POISSON MODEL IN STATISTICAL PROCESS CONTROL

Ângelo Márcio Oliveira Sant'Anna
Pontifical Catholic University of Parana
Imaculada Conceição Street, 1155, Prado Velho, Curitiba, PR
angelo.santanna@pucpr.br

Danilo Marcondes Filho
Federal University of Rio Grande do Sul
Bento Gonçalves Avenue, 9500, Agronomia, Porto Alegre, RS
marcondes.danilo@gmail.com

ABSTRACT

Control charts based on the Poisson distribution for monitoring series of counts typically arising in the quality control are presented. The use of the regression model is used by practitioners in industry when investigating one or more variables under the assumption of some probability distribution for the data. However, many processes have variables representing count data arise, in which case a Poisson distribution is more appropriate. The scheme is based on the residuals deviance for detecting any disturbance in the control variables. This paper presents the control charts for industrial processes that involve count variables. We illustrated the performance of Poisson model-based control charts to case study based on real process. Using Monte Carlo simulations run-length properties of the proposed schemes are investigated.

KEYWORDS. Statistical quality control, model-based control chart, Poisson distribution.

Main area (Statistics / OR Industry)

1. Introduction

Control charts require that the data monitored be independent and normally distributed around the constant mean of a reference model (MONTGOMERY, 2005). However, adjustments in one or more process control variables often result in alterations to the reference model, resulting in a different mean for each new adjustment. This demands specific control charts to monitor each adjustment in the control variables, thereby hindering the process of constructing traditional control charts.

Regression charts with Normal response are extensiveness found in the statistics and engineering literature, and it to use with Poisson response is growing in the literature. The procedure for monitoring a response variable as function of control variable was initially proposed by Mandel (1969) and was called "regression control chart." As an extension of Mandel's proposal, Hawkins (1993) then proposed the regression-adjustment chart in order to monitor processes with multiple control variables. This monitoring procedure requires that a suitable regression model be implemented by using multiple linear regressions, or by ordinary least squares (OLS) regression. Alternative models such as generalized linear models (GLM) can be applied in the case of non-normally distributed response variables. A GLM can be used to build models for any response variables from many different distributions.

Skinner *et al*. (2003) proposed a model-based control chart for processes with response variables following Poisson distributions. Jearkpaporn *et al*. (2003) proposed the use of GLM-based control charts for monitoring Gamma distributed response variables. Shu et al. (2008) reviewed the literature on regression control charts and their importance in process data, since practical applications arise from regression models.

This paper presents the control charts for industrial processes that involve count variables. A model-based procedure is used for deviation residuals, which is a likelihood ratio statistic when control variables are measurable. We illustrated the performance of Poisson model-based control charts through a case study based on a real process. Using Monte Carlo simulations run-length properties of the proposed schemes are investigated.

2. Generalized Linear Models

The generalized linear model (GLM) is a class of regression models appropriate for investigating the effect of control variables on a response variable in a non-normal distribution. The GLM was developed by Nelder and Wedderburn (1972) and these models are based on probability distributions with an unknown location parameter (θ), assuming that it belongs to the exponential family. This family includes the following distributions: Normal, Binomial, Poisson, Gamma, Exponential and others. The probability density function of the exponential family is most commonly seen, as in Eq. (1):

$$f(y;\theta,\phi) = exp[a(\phi)^{-1} - (y\theta - b\theta) + c(y\phi)] \tag{1}$$

where $a(\cdot)b(\cdot)c(\cdot)$ are known functions, θ is the location parameter and $\phi > 0$ is the dispersion parameter of the probability distribution.

GLM models are structured around three components: (i) random component, which identifies the probability distribution of the response variable in which y belongs to the exponential family, (ii) a systematic component, which specifies the structure for the control variables, which is used as a linear predictor η, and (iii) Link function - describes the functional relationship between the systematic component and value expected of random component (the mean μ of the response variable y).

The systematic component that makes up the regression model is the structure of control variables as a linear sum η and the relationship between variables in a GLM can be expressed by a known function g(.), called the link function (McCullagh & Nelder, 1989), which describes the functional relationship (connection) between the mean μ and the linear predictor (η), making the mean response variable as in Eq. (2),

$$g(\mu) = \eta = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k \qquad (2)$$

where β′s are unknown coefficients and k's are control variables. Sant'Anna and Caten (2010) suggest that the application of the GLM for modeling data process allows for better precision of estimates, nonlinear relationship between the variables and predicting their behavior.

2.1. The Poisson distribution

The model-based control chart is used in monitoring processes where the response variables undergo frequent change due to adjustments in one or more control variables. This model is based on the likelihood ratio statistic for a data set where the mean is a function of a linear combination of control variables. In this case the Poisson distribution is considered here, as it is known to model count of defects per items well.

Let y be a discrete random variable that measures the number of defects (y) in a set independent data by determined time, y = 0, 1, 2,... . The probability of y ($P\{y_i = y\}$) is defined by the Poisson distribution,

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \quad ; \quad \lambda > 0 \qquad (3)$$

The Poisson probability density function belonging exponential family is given Eq.4:

$$f(y;\lambda) = exp[ylog\lambda - \lambda + logy!] \qquad (4)$$

where λ is the location parameter (mean) of the Poisson distribution, y the response variable (Johnson *et al.* 2005). Consider S response variables, y_1, y_2,..., y_s, distributed according to a Poisson distribution and mean λ = E(y). And these variables are function of a set measurable control variables x_1, x_2, ..., x_k with unknown coefficients β_1, β_2,..., β_k. If a logarithmic function is adopted as link function, we have that,

$$y \sim Poisson[exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)] \qquad (5)$$

The coefficients β′s of the GLM are estimated using the likelihood ratio statistic, where the values are obtained when maximizes the log-likelihood function. The deviance residuals are obtained from the likelihood ratio statistics and used in the model-based control charts for monitoring and controlling the process, in order to detect the change in mean or variability. Thus, we estimate the uncorrelated deviance residuals (r_d) for the Poisson regression model in Eq.7,

$$r_d^P = sign(y - \hat{\lambda})\{2[ylog(y/\hat{\lambda}) - y + \hat{\lambda}]\}^{1/2} \qquad (6)$$

where $\hat{\lambda}$ is the estimated mean from the model being studied. McCullagh & Nelder (1989) show that the deviance residuals r_d^P are asymptotically normally distributed with zero mean and unit variance.

3. The Control Chart Approach

GLM-based models predict the response variables based on the control variables and the use of deviance residuals estimated by maximum likelihood to monitor the response variability. Response variables in industrial applications usually follow a non-normal distribution with asymmetric shape, which compromises the precision of traditional model-based control charts. Here we describe the GLM-based control chart for monitoring Poisson response variables.

Deviance residuals are plotted in univariate Shewhart charts and the control limits are defined from residuals mean ($\bar{r}_d^P$) and variance $s^2(\bar{r}_d^P)$ generated by historical data, as:

$$CL = \bar{r}_d^P \pm w\sqrt{s^2(\bar{r}_d)} \tag{7}$$

where the constant w defines the width for control limits based on the average run length until a false alarm is verified (ARL_0). Epprecht *et al*. (2005) suggest that the frequency of false alarms may be used for faster detection of control losses without raising the cost of sampling. Skinner *et al*. (2003) state that deviance residuals capture modifications on the response variables when adjustments on the control variables are performed, enabling their use in charts for process monitoring.

3.1. Summary for application

The steps to implement model-based control charts are summarized here:

a) Obtain the historical data to industrial process;
b) Fit the GLM to the response and control variables from in-control historical data (Eq. 5).
c) Obtain the deviance residuals (r_d^P) for in-control process data from Eq. 6.
d) Obtain the control limits to control charts from Eq. 7.
e) Monitor the deviance residuals from future samples using the constructed control charts.

4. Numerical Application

The implementation of the control charts for monitoring is illustrated through a case study from an industrial process. The study was carried out in a plastic injection company that manufactures and supplies plastic components to White Family, where the effect of control variables on the defects in industrial “Headwaters” production was assessed. The Company has three plants located in the states of Paraná and Santa Catarina, and it has over 500 employees and is the market of plastic injection components for 10 years. The customers of this product are the main wallet plant.

For the case study on plastic injection, the number of defects follows a Poisson distribution with a mean (standard deviation) of 2.99 (±1.23) flaws per area. The control variables of the process were: injection speed (x_1) and injection pressure (x_2).

We assumed that the variables x_1 and x_2 are independent and identically distributed with Uniform distribution (0,3), and λ_i represents the mean given an observation vector (x_{1i}, x_{2i}), where for each i =1,...,n, the value y_i is simulated using the Poisson distribution parameter λ_i (Eq. 3). This configuration allows representative change to be generated in Y from shifts in the control variables. The goal was to monitor the performance of a response (count) variable that varies according to two continuous control variables.

4.1. Empirical Study

Step 1:

From the 1000 historical data, we estimated the regression model to represent the real industrial process of plastic injection, as the Eq. 8. The coefficients for the Poisson model were estimated by maximum likelihood estimation based on Wald's statistical test at 5% and have log link function. The coefficients are statistically significant to model (p<0.01) and goodness of fit had better fit (AIC = 6875 and Deviance 1058.4 (997df)).

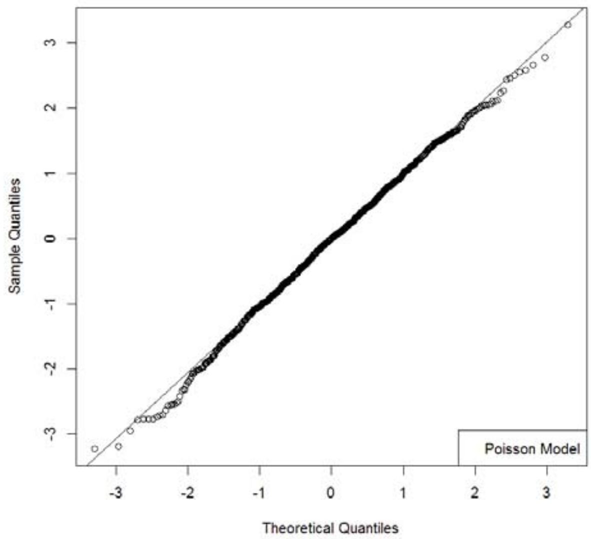
$$\lambda_i \sim Poisson[exp(0.994875 + 0.996597x_{1i} + 1.007360x_{2i})] \quad (8)$$

Step 2:

The deviance residuals were calculated. Figure 1 shows the adjustment proximity of their empirical distribution in relation to a QQ Normal probability plot. One will note that the Poisson residuals (r_d^P) distribution closely approximated the Normal curve as the theory predicts.

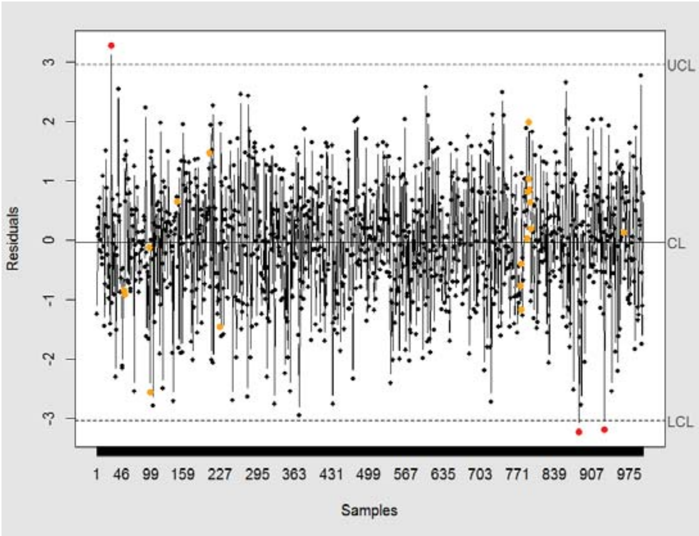
Step 3:

We obtain symmetric control limits for the Poisson model [Eq. 7]. Figure 2 shows the control chart obtained from 1000 in-control process data points for monitoring future samples. One can observe that the relative frequency of outliers in control chart [3/1000 in (a) and 4/1000 in (b)] is close to the theoretical probability of false alarms used in adjusting the chart (α=0.0027).



(a)

Figure 1. QQ Normal plot for deviance residuals from Poisson model



(a)

Figure 2. Poisson model-based control chart to in-control process

5. Performance Analysis

The performance analysis was development the Monte Carlo simulations to calculate the run-length properties of the proposed control chart using computational statistical program **R**® available in open source. We simulated 5000 replications for each trial with n=1000 observations to analyze Poisson model-based control chart performance in two scenarios: in-control and out-of-control processes. Cescato and Lemgruber (2011) assert that the simulation method is reliable, since it presents results very similar to those provided by the regression model.

For the in-control process, we evaluate the estimate average run-length until a false alarm (ARL_0), and for out-of-control, we evaluate the average run-length to detect shifts imposed by coeficients model (ARL_1). The shift sizes were in the β_0 ($\varphi = 1\sigma_y$ and $2\sigma_y$) and in the β_1 (Δ=0.1 and 0.3). The Table 2 shows the ARL performance for the models for in-control and out-of-control simulated process. Notice in the first row that the ARL_0 performance from the Poisson Model is close to the nominal value 370 (ARL_0=370 for α=0.0027).

One can notice the performance in the Poisson model for identifying the simulated disturbances, independent of the shift size imposed in β_0 and β_1. As expected, the overall results indicate that the Poisson model is good accuracy at detecting changes in count response.

Table 1. ARL (and ARL standard errors) for Poisson model

False Alarm	Shift Size	Shift Type	Poisson Model
ARL_0	–	–	344.97 (0.0548)
ARL_1	$1\sigma_y$	β_0	6.79 (0.000088)
ARL_1	0.1	β_1	4.34 (0.000031)
ARL_1	$2\sigma_y$	β_0	2.21 (0.000011)
ARL_1	0.3	β_1	1.76 (0.0000045)

6. Conclusions Remarks

A control chart using regression model is a quality tool for monitoring response variables that vary as a result of modifications in the control variables. The tool consists in building a regression model and estimating the control limits using the deviance residuals from the model.

We applied control charts to monitor an industrial process with historical count data from plastic injection process to evaluate the performance of Poisson model. A performance study was conducted to evaluate the average run-length of control chart in two different scenarios: in-control (ARL_0) and out-of-control (ARL_1). In this study, different shift sizes in coefficients β_0 and β_1 were imposed to demonstrate the improved performance of the Poisson model-based control chart.

The relationship between the process variables and the regression model estimated should be adequate for analyzing the industrial process and monitoring the disturbance on the future count data.

Natural extensions to the research presented here would include the development of an economic design for this control charts.

7. Acknowledgment

The authors are grateful for the support by Pontifical University Catholic of Parana (PUCPR/Brazil) and the opportunity to work with the Plastic Company.

Referências

Asadzadeh, S,; Aghaie, A. & Shahriari, H. (2009). Monitoring Dependent Process Steps using Robust Cause-selecting Control Charts, *Quality and Reliability Engineering International*, 25, 851–874.

Cescato, C.D., Lemgruber, E.F., (2011). Valuation of American interest rate options by the least-squares monte carlo method. *Pesquisa Operacional*. 31. 521-541.

Epprecht, E.K., Costa, A.F.B., Mendes, F.C.T., (2003). Gráficos adaptativos de controle por atributos e seu projeto na prática, *Pesquisa Operacional*, 25, 113–134.

Hawkins, D.M. (1993). Regression adjustment for variables in multivariate quality control, *Journal of Quality Technology*, 25, 170–182.

Jearkpaporn, D.; Montgomery, D.C.; Runger, G.C. & Borror, C.M. (2003) Process monitoring for correlated Gamma-distributed data using generalized linear model-based control charts, *Quality and Reliability Engineering International*, 19, 477–491.

Johnson, N.L.; Kemp, A.W.; Kotz, S. (2005). Univariate Discrete Distributions. 3ª ed., New York: John Wiley & Sons, 646p.

Mandel, B.J. (1969) The regression control charts, *Journal of Quality Technology*, 1, 1–9.

McCullagh, P. & Nelder, J.A. (1989). *Generalized Linear Models*. 2ª ed., Chapman & Hall, London.

Montgomery, D.C. (2005). *Introduction to Statistical Quality Control*, 5ª ed., John Wiley & Sons, New York.

Nelder, J.A.; Wedderburn, R.W.M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society A*, 135, 370-384.

Sant'Anna, A.O.M., Caten, C.S. 2010. Modelagem da fração de não-conformes em processos industriais. *Pesquisa Operacional*, 30, 53-72.

Shu, L.; Tsui, K.; Tsung, F. (2008) A review of regression control charts, *Encyclopedia of Statistics in Quality and Reliability*, v. 260, 1-9, 2008.

Skinner, K.R.; Montgomery, D.C. & Runger, G.C. (2003) Process monitoring for multiple count data using generalized linear model-based control charts, *International Journal of Production Research*, 41, 1167–1180.