Phase II Monitoring of Generalized Linear Profiles Using Weighted Likelihood Ratio Charts

Dequan Qi, Zhaojun Wang, Xuemin Zi, and Zhonghua Li

Notes by Taylor Grimm

July 31, 2023

Background Information

Statistical process monitoring (SPM) is a popular approach for detecting abnormal behavior in one or more variables in a process over time.

In SPM, there are two phases we talk about: phase I and phase II.

- Phase I is the phase where we define our model parameters and determine control limits (What is considered normal vs. abnormal?)
- Phase II is when we monitor new observations for abnormal behavior using the control limits previously defined (Is a new observation considered normal or abnormal when compared to the Phase I information?)

This paper focuses on monitoring generalized *linear profiles*, which is where we look for changes not in a set of response variables of interest, but in the relationships (profiles) between a set of explanatory variables and the responses.

Statistical Model

1. Response variables $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN})^T$ from an exponential family of the form

$$f(y_{ij}; \theta_{ij}) = \exp\left[y_{ij} b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})\right], i = 1, \ldots, t, j = 1, \ldots, N,$$

   where $b(\cdot), c(\cdot)$, and $d(\cdot)$ are known functions and θ_{ij}'s are the parameters of the exponential family of distributions.

2. Explanatory variables

$$\mathbf{X}_i = \begin{pmatrix} X_{i1}^T \\ \vdots \\ X_{iN}^T \end{pmatrix} = \begin{pmatrix} x_{i11} \cdots x_{i1p} \\ \vdots \qquad \vdots \\ x_{iN1} \cdots x_{iNp} \end{pmatrix}$$

   where $X_{ij}^T = (x_{ij1}, \cdots, x_{ijp},\ i = 1, \ldots, t, j = 1, \ldots, N$ can be combined linearly with coefficients $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T \quad (p < N)$ to form the linear predictor $\eta_{ij} = X_{ij}^T \boldsymbol{\beta}$.

3. Monotone link function $g(\cdot)$ such that

$$g(\mu_{ij}) = \eta_{ij} = X_{ij}^T \boldsymbol{\beta}, i = 1, \ldots, t, j = 1, \ldots, N,$$

   where $\mu_{ij} = E(Y_{ij})$.

We are specifically looking for changes in $\boldsymbol{\beta}$. We assume (from Phase I) that we know the true in-control $\boldsymbol{\beta}$, denoted $\boldsymbol{\beta}_{IC}$, and are looking to test

$$H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_{IC},$$
$$H_a : \boldsymbol{\beta} \neq \boldsymbol{\beta}_{IC}$$

at each time point.

Existing Work

For the ith profile, the log-likelihood function is given by

$$l_i(\boldsymbol{\beta}) = \sum_{j=1}^{N} [y_{ij} b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})],$$

and the maximum likelihood estimate for $\boldsymbol{\beta}$ can be obtained by

$$\mathbf{b}_i^{(m)} = \mathbf{b}_i^{(m-1)} + [\Im_i^{(m-1)}]^{-1} U_i^{(m-1)},$$

where $\mathbf{b}_i^{(m)}$ is the vector of estimates of $\boldsymbol{\beta}$ at iteration m, $[\Im_i^{(m-1)}]^{-1}$ is the inverse of the information matrix, and $U_i^{(m-1)}$ is the score vector (the gradient of the log-likelihood at time i).

The likelihood ratio test (LRT) monitoring statistic is given by

$$LRT_i = 2[l_i(\hat{\boldsymbol{\beta}}_i) - l_i(\boldsymbol{\beta}_{IC})], \quad i = 1, 2, \ldots.$$

LRT-EWMA Control Chart

The LRT_i values are first normalized (and termed NL_i), and the exponentially weighted moving average (LRT-EWMA) statistic is computed as

$$LE_i = \lambda NL_i + (1 - \lambda) LE_{i-1}, i = 1, 2, \ldots,$$

where λ is the smoothing parameter and $LE_0 = 0$.

The multivariate EWMA (MEWMA) statistic can be computed in the following steps:

1. $Z_i = (\mathbf{X}_i^T W \mathbf{X}_i)^{1/2} (\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_{IC})$, where W is an $n \times n$ diagonal matrix.

2. $E_i = \lambda Z_i (1 - \lambda) E_{i-1}, i = 1, 2, \ldots.$

3. The MEWMA monitoring statistic is now calculated by $M_i = E_i^T E_i$.

Proposed WLRT Control Chart

Up to time point t, we can derive the weighted-log-likelihood function as

$$wl_t(\boldsymbol{\beta}) = \sum_{i=0}^{t} w_i l_i(\boldsymbol{\beta}) = \sum_{i=0}^{t} w_i \left\{ \sum_{j=1}^{N} [y_{ij} b(\theta_{ij}) + c(\theta_{ij}) + d(y_{ij})] \right\},$$

where $w_0 = (1-\lambda)^t, w_i = \lambda(1-\lambda)^{t-i}, i = 1, \ldots, t$, and $\lambda \in (0,1)$ is a smoothing parameter. The observations of $(\mathbf{X}_0, \mathbf{Y}_0)$ are a pseudo "sample" from the IC dataset.

For a given λ, the WLRT statistic is

$$W_t = 2[wl_t(\hat{\boldsymbol{\beta}}_t) - wl_t(\boldsymbol{\beta}_{IC})], \tag{1}$$

where $\hat{\boldsymbol{\beta}}_t = \operatorname{argmax}_{\boldsymbol{\beta}} wl_t(\boldsymbol{\beta})$ is the maximum weighted likelihood estimator of $\boldsymbol{\beta}$ (algorithm given in the Appendix).

We can then flag observations as out-of-control when W_t exceeds a predetermined upper control limit.

- Upper control limit for WLRT chart is determined algorithmically to ensure a desired IC average run length (ARL).

Simulation Study

The WLRT control chart is compared to the following methods:

- EWMA-GLM
- LRT
- LRT-EWMA
- MEWMA

The following data is generated:

- Response variables Y_{ij} are independent Poisson random variables, $j = 1, \ldots, 10$.
- Explanatory variables X_{ij} such that

$$\begin{aligned}(X_{i1}, X_{i2}, \ldots, X_{i10}) &= \begin{pmatrix} x_{i11} & x_{i21} & \cdots & x_{i10,1} \\ x_{i12} & x_{i22} & \cdots & x_{i10,2} \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0.1 & 0.2 & \cdots & 1.0 \end{pmatrix}\end{aligned}$$

  when the design points are fixed

  Otherwise, for the ith profile, 9 different design points randomly take values in the above equation.

- The log link function such that $g(\mu_{ij}) = \log(\mu_{ij}) = X_{ij}^T \boldsymbol{\beta}$, where $\mu_{ij} = E(Y_{ij})$.

Assume that our IC parameters $\boldsymbol{\beta}_{IC}$ are $(1,1)^T$. The OC profile parameters at the ith sample profile are

$$\boldsymbol{\beta}_i = \begin{cases} \boldsymbol{\beta}_{IC}, & i = 1, \ldots, \tau \\ \boldsymbol{\beta}_{OC} = \boldsymbol{\beta}_{IC} + \Delta, & i = \tau + 1, \ldots, \end{cases}$$

where $\Delta = (\delta_1\sigma_1, \delta_2\sigma_2)^T, \delta_1 \neq 0$ or $\delta_2 \neq 0$, and $\sigma_1 = 0.35181, \sigma_2 = 0.50947$ are the standard deviation of the MLE of the profile parameters.

Control limits h for each chart are adjusted to make the IC ARL (ARL_0) as close to 370 as possible.

- Average run length (ARL) measures the average number of observations before an alarm is raised. An IC ARL of 370 implies that we expect 1 alarm, on average, for every 370 IC observations, which implies a false alarm rate of $1/370 = 0.0027$.

Fixed Design Points

In-control Performance

- $F_{30} = P_{IC}(\text{RL} \leq 30)$ examines false alarms for the first 30 observations
- 5000 replications

Ideally, the IC run length distribution should be close to geometric.
When the run length distribution is geometric, the standard deviation of the run length (SDRL) should be approximately equal to ARL_0, and Q(.10), Median, Q(.90) and F_{30} are about 38, 256, 850 and 0.080 respectively.

Table 2
IC comparisons.

	h	ARL_0	SDRL	Q(.10)	Median	Q(.90)	F_{30}
EWMA-GLM0.05	9.90756	370	520	2	154	1040	0.355
EWMA-GLM0.2	11.60040	370	409	4	239	922	0.183
LRT	11.89143	370	369	38	258	850	0.083
LRT-EWMA0.05	0.68820	370	348	63	261	813	0.022
LRT-EWMA0.2	1.47790	370	370	43	256	850	0.069
MEWMA0.05	0.31650	370	338	57	268	826	0.030
MEWMA0.2	1.55000	370	353	43	265	839	0.068
WLRT0.05	0.22710	370	369	47	255	853	0.055
WLRT0.2	1.22170	370	371	40	257	845	0.075

- The EWMA-GLM control chart has very large short-run false alarms. For example, F_{30} can be as large as 0.183 when $\lambda = 0.2$, and 0.355 when $\lambda = 0.05$.

Out-of-control Performance

The "true" detection ability is given by

$$\gamma_t = Pr_{\mathrm{OC}}(\mathrm{RL} \leq t) - Pr_{\mathrm{IC}}(\mathrm{RL} \leq t),$$

and the relative mean index (RMI) is

$$\mathrm{RMI} = \frac{1}{M} \sum_{i=1}^{M} \frac{\mathrm{ARL}_{\Delta l} - \mathrm{MARL}_{\Delta l}}{\mathrm{MARL}_{\Delta l}},$$

where

- M is the total number of shifts considered
- $\mathrm{ARL}_{\Delta l}$ is the ARL_1 of the given control chart when detecting a parameter shift of magnitude Δl
- $\mathrm{MARL}_{\Delta l}$ is the smallest among all ARL_1 values of the charts considered when detecting the shift Δl.

For comparing methods,

- a larger γ_t is better
- a smaller RMI is better

We also compare the conditional expected delay (CED) which is defined by

$$\mathrm{CED} = E[\mathrm{RL} - \tau | \mathrm{RL} > \tau].$$

A control chart with a smaller CED is considered better than another one (smaller delay). Figure 2 shows the true detection capability for selected δ_1, δ_2 pairs for different methods when $t \leq 100$. (for true $\tau = 0$)

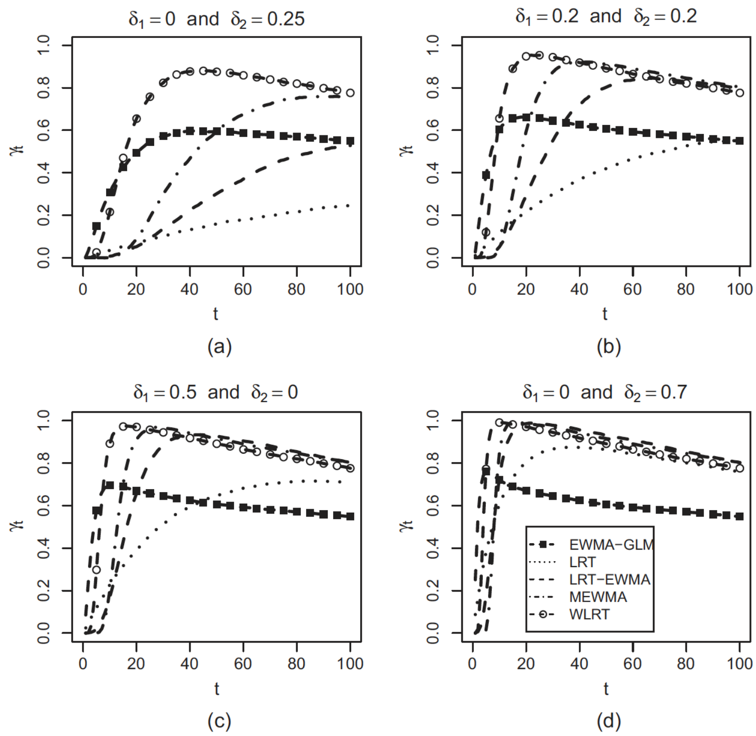

Fig. 2. The "true" detection capability for the Poisson profiles ($\lambda = 0.05$). The legend in the last plot is applicable for all the others.

- EWMA-GLM ($\lambda = 0.05$) performs the best overall
- LRT control chart performs better at detecting large shifts, while LRT-EWMA, MEWMA, and WLRT perform better at detectiong small to medium shifts
- Performance for LRT-EWMA, MEWMA, and WLRT depends on λ
  - smaller λ: better for detecting small shifts
  - larger λ: better for detecting larger shifts

Example

Multinomial logistic regression with a response variable Y (different categories for different classes of quality) with four categories with probabilities $\pi_1, \ldots, \pi_4$.

Three covariates:

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

where x_1, x_2, and x_3 take values $-1, 0$, and 1.
At each time t, we obtain a data set of 25 observations taken randomly from each of the $3^3 = 27$ possible combinations of the three covariate values.

- Newton-Raphson is used to estimate model parameters, and control limits are adjusted to make ARL_0 as close to 370 as possible based on 5000 replicates.
- The first 20 profiles are generated from IC ($\delta = 1$) conditions, and the rest are from OC ($\delta = 1.6$) conditions.
- λ is chosen as 0.1 for LRT-EWMA and WLRT

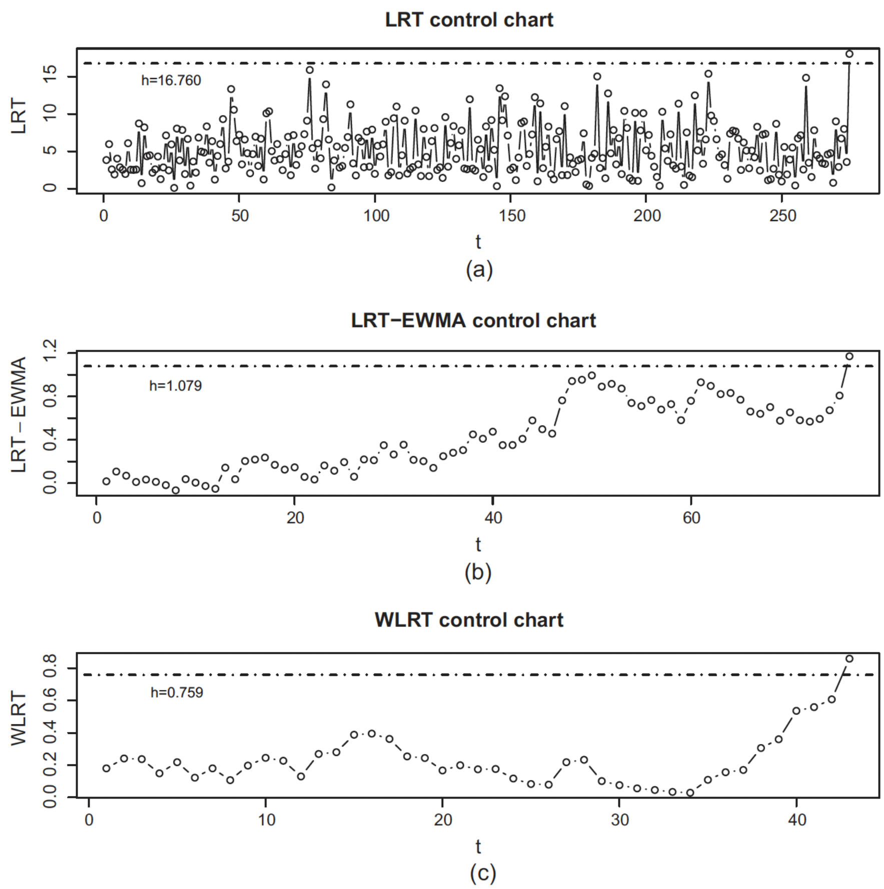

Fig. 4. The LRT, LRT-EWMA and WLRT control charts for the multinomial profiles.

"we can see that the performance of the WLRT chart is satisfactory"