T^2 Based Methods for Monitoring Gamma Profiles

Amirhossein Amiri and Mehdi Koosha
Department of Industrial Engineering
Shahed University, Tehran, Iran

Armaghan Azhdari
Department of Industrial Engineering
PNU, Tehran, Iran

Abstract

Profile is a relationship between a response variable and one or more explanatory variables that should be monitored over time. There are many researches in this area but in most of them, it is assumed that the response variable is normally distributed. However, sometimes this relationship is better characterized by Generalized Linear Models. Yeh et al. (2009) proposed five T^2 based methods for monitoring of logistic regression profiles in which the response variable follows Bernoulli distribution. This paper evaluates the performance of the T^2 methods for monitoring Gamma response profiles in Phase I via simulation studies. The evaluation is done under different magnitudes of step shifts and drifts in terms of the signal probability.

Keywords
Statistical process control; Profile monitoring; T^2 control chart; Gamma profile; Phase I

1. Introduction
In many real case problems, the relationship between response variable and explanatory variables which is called as profile is desirable to be monitored over time instead of the response variable itself. According to the type of this relationship, profiles are classified into categories such as Simple linear profiles, multiple linear profiles, Polynomial profiles, multivariate linear profiles, non-linear profiles, Logistic profiles, and so on. Many studies have been done by researchers for monitoring different types of profiles. For example Kang and Albin (2000), Kim et al. (2003), Mahmoud and Woodal (2004), Mahmoud et al. (2007), Saghaei et al. (2009) Chen and Nembhard (2011) and Noorossana et al. (2004) have proposed some methods for monitoring simple linear profiles.

Zou et al. (2007), and Mahmoud (2008) proposed some approaches for monitoring multiple linear profiles. Amiri et al. (2010) and Kazemzadeh et al. (2008) have contributed in the area of polynomial regression profiles. Eyvazian et al. (2011) and Noorossana et al. (2010) have proposed methods on monitoring multivariate linear profiles. Williams et al. (2007) and Vaghefi et al. (2009) developed some methods for monitoring nonlinear profiles in Phases I and II, respectively.

In all of the aforementioned researches, distribution of the response variable is assumed to be normal. However, in some real case problems, response variable follows other Exponential family distributions such as Bernoulli, Poisson, Exponential, Gamma, and etc. Yeh et al. (2009) proposed five T^2 based methods in order to monitor logistic profiles in which the response variable follows Bernoulli distribution. Amiri et al. (2011) evaluated two of the best T^2 methods proposed by Yeh et al. (2009) for monitoring Poisson regression profiles.

This paper concentrates on phase I monitoring of Gamma regression profiles. Gamma distribution is sum of m identical independent exponential variables. There are many real cases in which the response variable follows Gamma distribution. For example when each observation represents the time interval between the occurrence of two successive defects in a process, the observations follow exponential distributions and if we consider the sum of these exponential variables in each level of the explanatory variable as an observation, then each observation follows Gamma distribution.

The rest of this paper is classified as follows: Section 2 proposes the procedure of parameter estimation in Gamma regression profiles. The T^2 methods considered in this paper are explained in Section 3. Section 4 is dedicated to the numerical example to evaluate the performance of the proposed methods. Our concluding remarks are given in Section 5.

2. Gamma Regression Parameter Estimation

Suppose that there are m independent profiles (i=1, 2,…., m) and the observations in each level of the explanatory variable (j=1,2,…,n) are repeated for k (k=1,2,…,K) times. Vector of explanatory variables for each profile is as follows:

$$x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]$$

Let Z_{ijk} be the kth observation of jth setting level in i^{th} profile which is assumed to follow an exponential distribution with mean $\frac{1}{\lambda_{ij}}$. We define Y as sum of K observations in j^{th} level of ith profile. Hence,

$$Y_{ij} = \sum_{k=1}^{K} Z_{ikj}, \tag{1}$$

As Y_{ij} is sum of K exponential observations, it follows Gamma distribution. So the density probability function of Y_{ij} is as follows:

$$f_{Y_{ij}}(y_{ij}) = \frac{\lambda_{ij}^{K}}{\Gamma(K)} y_{ij}^{K-1} e^{-\lambda_{ij} y_{ij}} \tag{2}$$

Hence, $E(Y_{ij}) = \frac{K}{\lambda_{ij}}$ and $\mathrm{var}(Y_{ij}) = \frac{K}{\lambda_{ij}^2}$. In addition, let $\eta_{ij} = \beta_0 + \beta_1 x_{ij}$.

We use Log link function (g) to make a relationship between λ_{ij} and η_{ij}. Hence,

$$\eta_{ij} = g(\lambda_{ij}) = \log\left(E(Y_{ij})\right) = \beta_0 + \beta_1 x_{ij}, \tag{3}$$

consequently,

$$\lambda_{ij} = \frac{K}{\exp\left(\eta_{ij}\right)}, \tag{4}$$

where β_0 and β_1 are the regression parameters of each profile.

Since Y_{ij} in each experimental setting level is independent to other levels' observations, the joint likelihood function of $Y_{i1}, Y_{i2}, \ldots, Y_{in}$ is as follows:

$$L(\lambda_{ij}, y_{ij}) = \frac{\prod_{j=1}^{n} \lambda_{ij}^{K}}{[(K-1)!]^n} \prod_{j=1}^{n} y_{ij}^{K-1} e^{\sum_{j=1}^{n} -\lambda_{ij} y_{ij}} \tag{5}$$

By taking logarithm of both sides of Eq. (5) and replacing λ_{ij} with η_{ij}, the following equation is obtained:

$$\begin{aligned} \log[L(\lambda_{ij}, y_{ij})] &= K \sum_{j=1}^{n} \log\left(K \Big/ \exp\left(\eta_{ij}\right)\right) + \\ (K-1) \sum_{j=1}^{n} \log(y_{ij}) &- K \sum_{j=1}^{n} y_{ij} \Big/ \exp\left(\eta_{ij}\right) - n \log\left(K-1\right)! \end{aligned} \tag{6}$$

By derivation of Eq. (6) with respect to β, we obtain the score function in Eq. (7).

$$\frac{\partial \log[L(\lambda_{ij}, y_{ij})]}{\partial \hat{\beta}} = -K \sum_{j=1}^{n} x_{ij} + \sum_{j=1}^{n} y_{ij} x_{ij} \lambda_{ij} \tag{7}$$

For obtaining the estimations of parameters, we should put the computed score function equal to zero and multiply both sides to $\exp\left(\eta_{ij}\right) \Big/ K$.

So the Maximum Likelihood Estimator (MLE) of β is the result of the following score function:

$$X_i^T(\mathbf{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{8}$$

where $\mu_i = exp(\eta_i)$, $X_i = \begin{bmatrix} \mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1},\mathbf{1} \\ x_i \end{bmatrix}$, $\mathbf{0} = (0,...,0)^T$, $\boldsymbol{\mu}_i = (\mu_{i1},....,\mu_{in})$ and $\mathbf{Y}_i = (Y_{i1},....,Y_{in})$. By applying the iterative weighted least square estimation method which is described below, the MLE estimators of the regression parameters can be obtained. The estimation procedure of Gamma regression parameters are explained as follows:

Step 1: Set L=0. Consider $\hat{\beta}^{(0)}$ as an initial estimate of β that can be estimated by any method such as MLE.

Step 2: Compute $\eta_i^{(L)}$ by using Eq. (3). Then compute the vector of λ_i as follows:

$$\lambda_i^{(L)} = K \big/ \exp(\eta_i^{(L)}). \tag{9}$$

Note that Eq. (9) is derived from Log link function.

Step 3: Define a diagonal matrix of W by using Eq. (10).

$$W^{(L)} = diag\left(\frac{K}{\left(\lambda_i^{(L)}\right)^2}\right) \tag{10}$$

So W is a $n \times n$ diagonal matrix which its main diameter's elements are $\frac{K}{\left(\lambda_i^{(L)}\right)^2}$.

Step 4: Compute the adjusted dependent vector of $q_i^{(L)}$ by Eq. (11).

$$q_i^{(L)} = \eta_i^{(L)} + (W^{(L)})^{-1}(y_i - \mu_i) \tag{11}$$

Step 5: Update the estimation of β by Eq. (12)

$$\hat{\beta}_i^{(L+1)} = (X^T W^{(L)} X)^{-1} X^T W^{(L)} q_i^{(L)} \tag{12}$$

if $\hat{\beta}^{(L+1)} - \hat{\beta}^{(L)} < \varepsilon$, then $\hat{\beta}^{(L+1)}$ can be considered as the final estimation of β. Otherwise set L=L+1 and go to step 2. ε is a predetermined small number (for example 0.005).

3. T^2 Methods

As mentioned former, Yeh et al. (2009) proposed five T^2 based methods in order to monitor logistic profiles. We apply these methods for the case that the observations follow Gamma distribution. The methods will be described below:

3.1 T^2based on sample mean and covariance matrix (T_H^2)

In this method, the covariance matrix is estimated as follows:

$$S_H = \frac{1}{m-1}\sum_{i=1}^{m}\left(\hat{\beta}_i - \overline{\beta}\right)\left(\hat{\beta}_i - \overline{\beta}\right)', \tag{13}$$

where $\hat{\beta}_i$ is vector of estimated parameters for i^{th} profile and $\overline{\beta}$ is the average of estimated vectors among all m profiles. The T^2 statistic of this method can be computed as Eq. (14).

$$T_{H,i}^2 = (\hat{\beta}_i - \overline{\beta})' S_H^{-1} (\hat{\beta}_i - \overline{\beta}), \tag{14}$$

3.2 T^2 based on the sample average and intra-profile pooling (T_I^2)

The statistic of this method is shown in Eq. (15).

$$T_{I,i}^2 = (\hat{\beta}_i - \overline{\beta})^T S_I^{-1} (\hat{\beta}_i - \overline{\beta}), \tag{15}$$

where $S_I = \frac{1}{m}\sum_{i=1}^{m} \text{var}(\hat{\beta}_i) = \frac{1}{m}\sum_{i=1}^{m} (X' W_i X)^{-1}$

3.3 T^2 based on sample average and moving ranges (T_R^2)

In this method, the T^2 statistic that should be mapped on the control chart is shown in Eq. (16)

$$T_{R,i}^2 = (\hat{\beta}_i - \bar{\beta})^T S_R^{-1} (\hat{\beta}_i - \bar{\beta}), \tag{16}$$

where $S_R = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} (\hat{\beta}_{i+1} - \hat{\beta}_i)(\hat{\beta}_{i+1} - \hat{\beta}_i)^T$

3.4 T^2 based on Minimum Volume Ellipsoid (T_{MVE}^2)

For this method, the T^2 statistic is given in Eq. (17)

$$T_{E,i}^2 = (\hat{\beta}_i - \hat{\beta}_E)' S_E^{-1} (\hat{\beta}_i - \hat{\beta}_E), \tag{17}$$

where β_E and S_E are estimated by the MVE method.

3.5 T^2 based on minimum covariance determinant (T_{MCD}^2)

By applying Eq. (18), one can compute theT^2 statistic for this method.

$$T_{D,i}^2 = (\hat{\beta}_i - \hat{\beta}_D)' S_D^{-1} (\hat{\beta}_i - \hat{\beta}_D) \tag{18}$$

β_D and S_D are estimated by the MCD method.
Details about the algorithms of MVE and MCD are presented in Vargas (2003).

4. Performance Evaluation of the Methods

In this section, the performance of the methods in detecting step shifts and drifts is evaluated using simulation studies. The vector of explanatory variable is adopted from Yeh et al. (2009). So:

$$\mathrm{X} = [\mathrm{Log}(0.1), \mathrm{Log}(0.2), \ldots\ldots, \mathrm{Log}(0.9)]$$

β_0 and β_1 are assumed to be 3 and 2, respectively when the profiles are under statistical control. m and K are considered to be equal to 30. We consider two kinds of shifts including step shifts and drifts in this paper. So, we first explain the concepts of "step shifts" and "drifts".

4.1 Step shifts

In this type of shifts, the mean of the regression parameters is in statistical control up to a special profile. After that profile, the mean of the regression parameters changes to another value and remains in that level up to a time a corrective action is implemented. Hence, $\beta_{0,out} = 3 + 1 \times \sigma_1$ and $\beta_{1,out} = 2 + 3 \times \sigma_2$

σ_1 and σ_2 can be computed as follows (Yeh 2009):

$$\begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} = (X^T W X)^{-1} = \begin{pmatrix} 0.0413 & 0.0617 \\ 0.0617 & 0.1595 \end{pmatrix}$$

In this paper we have applied the step shifts on the second half of profiles (16th to 30th profiles).

4.2 Drifts

In this type of shift, the regression parameters are in statistical control and from a special point, the mean of the regression parameters changes with a linear trend. So the magnitude of shift increases by incensing in the sample number. For example (1,3) means that:

$$\beta_{0,out} = 3 + \left(\frac{i-1}{m-1}\right) \times 1 \times \sigma_1 \text{ and } \beta_{1,out} = 2 + \left(\frac{i-1}{m-1}\right) 3 \times \sigma_2$$

Drifts are implemented from the second profile in this paper.

Performance of the methods under various step shifts and drifts are shown in Tables 1 and 2, respectively. Note that we use power of the control chart criterion which determines the probability of detecting each shift in 10000 replications as the basis of our comparisons.

Table 1: Performance of the T^2 methods under step shift

	Methods				
	T_I^2	T_R^2	T_H^2	T_{MVE}^2	T_{MCD}^2
UCL / Shifts	40.5000	12.3944	10.5336	18.0500	58.0004
0,0	0.0574	0.0508	0.0536	0.0507	0.0523
1,1	0.0624	0.5358	0.0288	0.0283	0.0533
1,0	0.1282	0.1200	0.0490	0.0387	0.0540
1,2	0.0550	0.9694	0.0240	0.0370	0.0517
2,3	0.0560	0.9972	0.0220	0.0323	0.0563
2,0	0.3832	0.3316	0.0300	0.0353	0.0399
1,3	0.0772	0.9984	0.0228	0.0278	0.0382
2.5,0	0.5760	0.4784	0.0270	0.0397	0.0419
3,0.3	0.6862	0.4566	0.0290	0.0340	0.0536
3,0	0.7588	0.6107	0.0256	0.0343	0.0545

Table 2: Performance of the T^2 methods under drift

	Methods				
	T_I^2	T_R^2	T_H^2	T_{MVE}^2	T_{MCD}^2
UCL / Shifts	40.5000	12.3944	10.5336	18.0500	58.0004
0,0	0.0574	0.0508	0.0536	0.0507	0.0523
1,1	0.0570	0.3290	0.0422	0.0510	0.0529
1,0	0.0870	0.0822	0.0570	0.0547	0.0532
1,2	0.0526	0.9672	0.0264	0.0480	0.0519
2,3	0.0610	0.9996	0.0282	0.0530	0.0535
2,0	0.2128	0.1822	0.0492	0.0457	0.0499
1,3	0.0526	1.0000	0.0294	0.0557	0.0523
2.5,0	0.3054	0.2832	0.0420	0.0537	0.0517
3,0.3	0.3690	0.2378	0.0432	0.0383	0.0492
3,0	0.4188	0.3982	0.0428	0.0507	0.0511

Results show that T_I^2 performs better than the other T^2 based methods under individual step shift and drift in β_0. Under simultaneous shifts in $\beta_0$and β_1, the T_R^2 method outperforms the other methods.

5. Conclusions

There are many real case problems in which the response variable of the profiles follows Gamma distribution. For example when each observation describes the interval between the occurrence of two defects, the sum of k observations for each level of the explanatory variable follows Gamma distribution. In this paper, we extend and evaluate the performance of the five T^2 methods proposed by Yeh et al. (2009) for monitoring Gamma profiles. As a matter of fact, the procedure of parameter estimation for Gamma profiles is the other contribution of this paper.

Results showed that the T_I^2 method performs better than the other T^2 methods in the case of individual step shifts and drifts in β_0. However, T_R^2 performs better than the other T^2 methods when simultaneous shifts, either step shifts or drifts, occur in the regression parameters $\beta_0$and β_1.

References

Amiri A., Jensen W. A., Kazemzadeh R. B., A Case Study on Monitoring Polynomial Profiles in the Automotive Industry, *Quality and Reliability Engineering International*, vol. 26, pp. 509-520, 2010.

Amiri, A., Koosha, M., Azhdari, A., Profile monitoring for Poisson response, *Proceedings of IEEM conference, IEEE*, Singapore, 2011.

Chen, S., Nembhard H. B., A High Dimensional Control Chart for Profile Monitoring, *Quality and Reliability Engineering International*, vol. 27, pp. 451-464, 2011.

Eyvazian M, Noorossana R, Saghaei A., Amiri A., Phase II Monitoring of Multivariate Multiple Linear Regression Profiles, *Quality and Reliability Engineering International*, vol. 27, pp. 281-296, 2011.

Kang, L., Albin, S. L., On-Line Monitoring when the Process yields a Linear Profile, *Journal of Quality Technology*, vol. 32, pp. 418-426, 2000.

Kazemzadeh, R. B., Noorossana, R., Amiri, A., Phase I Monitoring of Polynomial Profiles, *Communications in Statistics: Theory and Methods*, vol. 37, pp. 1671-1686, 2008.

Kim, K., Mahmoud, M. A., Woodall, W. H., On the Monitoring of Linear Profiles, *Journal of Quality Technology*, vol. 35, pp. 317-328, 2003.

Mahmoud, M. A., Phase I Analysis of Multiple Linear Regression Profiles, *Communications in Statistics: Simulation and Computation*, vol. 37, pp. 2106-2130, 2008.

Mahmoud, M. A., Parker, P. A., Woodall, W. H., Hawkins, D. M., A Change Point Method for Linear Profile Data *Quality and Reliability Engineering International*, vol. 23, pp. 247-268, 2007.

Mahmoud, M. A., Woodall, W. H., Phase I Analysis of Linear Profiles with Calibration Applications, *Technometrics*, vol. 46, pp. 380-391, 2004.

Noorossana, R., Amiri, A., Vaghefi, S.A. and Roghanian, E., .Monitoring Quality Characteristics Using Linear Profile, *Proceedings of the 3rd International IndustrialEngineering Conference*, Tehran, Iran, pp. 246-255, 2004.

Noorossana, R., Eyvazian M., Amiri A., Mahmoud M.A., Statistical Monitoring of Multivariate Multiple Linear Regression Profiles in Phase I With Calibration Application, *Quality and Reliability Engineering International* vol. 26, pp. 291-303, 2010.

Saghaei, A., Mehrjoo, M., Amiri, A., A CUSUM-based Method for Monitoring Simple Linear Profiles, *International Journal of Advanced Manufacturing Technology*, vol. 45, pp. 1252-1260, 2009.

Vaghefi, A., Tajbakhsh, S. D., Noorossana, R., Phase II Monitoring of Nonlinear Profiles, *Communications in Statistics: Theory and Methods*, vol. 38, pp. 1834-1851, 2009.

Vargas J. A., Robust Estimation in Multivariate Control Charts for Individual Observation, *Journal of Quality Technology*, vol. 35, pp. 367-376, 2003.

Williams, J. D., Woodall, W. H., Birch, J. B., Statistical Monitoring of Nonlinear Product and Process Quality Profiles, *Quality and Reliability Engineering International*, vol. 23, pp. 925-941, 2007.

Yeh, A. B., Huwang, L., Li, Y. M., Profile Monitoring for a Binary Response, *IIE Transactions*, vol. 41, pp. 931-941, 2009.

Zou, C., Tsung, F., Wang, Z., Monitoring General Linear Profiles using Multivariate Exponentially Weighted Moving Average schemes, *Technometrics*, vol. 49, pp. 395-408, 2007.