T^2 Control Charts with Variable Dimension

Aparisi, Francisco
Departamento de Estadística e I. O. Aplicadas y Calidad. Universidad Politécnica de Valencia.
Camino de Vera s/n
46022 Valencia, Spain
faparisi@eio.upv.es

Epprecht, Eugenio K.
Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, Brazil
eke@puc-rio.br

1. Introduction

Multivariate statistical process control is a mature tool used frequently in industry to monitor the quality of a process. Let *p* denote the number of variables to be controlled simultaneously. In order to control these *p* variables two main approaches can be employed: 1.- A univariate control chart for each variable (multiple scheme). 2.- A single multivariate control chart.

The first option is discussed in Aparisi *et al*. (2010) for the optimum use of several $\overline{X}$ control charts; Epprecht *et al*. (2011) discuss the use of optimized multiple EWMA control charts. The multivariate SPC has been widely discussed in the bibliography, being the T^2 and MEWMA control charts the most employed alternatives (Montgomery, (2008)).

In this paper, we consider the use of the T^2 control chart when the number of variables to monitor is variable. This approach is useful when there is a set of p_1 variables that are easy to monitor or whose measurement is cheap, against a set of p_2 variables, $p = p_1 + p_2$, that are difficult and/or expensive to monitor. However, the information that these p_2 variables provide is important to detect quickly the process quality shift. Therefore, there are cases when controlling the whole set of *p* variables may be difficult or expensive, but controlling sometimes p_1 variables, and only when the process seems to have a problem controlling the full set of *p* variables, may be a cheaper option on average and very efficient, as we are going to show in this work.

As an example, we may consider a process that produces an expensive electronic component and has three (correlated) quality variables to be monitored. Two of these are voltages, which are cheap and easy to measure. In addition, there is third variable, whose measurement is destructive. For example, the voltage that will burn a part of component. In other cases, the additional variables could need, for example, a laboratory analysis that may be difficult, slow, expensive, etc. However, as it shown later, these additional variables may help to detect the out-of-control state of the process very efficiently.

2. New VDT^2 and DDT^2 control charts

The standard T^2 control chart procedure involves periodically taking a sample of size *n*, and next, calculating the sample average vector $\vec{\overline{X}}$ for the *p* variables to be monitored. The T^2 statistic is then computed as $T_i^2 = n(\vec{\overline{X}}_i - \mu_0)' \Sigma^{-1} (\vec{\overline{X}}_i - \mu_0)$ where $\vec{\mu}_0' = (\mu_{0,1}, \mu_{0,2}, \dots, \mu_{0,p})$ is the mean vector and Σ is the covariance matrix, both when the process is in control. The charting statistic T_i^2 is called Hotelling´s T^2 statistic and is distributed as a chi-square variate with *p* degrees of freedom. When the process is in control ($\vec{\mu}_i = \vec{\mu}_0$), there is a probability α that this statistic exceeds a critical point $\chi^2_{p,\alpha}$, so that the overall error rate can be maintained exactly at the level α by the use of a Shewhart type χ^2 control chart for T_i^2 with an upper control limit (CL) of $\chi^2_{p,\alpha}$. If $\vec{\mu}_i \neq \vec{\mu}_0$, T_i^2 is distributed as a non-central chi-squared distribution with p degrees of freedom and with non-centrality parameter $\lambda = n(\vec{\mu}_i - \vec{\mu}_0)' \Sigma^{-1} (\vec{\mu}_i - \vec{\mu}_0)$, $\lambda = nd^2$, where d is the Mahalanobis' distance of $\vec{\mu}_1$ with respect to $\vec{\mu}_0$.

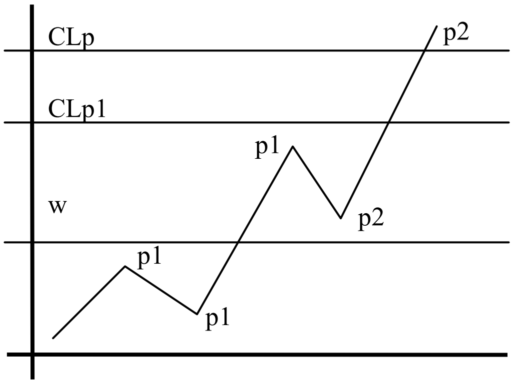

Figure 1. VDT^2 Control Chart

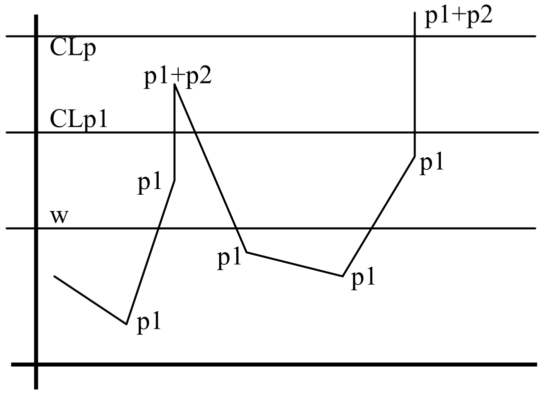

Figure 2. DDT^2 Control Chart

However, as we have commented in the introduction, sometimes p_2 out of the p variables are difficult and/or expensive to measure, while the remaining p_1 variables are easy and cheap to measure. Hence, the total set of p variables *can be* separated in two groups, where p_1 variables are always monitored and all the $p = p_1 + p_2$ variables are monitored only when there is a warning about the stability of the process.

Two charts are proposed. The Variable-Dimension T^2 control chart (VDT^2) is shown on Figure 1. The procedure employed in this chart is similar to the one of the variable sampling T^2 control chart (Aparisi, 1996). The monitoring begins with a sample where only p_1 variables are measured. If the T^2 value computed for these p_1 variables (T^2_{p1}) exceeds the control limit, CL_{p1}, then it is assumed that the process is out of control. If $T^2_{p1} < w$, where w is a warning limit, then for the next sample only the set of p_1 variables will be measured. If $w \leq T^2_{p1} \leq CL_{p1}$ then in next sample all the variables $p = p_1 + p_2$ will be measured to compute T^2_p. When all the variables are measured the control limit is CL_{p2}. If $T^2_p < w$ then in next sample only the set of p_1 variables is considered.

This chart is a simplification of the chart when two warning limits are present, w_{p1}, when p_1 variables are considered, and w_p, when all the variables are measured. Our results show that although a chart with two warning limits shows better performance, the improvement is not large.

The second chart proposed is the Double-Dimension T^2 control chart (DDT^2), similar to the double sampling T^2 control chart (Champ and Aparisi, 2008), where for the DDT^2 chart the variable parameter is the number of variables to be measured (see Figure 2). In this case, if $w \leq T^2_{p1} \leq CL_{p1}$ the rest of variables (set p_2) of the sample is measured and combined with the measurements of the set of p_1 variables in order to obtain T^2_p, and this value is plotted in the same horizontal position on the chart, i.e, the decision of measuring all the variables is taken for the same sample, not waiting until the next sampling time.

3. ARL computation

The ARL of the VDT^2 chart is calculated employing a Markov's chain. In the case of only one warning limit, the following states are defined:

State 1: $T^2 < w$ (and the next sample will contain only p_1 variables)
State 2: $w \leq T^2 < CL$ (and the next sample will contain all p variables)
State 3: $T^2 > CL$ (signal; absorbing state).

Consider the transition probability matrix for a given shift, *d*

$$P_d = \begin{pmatrix} P_{1,1}^d & P_{1,2}^d & P_{1,3}^d \\ P_{2,1}^d & P_{2,2}^d & P_{2,3}^d \\ 0 & 0 & 1 \end{pmatrix}$$

For example $P_{1,2}^d$ is the transition probability from state 1 to state 2, that is, when p_1 variables are measured, T_{p1}^2 falls in zone [*w, CL*] when there is a shift *d*.

$$P_{1,2}^d = P(w < T_i^2 < CLp_1 \mid p_1, d) = P(w < \chi_{p1}^2(\lambda) < CLp_1)$$

where $\lambda = n\ d^2$.

When the zero-state scenario si considered $ARL(d) = \vec{B}(I - Q_d)^{-1}\vec{1}$, where I is the 2×2 identity matrix, $\vec{1} = (1 \quad 1)^T$ and $\vec{B} = (b_1, \quad b_2)$ is the 1×2 vector of initial state probabilities. Hence $b_1 + b_2 = 1$. Q_d is the 2×2 probability transition matrix where the elements associated with the absorbing state have been deleted. I is the identity matrix of order 2. The ARL for the steady-state case is $ARL(d) = \vec{B}(I - Q_d)^{-1} / ARL_0$.

ARL calculations for the DDT^2 control chart do not need a Markov chain approach. In this case, ARL = 1 / (1 – *P*(*no signal*)) where *P*(*no signal*) is the probability that the chart does not signal, given by:

$$P(no\,signal) = \int_0^w f_{\chi_{p1}^2}(v)dv + \int_w^{CLp_1} f_{\chi_{p1}^2}(v)\int_0^{\frac{CL_p - v}{p_2}} f_{\chi_{p2}^2}(u)dudv$$

In this paper, ARL results for the DDT2 chart are not shown due to length limitations.

4. Optimization and Performance Comparison

Software for finding the best parameters for VDT^2 and DDT^2 control charts employing Genetic Algorithms (GA) has been developed. Figure 3 shows the software for optimizing the VDT^2 chart with one warning limit.

The optimization is carried out for the following case: $p_1 = 2$, $p_2 = 2$, $p = 2 + 2 = 4$. Desired in-control ARL: 400. The user has to specify the shift for which the out-of-control ARL should be minimized. This specified out-of-control mean vector is at a Mahalanobis distance d_{p1} for the set of p_1 variables and at another Mahalanobis' distance d_p for the set of p variables. It results obvious that VDT^2 and DDT^2 charts are only useful when $d_p > d_{p1}$, i. e., when adding more variables to the monitoring increases the performance of the scheme. In our example, the shift selected for having lowest out-of-control ARL has a shift size $d_{p1} = 0.5$ Mahalanobis' distance when $p_1 = 2$ variables are measured and a shift size $d_p = 1$ when $p = 4$ variables are monitored.

Figure 3 shows the result of the optimization for VDT^2 control chart. The resulting warning limit is $w = 2.61$, which has an in-control probability left tail equal to 0.729. The control limit for p_1 variables is 40.95, right tail probability = 0, which means that no control limit is needed for p_1 variables. This is a common result for this scheme, and simplifies the use of the chart, as only the warning limit and one control limit (for the p variables) are required. The control limit for the $p = p_1 + p_2$ variables is 14.44, with a right tail probability of 0.00602. This chart has an in-control ARL of 399.94. The ARL for the given shift is 101.31.

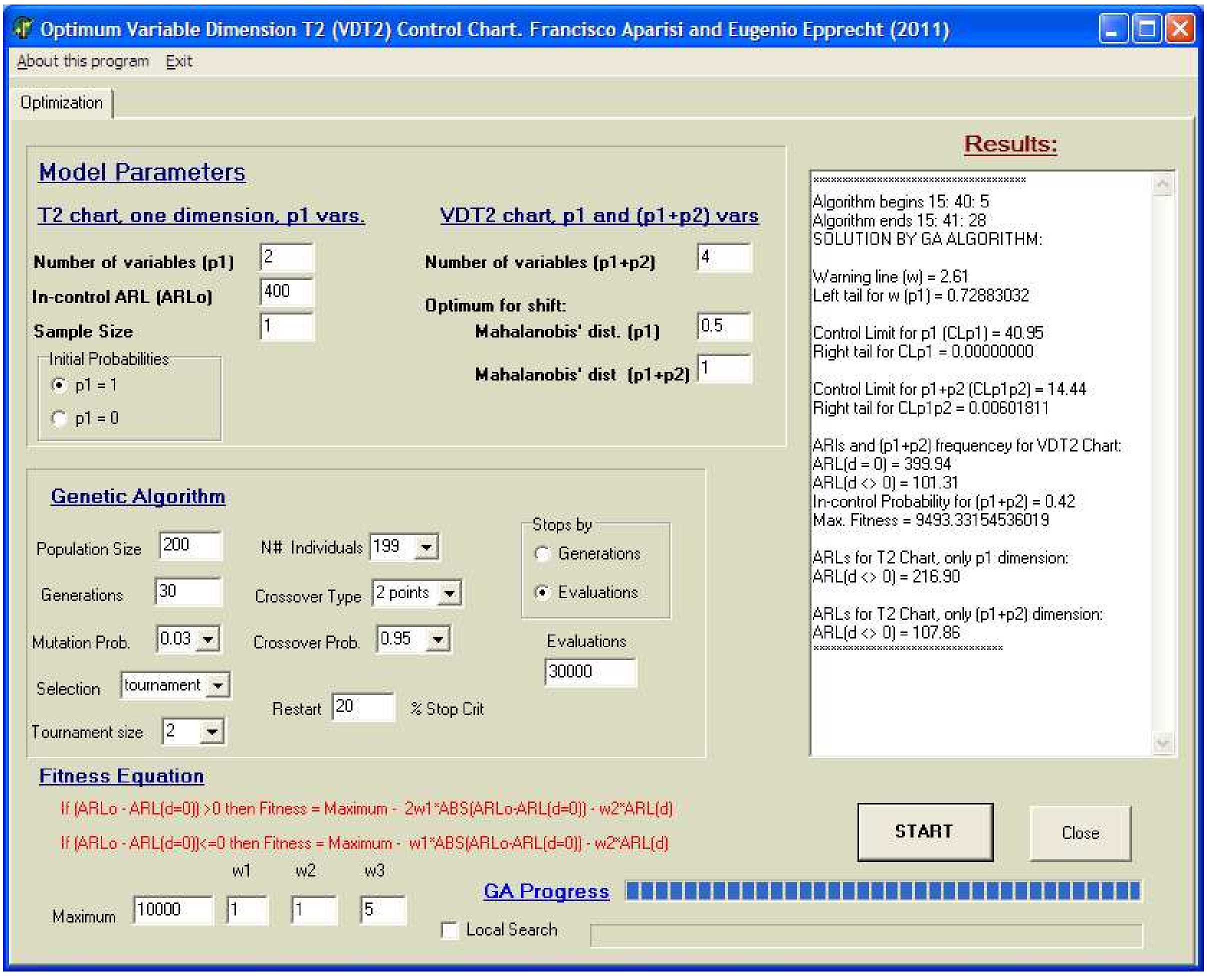

Figure 3. Software for the optimization of the VDT^2 control chart

The next step is to compare the performance of this chart with that of a standard T^2 control chart where always p_1 variables are measured and versus a T^2 control chart where always all the p variables are monitored at every sampling time. The T^2 control with p_1 variables has an out-of-control ARL of 216.90, more than twice the ARL of the VDT^2 control chart. In addition, the VDT^2 outperforms the control chart when all the variables are always measured, having this chart an out-of-control ARL = 107.86, although the improvement is marginal. However, the sampling cost of VDT^2 versus a T^2 employing all the variables may be quite smaller. How much smaller? The software returns the percentage of times than we must sample all the variables when the process is in control, 42%. Therefore, it is easy to compute the average cost of sampling of the VDT^2. If the cost of measuring the p_1 variables is c_1 and the cost of measuring the p_2 variables is c_2, then the average cost per sample is $c_1 + 0.42c_2$ (against a fixed cost of $c_1 + c_2$ of the T^2 chart with which all p variables are always measured — a relative economy of $58c_2 / (c_1 + c_2)$% per sample).

The optimum VDT^2 has an average cost of sampling that is 0.42(sampling cost for p) + 0.58(sampling cost for p_1) per sample. In some cases this cost could be still high. Another version of the software has been written where the percentage of times that all the variables are measured when the process is in control is a restriction for the optimization. This software is not shown here due to paper length limitations. If for example, this percentage is fixed at 20%, in order to reduce average sampling costs, the optimization returns: Warning limit: w = 3.82. Control limit for p_1 variables = 24.35, right tail probability = 0.00000470 (practically the control limit can be set to infinite). Control limit for $p = p_1 + p_2$ variables = 12.80, right tail probability = 0.01229719. This chart has an in-control ARL = 400. The ARL of this scheme need to detect the shift is 105.74. That means that with only a marginal reduction in performance the sampling cost could be significantly reduced.

Acknowledgments

This work has been supported by the Ministry of Education and Science of Spain, research project number DPI2009-09925 and the support CNPq (the Brazilian Council for Scientific and Technological Development), project number 201811/2007-3.

REFERENCES (RÉFERENCES)

Aparisi, F. (1996). Hotelling's T2 control chart with adaptive sample sizes. *International Journal of Production Research*, 34, 2853-2862.

Aparisi, F., De Luna, M., Epprecht, E. (2010). Optimisation of a set of $\overline{X}$ or principal components using genetic algorithms. *International Journal of Production Research*, 48(18), 5345-5361.

Champ, C. W. and Aparisi, F. (2008). Double Sampling Hotelling's T2 Charts. *Quality and Reliability Engineering International*, 24, 153-166.

Epprecht, E., De Luna, M. Aparisi, F. (2011). Joint EWMA Charts for Multivariate Process Control: Markov Chain and Optimal Design. *International Journal of Production Research*, Accepted for publication.

Montgomery, D. C. (2008). *Introduction to Statistical Quality Control*, 6th Edition. Willey, New York.