- [10] Qi, D., Wang, Z., Zi, X. and Li, Z. (2016) Phase II Monitoring of Generalized Linear Profiles Using Weighted Likelihood Ratiocharts. *Computers & Industrial Engineering*, **94**, 178-187. <https://doi.org/10.1016/j.cie.2016.01.022>
- [11] 王兆军. 关于动态质量控制图的设计理论[J]. 应用概率统计, 2002, 18(3): 316-333.
- [12] 尚云艳. 基于似然比的指数加权滑动平均动态控制图[J]. 南开大学学报(自然科学版), 2021, 54(1): 41-47.
- [13] Kazemzadeh, R.B., Amiri, A. and Kouhestani, B. (2016) Monitoring Simple Linear Profiles Using Variable Sample Size Schemes. *Journal of Statistical Computation & Simulation*, **86**, 2923-2945. <https://doi.org/10.1080/00949655.2016.1138115>
- [14] Mohammadzadeh, M., Yeganeh, A. and Shadman, A. (2021) Monitoring Logistic Profiles Using Variable Sample Interval Approach. *Computers & Industrial Engineering*, **158**, 107438. <https://doi.org/10.1016/j.cie.2021.107438>
- [15] Goeman, J.J. and Cessie, S. (2016) A Goodness-of-Fit Test for Multinomial Logistic Regression. *Biometrics*, **62**, 980-985. <https://doi.org/10.1111/j.1541-0420.2006.00581.x>