Monitoring a Poisson Vector with a Linear Combination of Its Elements

Eugenio Epprecht[1], Francisco Aparisi[2], Sandra García-Bustos[3]
[1]Pontificia Universidade Católica do Rio de Janeiro/Depto de Eng. Industrial, Rua Marquês de São Vicente, 225, 22451-900, Rio de Janeiro, Brazil
[2]Universidad Politécnica de Valencia/Depto de Estadística, Camino de Vera, 46022 Valencia, Spain
[3]Escuela Superior Politécnica del Litoral/Facultad de Ciencias Naturales y Matemáticas, Km 30.5 Via Perimetral, Guayaquil, Ecuador

Abstract
We propose and analyze a new chart for the multivariate process control of Poisson variables: the Linear Combination of Poissons (LCP) chart. The implementation parameters (coefficients of the linear combination and control limit) of this chart can be obtained running (in Windows) user-friendly software developed by the authors. The program employs genetic algorithms to minimize the out-of-control ARL for a given shift, under the constraint of a desired in-control ARL. The new chart shows very good performance when compared with its competitors. An EWMA version of it has also been developed.

Keywords: multivariate Poisson distribution, statistical process control, SPC, optimization, optimal design, EWMA, genetic algorithms, performance, ARL

Introduction

There is a vast body of literature on multivariate statistical process control (MSPC) by variables; on the other hand, works on MSPC by attributes are scarce. We may cite Patel (1973), Lu et al. (1998), Skinner et al. (2003), Chiu and Kuo (2008), Ho and Costa (2009), Laungrungrong et al. (2011), Dourodyan and Amiri (2013), Aparisi et al. (2014). Although this list is not intended to be exhaustive, it is representative, and gives an idea of the number of papers on the subject, which is some orders of magnitude smaller than the number of papers on MSPC by variables.

We are interested in the monitoring of multivariate Poisson processes. Specifically, we consider processes that can be well represented by the model proposed by Holgate (1964). He considered that each one of the p (observable) Poisson variables X_i (i = 1, 2, ..., p) is the sum of two non-observable and independent Poisson variables: a common component Y_0 and an individual component Y_i (i = 1, 2, ..., p). In addition, all Y_i variables are independent of each other. As a result, the covariance between any pair of observable variables X_i and X_j is the variance of Y_0.

In formal notation, denoting by λ_i the mean of Y_i (and by λ_0 the mean of Y_0), we have:

e-mails: [1] eke@puc-rio.br [2] faparisi@eio.upv.es [3] slgarcia@espol.edu.ec

$$X_i = Y_0 + Y_i \qquad i = 1, 2, ..., p \tag{1}$$

$$\mathrm{E}(X_i) = \lambda_0 + \lambda_i \qquad i = 1, 2, ..., p \tag{2}$$

$$Cov(X_i, X_j) = \lambda_0 \qquad i = 1, 2, ..., p, \quad i \neq j \tag{3}$$

$$\rho(X_i, X_j) = \frac{\lambda_0}{\sqrt{(\lambda_0 + \lambda_i)(\lambda_0 + \lambda_j)}} \qquad i = 1, 2, ..., p, \quad i \neq j \tag{4}$$

This is not the only possible way in which the elements of a Poisson vector can be correlated. For instance, negative correlations cannot be represented within the framework of this model. It is, however, the model that has been assumed in virtually any published work on the monitoring of a multivariate Poisson process, and can adequately represent several real situations. The consideration of other models (and the development of appropriate monitoring methods for them) remains an open issue for research.

Based on Holgate's model, Chiu and Kuo (2008) proposed to chart the sum of the values of the different X_i variables, in what they called the MP chart. They also compared the performance of this chart with the Multiple Scheme (a set of *c* charts, one for each variable). Ho and Costa (2009) considered other two charts, applied to the case of two Poisson variables following Holgate's model: the DX chart, on the difference between them, and the MX chart, on the maximum of their values in the sample. They compared the performances of these two charts, the MP chart and the Multiple Scheme. Aparisi et al. (2013) presented a program (running in Windows®, user-friendly and available upon request) that optimizes (using Genetic Algorithms) and compares the performances of these four schemes, and is intended to assist the user in the choice of the best monitoring scheme for his/her particular problem. Laungrungrong et al. (2011) proposed and evaluated a version of the MEWMA chart of Lowry et al. (1992) based on Holgate's multivariate Poisson model rather than on the normal distribution. They found that although the out-of-control performance of the two schemes is similar, the in-control performance of the new version is superior. More recently, the same authors developed a one-sided version of this scheme (Laungrungrong et al., 2014).

We do not detail the works in other references given (Patel, 1973; Lu et al., 1998; Skinner et al., 2003; Dourodyan and Amiri, 2013) because they considered different types of variables/processes, e.g. binomial variables, independent Poisson variables or a mix of discrete and continuous variables.

This paper briefly presents an alternative statistic for multivariate Poisson monitoring we recently proposed, namely the linear combination of the Poisson variables (LCP), as well as two monitoring schemes based on it: a Shewhart-type chart (Epprecht et al., 2013) and an EWMA chart (García-Bustos et al., 2014), and summarizes the findings of the performance analyses carried out. The idea for the statistic came from the observation that the sum of variables in the MP chart and the difference between the variables in the DF chart are particular cases of linear combinations; the linear combination, being more general (and also applicable to more than two variables, in contrast with the DF chart), should be more flexible and exhibit better performance, which the performance analysis indeed confirmed.

The remainder of this paper presents the LCP statistic, describes the charts, the mathematical models for computing the ARLs, and two programs we developed for optimizing the design of the charts. Then follows a synthetic presentation of the results of comparative performance analyses, and the general conclusions. For details, the reader is referred to our two papers mentioned in the preceding paragraph.

The LCP chart

The LCP chart (Epprecht et al., 2013) is a two-sided Shewhart-type control chart on which, at each sample, the statistic

$$LCP = \sum_{i=1}^{p} a_i X_i \qquad (2)$$

is calculated and plotted. A point plotting outside the control limits is an alarm.

The values of the coefficients a_i and of the control limits UCL and LCL are design parameters, to be determined so as to give the desired in-control ARL and to minimize the out-of-control ARL for a specified shift. Also, for practical reasons (to limit the search region for the optimization algorithm) the values a_i are constrained to lie in the interval $[-1,\ 1]$, without loss of flexibility for the design since the control limits are determined accordingly. It is only a matter of scale.

A number of comments are in order: (i) Even when the purpose is to detect changes of the Poisson rates in one direction (e.g. increases), the chart may need a pair of limits, if some of the a_i's are negative and others positive. (ii) In contrast with most charts for attributes (in which, due to the discrete nature of the charting statistic, the in-control ARL can only take discrete values and it is not generally possible to match a desired ARL_0 value), the chart can be designed to match any ARL_0 required by an appropriate choice of the coefficients and control limits. This is an advantage of the proposed chart. (iii) There is no closed-form expression for determining the coefficients of the linear combination and control limits for the chart; these parameters have to be determined by a search algorithm that minimizes the ARL for a given shift in the means of the *Y* variables subject to a constraint in the in-control ARL. (iv) Even if the *Y* variables are not directly observable, estimates of their in-control means can be obtained as a function of estimates of the means and covariance of the *X* variables, using the relations expressed in Equations (2) and (3).

The EWMA LCP chart

The EWMA LCP chart (García-Bustos et al., 2014) uses as statistic an exponentially weighted moving average of the successive LCP values, in the usual way:

$$EWMALCP_t = rLCP_t + (1-r)(EWMALCP_{t-1}) \qquad \text{for } t = 1, 2, \ldots \qquad (6)$$

where LCP_t is the value, in the *t*-th sample, of the *LCP* statistic given in (2), r is the smoothing constant, ($0 < r \leq 1$) and the initial value $EWMALCP_0$, is the in-control expected value of *LCP*, given by

$$EWMACP_0 = \sum_{i=1}^{p} a_i E(X_i) \qquad (7)$$

In practice, if the expected values $E(X_i)$ are unknown, they may be directly estimated from historical in-control data by their averages $\overline{X_1}$, $\overline{X_2}$, ..., $\overline{X_p}$.

The chart has two control limits, *UCL* and *LCL*. These limits are not necessarily symmetric with respect to the expected value of *LCP*, given by (7), when the chart is optimized against a given shift. (This is also true of the LCP chart).

ARL computation

Under the assumption of serially independent data, the ARL of the LCP chart is the reciprocal of the signal probability. It is easier to calculate the probability of the complementary event, namely, that LCP lies between the control limits. This is the sum of the joint probabilities of all values of the tuple (or vector) $[x_1, x_2, ..., x_p]$ that yield a value of *LCP* between *LCL* and *UCL*. Note that since $X_i = Y_0 + Y_i$, *LCP* can be alternatively expressed as a linear combination of the Y_i's, with the advantage that because the Y's are independent of each other, the joint probability of any particular $[y_0, y_1, y_2, ..., y_p]$ point is simply the product of the univariate Poisson probabilities. The cumbersome analytical expression is omitted here (the interested reader is referred to Epprecht et al., 2013), but its calculation is algorithmically quite simple, and can be made by p+1 nested loops, one for each Y_i variable (including Y_0). The minimum and maximum values of each variable in its loop are not fixed, but are defined at each time as a function of the particular values assumed in that iteration by the variables in the loops external to it, so that *LCP* remain between *LCL* and *UCL*.

The computation of the ARL of the EWMA LCP chart requires a Markov chain model, which is however similar to the ones used in the literature for several other EWMA charts, since the early work of Lucas and Saccucci (1990). The basic idea is to discretize the continuous interval of variation of the EWMA statistic between the control limits by dividing it in a large number of subintervals, each one corresponding to a transient state. Then the one-step transition probability from state j to state k (that is, the probability that the EWMA statistic in sample t falls in subinterval k given that the EWMA statistic in sample t–1 was in subinterval j) is the probability that the sample statistic that will be smoothed (in our case, the LCP_t statistic) falls between a minimum and a maximum value that are determined by a trivial manipulation of the smoothing equation (in our case, Eq. (6)), under the simplifying assumption that the value of the EWMA statistic in sample t–1 coincided with the midpoint of its respective subinterval j. In the case of the EWMA LCP chart, the matrix equations are exactly the same as in these previous works. The only difference lies in the calculation of the transition probabilities: the probabilities that LCP_t lie between a minimum and a maximum value determined from Equation (6) are calculated using the same algorithm used in the case of the (Shewhart-type) LCL chart to compute the probability that *LCL* falls in a given interval.

Design and optimization

There is no closed-form expression for the design of the LCP or the EWMA LCP chart either. Given a specified ARL_0 value, there are multiple solutions (values of the coefficients of the linear combination, control limits and, in the case of the EWMA LCP chart, smoothing constant r) that result in the given ARL_0. A good design has to be obtained by search. On the other hand, this has the advantage of enabling the optimization of the charts' out-of-control performance.

We developed user-friendly software running in Windows, which performs the search (using genetic algorithms) to solve the problem of minimizing the ARL for a given shift (specified by the user) in the vector of means of the Y_i's, subject to the constraint that the in-control ARL should not be smaller than a given ARL_0 value. As already mentioned, the LCP and EWMA LCP charts can precisely match the required ARL_0 value, so the solution is on the contraint. A specific program was developed for each chart. In the case of the EWMA LCP chart, the ARL minimized is the steady-state ARL (the zero-state ARL is close to it, and given in the program output). The maximum number of observable variables (X_i's) allowed is three.

The data to be input by the user is then the ARL_0, the in-control means of the Y variables and the shift vector. Note that, even if the Y variables cannot be directly observed, their means can be estimated from historical data using the relations in (2) and (3).

Figure 1 shows the user interface of the program for optimizing the LCP chart. The interface of the program for optimizing the EWMA LCP chart is similar. Detailed descriptions and examples of use of the programs are given in Epprecht et al. (2013) and García-Bustos et al. (2014).

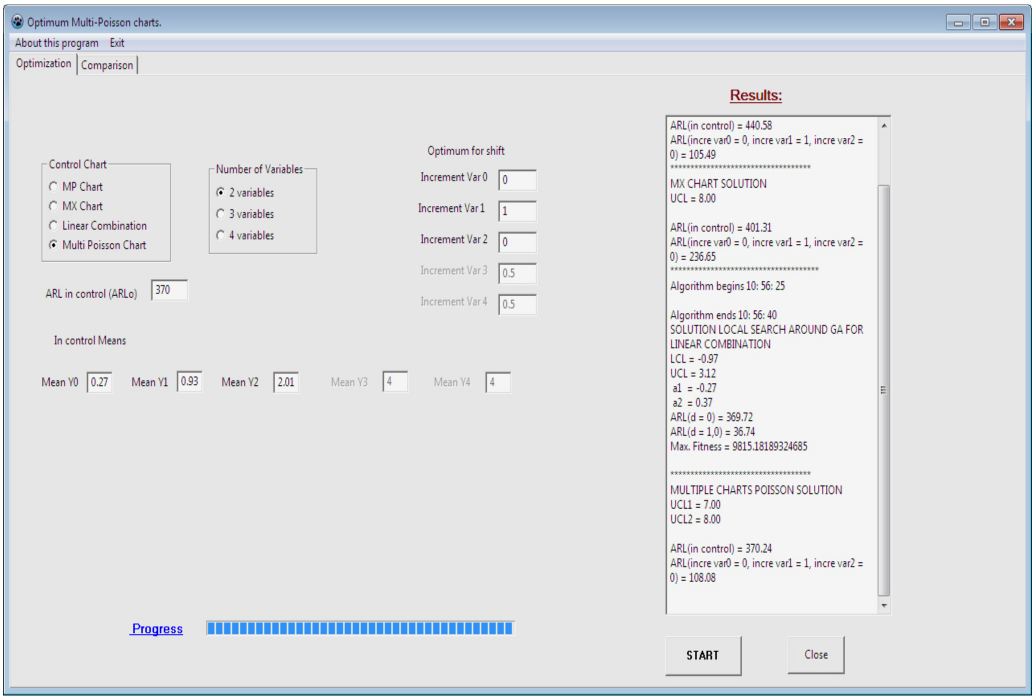

Figure 1: Program for optimizing the LCP chart – user interface

Performance analysis and comparison

The results of performance and sensitivity analyses of the LCP and EWMA LCP charts are presented in detail in the papers cited (Epprecht et al., 2013; García-Bustos et al., 2014). Numerical results for a comprehensive spectrum of cases could not be presented without the use of many tables, so for reasons of space and of keeping this exposition simple, we refer the reader interested in that to those papers; and here we only summarize the general results.

As predictable, the charts are not directionally invariant. In the case of the LCP chart, however, the optimal design depends only on the direction of the shift, and not on its magnitude. That is, if the LCP chart is optimized for a pair of shifts along the same direction but with different magnitudes, the results will be equivalent. By "equivalent" we mean that the coefficients and control limits of the two solutions will be proportional; in other words, they will differ only by the scale. This because the genetic algorithm may generate solutions in different scales. But for any given shift, the ARLs of the two charts will coincide.

In the case of the EWMA chart, however, the existence of the smoothing constant provides an additional flexibility, in that this constant and the control limits can be tuned together in order to optimize the performance for the magnitude of the relevant shift without compromising the in-control performance. So it the size of the shift for which the chart is optimized affects its design and performance.

This difference between the LCP and the EWMA LCP charts is analogous to the one between univariate Shewhart and EWMA charts, in which case the only component of the shift is its size. Given a desired ARL_0 (and sample size), the design of a Shewhart chart is fixed (the only parameter to be chosen are the control limits) and cannot be optimized in terms of ARL (unless the sample size and sampling interval are treated as decision variables); in contrast, EWMA charts can be optimized because there is an infinity of combinations of values of smoothing constant and control limits that yield the desired ARL_0. Now, in the case of the LCP and EWMA LCP charts, the shift has a new element, its direction, which allows optimization of both charts with respect to it, but regarding the shift magnitude the situation is analogous to the one of univariate charts.

Further analysis of the performance of the EWMA LCP chart has shown that (similar to univariate EWMA charts) its performance for large shifts is never poor. So if the size of the shift is not predictable, and protection is needed against different magnitudes of shifts, a good design strategy is to optimize it for the smallest relevant shift.

Regarding the comparison of performance with the competing monitoring schemes, the program for designing the LCP chart also gives the designs and ARLs of the MP chart, MX chart and multiple univariate Poisson charts scheme, for comparison and choice of the best scheme for the particular shift the user is concerned with. In addition, a new window shows the ARL curves of the various schemes as a function of the shift in the mean of one of the variables when the shifts in the means of the remaining variables are kept fixed, and also a plot of the regions where each scheme is the best one, considering the shifts in the means of two of the variables when the shifts in the means of the remaining ones are fixed. Figure 2 shows this window with an example of output.

The dimensionality of the problem renders difficult to give a comprehensive picture of the ARL surface of the charts. The results can however be summarized saying that, from a large number of different cases analyzed, the LCP chart outperformed its competitors (the MP chart, the MX chart and the multiple univariate Poisson charts) in the large majority of cases, and almost always quite significantly (giving in some cases ARL three times smaller); and that in the few cases where it was not the best scheme, the differences in ARL were small.

As to the EWMA LCP chart, it provides a substantial gain in performance with respect to the LCP chart. The reduction in the out-of-control ARLs range from 15% to even more than 70%, depending on the case. The few cases where the LCP chart outperforms the

EWMA LCP chart are cases of very large shifts, in which both charts have already quite small ARLs (a little larger than 1.0). In addition, the EWMA LCP chart performance is fairly robust regarding the shifts in the individual components of the mean vector of the Y_i variables, although somewhat more sensitive to the value considered for the shift in the mean of Y_0. In any case, though, it is much more efficient than the Shewhart-type LCP chart.

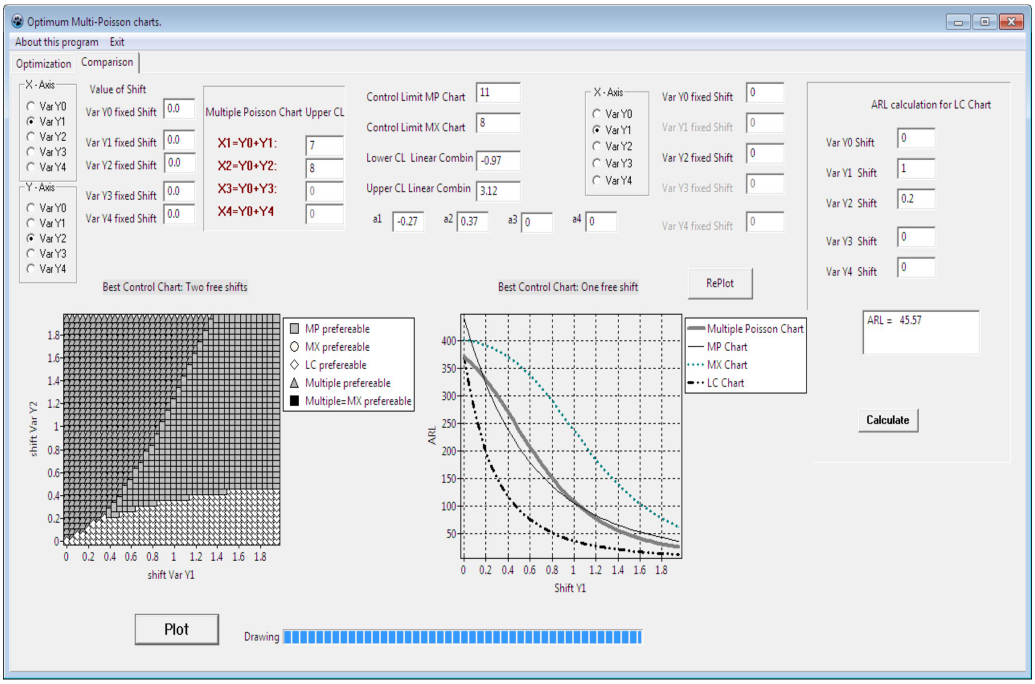

Figure 2: Comparison output window of the program for optimizing the LCP chart

Conclusions

The LCP statistic has shown its superiority over the alternative statistics in monitoring multivariate Poisson processes. Namely, the LCP chart has shown better out-of-control ARL performance than the MP, MX and multiple univariate Poisson charts. The ARL values of the LCP chart are sometimes one third of the ones of its best competitor. The cases where the LCP chart has not shown the best performance were very rare and its ARL was quite close to the one of the best scheme. In addition, with the LCP statistic, it is possible to match quite precisely the in-control ARL specified by the user of the chart — something that is generally not possible with control charts by attributes.

The EWMA version of the LCP chart has still better performance, with lower out-of-control ARLs in virtually any case. Only for very large shifts it is outperformed its Shewhart-type version. Note however that for such shifts the ARLs of both charts are already quite small, and close to each other. So, the EWMA LCP chart (optimized for the smaller shift that is deemed relevant) becomes an interesting *omnibus* control schema for shifts of different magnitudes.

The programs developed and made available by the authors make it easy to optimize the chart(s) for the desired shift, thus contributing to their practical applicability.

Acknowledgements

The first and second authors have been supported by the CNPq (the Brazilian Council for Scientific and Technological Development), projects numbers 454909/2013-6 and 307453/2011-1, respectively, and the third author has been supported by SENESCYT-Ecuador (National Secretary of Higher Education, Science, Technology and Innovation of Equator).

References

Aparisi, F., García-Bustos, S. and Epprecht, E. K. (2013). Optimum Multiple and Multivariate Poisson Statistical Control Charts. *Quality and Reliability Engineering International*, 30(2), pp. 221-234.

Chiu, J., and Kuo, T. (2008). Attribute Control Chart for Multivariate Poisson Distribution. *Communications in Statistics-Theory and Methods*. 37: 146-158.

Epprecht, E. K., García-Bustos, S., and Aparisi, F. (2014). Optimum Linear Combination of Poisson Variables for Multivariate Statistical Process Control. *Computers & Operations Research*, 40 (3021-3032).

Dourodyan, M. H. and Amiri, A. (2013). Monitoring multivariate-attribute processes based on transformation techniques. *International Journal of Advanced Manufacturing Technology* 69:2161-2172.

García-Bustos, S., Aparisi, F., and Epprecht, E. K. (2014). Optimum EWMA of Linear Combination of Poisson Variables for Multivariate Statistical Process Control. Submitted to the *International Journal of Production Research*.

Ho, L. L., Costa, A.F.B. (2009). Control charts for individual observations of a bivariate Poisson process. *International Journal of Advanced Manufacturing Technology* 43:744-755.

the program for optimizing **Holgate, P. (1964).** Estimation for the bivariate Poisson distribution. *Biometrika* 51: 241-245.

Laungrungrong, B., Borror, C. M., and Montgomery, D. C. (2011). EWMA Control Charts for Multivariate Poisson-distributed data. *International Journal of Quality Engineering and Technology*. 2(3), 185-211.

Laungrungrong, B., Borror, C. M., and Montgomery, D. C., (2014). A one-sided MEWMA control chart for Poisson-distributed data. *International Journal of Data Analysis Techniques and Strategies*, 6(1), p. 15-42. http://EconPapers.repec.org/RePEc:ids:injdan:v:6:y:2014:i:1:p:15-42.

Lowry, C. A., Woodall, W. H., Champ, C. W. and Ridgon, S. E. (1992). A multivariate exponential weighted moving average control chart. *Technometrics*, 34(1): 46-53.

Lu, XS., Xie, M., Goh, TN., & Lai, CD. (1998). Control chart for multivariate attribute processes. *International Journal of Production Research*. 36(12), 3477-3489.

Lucas, J. M., and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1), 1-12.

Patel, H. I. (1973). Quality control methods for multivariate binomial and Poisson distributions. *Technometrics* 15, 103-112.

Skinner, K. R., Montgomery, D. C., and 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.