Proceedings of the 2010 International Conference on Industrial Engineering and Operations Management
Dhaka, Bangladesh, January 9 – 10, 2010

Designing a Quality Control System for Mean Vector and Change Point Diagnosis in Multivariate Normal Process

Shima Mohebbi
Department of Industrial Engineering
K.N.Toosi University of Technology, Tehran 1999143344, Iran

Iraj Mahdavi and Zahra Hejazizadeh
Department of Industrial Engineering
Mazandaran University of Science and Technology, Babol 734, Iran

Bahram Sadeghpour Gilde
Department of Statistics
University of Mazandaran, Babolsar 47415, Iran

Abstract

This paper considers the problem of contamination in process distribution as well as identifying a combination of variables which are responsible for the change in process. Three interactive steps for the control of mean vector in a contaminated multivariate normal process are proposed. The first step is to estimate the process parameters by robust statistics. The control chart is introduced in the second step. In the third step, we consider a change point model to monitor a step change in mean vector. The simulation results reveal the efficiency of model for detecting a specific step change in the process.

Keywords

Contamination, Multivariate Normal, Change Point

1. Introduction

When, an industrial process is first subjected to statistical quality control procedures, an early step usually is the construction of control charts. A typical procedure might be to select 20-40 rational subgroups of observations and calculate the mean and range of each subgroup. The control limits can be established from the mean of the subgroup means in such a way that the subgroup statistics would fall outside these limits only rarely if the process were in a state of statistical control. When a control chart signals that an especial cause is present, process engineer must initiate a search for the especial cause of the process disturbance. The search will depend on the process engineers' expertise and knowledge of their process. Identifying which combination of the process variables is responsible for the change in the process allows engineers to be proactive in improving quality by preventing changes that lead to a poor quality, and in perpetuating those changes and optimizing the settings of variables that improve quality [9]. Hence, although, control charts are generally better at detecting isolated abnormal points and at detecting a major change quickly, a change point model is required to detect subtle changes which frequently missed by control charts. Hence, the two methods can be used in a complementary fashion. Additionally, knowing when a process changed would simplify the search for the special cause. If the time of the change could be determined, process engineers would have a smaller search window within which to look for the special cause [9]. Postulating the multivariate normal as the most widely used setting for multivariate SPC, Zamba & Hawkins [10] have classified the important departure from control for multivariate normal into 5 categories: 1) the mean vector changes while the covariance structure remains unchanged, 2) the covariance structure could be perturbed, 3) both mean and covariance could have a step change, 4) one or both of these parameters could drift, and 5) the distribution could change from normal to some other form. Here, we concentrate on the first case; when there is a shift in the mean vector. However, there are some occurrences to which one's attention should be drawn. For the control charts' procedure, classical estimates such as the sample mean and sample covariances can be adversely influenced by atypical data, called outlier, and often fail to provide good fits to the bulk of data. Thus, it is important that the charting procedure be as sensitive as possible to problem of control, while still maintaining a fixed limit on the number of false alarms [8].

Classical methods assume that the process distribution belongs to an exactly known parametric family of distribution, most likely normal distribution. At this point it may seem natural to think that an adequate procedure could be to test the hypothesis that the data are normal; if it is not rejected, we use the maximum likelihood estimate (MLE) for the fitted distribution. But this has the drawback that very large sample sizes are needed to distinguish the true distribution, especially the tail-precisely the region with fewer data-are most influential [6]. The presence of outliers tends to reduce the sensitivity of control charting procedures so that the detection of the outliers becomes less likely. There exist robust parameter estimates that provide a good fit to the bulk of data and are reliable particularly in the case of multivariate data.

In this paper we introduce a new system composed of three interactive steps. The first step uses robust statistics to estimate the process parameters. Afterwards, a control chart is introduced based on multivariate maximum likelihood ratio to monitor mean vector in the second step. The third step equips the control chart with a change point model to monitor the step change in mean vector. The proposed system will be appropriate when the following conditions are satisfied.

1) The control chart is to be established from the data at hand, rather than from known parameters. This will usually be the case when a process is newly brought under statistical quality control procedures.
2) There is possibility of outlier within the subgroup. If the distribution is known to be exactly normal (which would rarely be the case), then MLE will suffice. The term "outlier" is used to mean an observation that is unusually far from the rest of the data. This may indicate that the particular data point was drawn from a different population or that a sporadic especial cause was operating.
3) The mean vector of process alters after an unknown point at time τ. While this step change in the mean vector occurs, the process remains at the new level until control chart issue a signal.

In the next section, we present proposed quality control system through three steps.

2. Proposed Quality Control System

Our proposed quality control system has three interactive steps for the control of mean vector (See Figure 1).

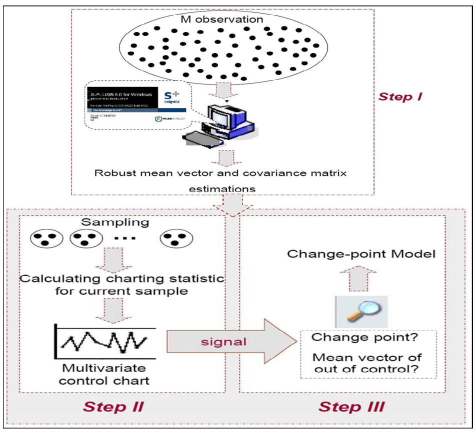

Figure 1: Proposed quality control system

The first step is to estimate mean vector and covariance matrix robustly using Stahel-Donoho estimators. Based on the outcomes of the first step and maximum likelihood ratio, the charting statistic can be constructed. If the control chart issues a signal, the change point model will monitor the step change in mean vector. In the next subsections, each step will be discussed in detail.

2.1 Step I: Robust Identification

It is possible that each process produces outliers. A simple way to handle outliers is to detect and remove them from the data set. Deleting an outlier still poses a number of problems. Since there is generally some uncertainty as to whether an observation is really atypical, there is a risk of deleting good observations which result in underestimating data variability. In addition, the result depends on the user subjective decisions and thereby it is difficult to determine the statistical behavior of the complete procedure. Therefore, the scholars have been motivated to apply statistical approaches for confronting outliers.

Here, we apply Stahel-Donoho estimator which is defined as a weighted mean and a weighted covariance matrix. The weight of each point is a function of an “outlyingness” measure. The outlyingness measure *r* is based on the idea that if a point is a multivariate outlier then there must be some one-dimensional projection of the data for which the point is a univariate outlier. Suppose $X = (x_1, x_2, ..., x_n)^T$ is a set of *n* points in R^p. The outlyingness *r* of each observation x_i is computed by finding the direction $a \in A$ where $A = \{a \in R^p \mid \|a\| = 1\}$ and r_i is given in formula (1):

$$r_i = \sup_{a \in A} \frac{\left| x_i' a - med \{x_j' a\}_{j=1}^n \right|}{MAD \{x_j' a\}_{j=1}^n} \tag{1}$$

The weight of each observation, w_i, is computed using the function of outlyingness as shown in equation (2):

$$w(r_i; C) = \begin{cases} 0 & if \left|\frac{r_i}{C}\right| > 1 \\ b_1 + b_2\left(\frac{r_i}{C}\right)^2 + b_3\left(\frac{r_i}{C}\right)^4 + b_4\left(\frac{r_i}{C}\right)^6 & if\, 0.8 < \left|\frac{r_i}{C}\right| \le 1 \\ 1 & if \left|\frac{r_i}{C}\right| \le 0.8 \end{cases} \tag{2}$$

where, $b_1 = -19.71879, b_2 = 82.30453, b_3 = -105.45267, b_4 = 42.86694.$

The tuning constant (*C*) is set, by default, to be the square root of the .95 quantile of a chi squared distribution with *p* degrees of freedom. The Stahel-Donoho estimator of location and scale are defined in equation (3) and equation (4) respectively:

$$\hat{\mu} = \frac{\sum_{i=1}^n w_i . x_i}{\sum_{i=1}^n w_i} \tag{3}$$

$$\hat{\Sigma} = \frac{\sum_{i=1}^n w_i .(x_i - \hat{\mu})(x_i - \hat{\mu})'}{\sum_{i=1}^n w_i}. \tag{4}$$

In this paper we suggest generate $p \times m$ matrix of process observations (which p is the numbers of process variables and m is the numbers of process observations) and then, this matrix is given to *S-PLUS* software- that based on Stahel-Donoho estimator- to estimate robust mean vector and robust covariance matrix. Therefore, we introduce the taken output from the S-PLUS as robust mean vector of process, $\hat{\mu}_{Robust} = \hat{\mu}_R$, and robust covariance matrix of process, $\hat{\Sigma}_{Robust} = \hat{\Sigma}_R$, and we will use them in step II , III.

2.2 Step II: Multivariate Control Chart

Based on the outcomes of the previous step, the charting statistic can be constructed. For this purpose, maximum likelihood ratio is applied to detect the mean vector deviation. The test hypotheses and simplified test statistics is given in equation (5) and (6) respectively:

$$\begin{cases} H_0 : \mu = \hat{\mu}_R \\ H_1 : \mu \ne \hat{\mu}_R \end{cases} \tag{5}$$

$$\Lambda = \frac{\max_{\Sigma} L(\hat{\mu}_R, \Sigma)}{\max_{\mu,\Sigma} L(\mu, \Sigma)} = \left(\frac{\left| \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' \right|}{\left| \sum_{j=1}^{n} (x_j - \hat{\mu}_R)(x_j - \hat{\mu}_R)' \right|} \right)^{\frac{n}{2}} \tag{6}$$

where, $\hat{\mu}_R$ is the robust estimation of process mean vector. If the value of this ratio is very small, the H_0 will be rejected in favor of H_1. Using the relationship between T^2 statistics and Λ [4] the rejection area can be obtained as shown in equation (7):

$$\Lambda^{2/n} < \left(1 + \frac{p}{(n-p)} F_{p,n-p}(\alpha) \right)^{-1} \tag{7}$$

Hence, we introduce the charting statistic for each subgroup and control limit as shown in formula (8):

$$\text{Charting statistic} = \frac{1}{\Lambda^{2/n}} = \left| \sum_{j=1}^{n} (x_j - \hat{\mu}_R)(x_j - \hat{\mu}_R)' \right| \Big/ \left| \sum_{j=1}^{n} (x_j - \bar{x})(x_j - \bar{x})' \right|$$
$$\text{Control limit} = 1 + \big(p/(n-p)\big) F_{p,n-p}(\alpha). \tag{8}$$

If $Charting \quad Statistics \geq 1 + \big(p/(n-p)\big) F_{p,n-p}(\alpha)$,then chart will signal.

2.3 Step III: Change Point Model

When the quality chart signals that a process change has occurred, the change point model is then applied to estimate the real time of change, τ ,and out-of-control mean vector, μ_a. Using the similar concept advocated by Harrell [3], change point and out-of-control mean vector can be estimated as given in equation (9) and equation (10) respectively:

$$\hat{\tau} = \arg \max_{0 \le c < T} \frac{n(T-c)}{2} \left[\left(\hat{\mu}_R - \overline{\overline{x_{T,c}}} \right)' \hat{\Sigma}_R^{-1} \left(\hat{\mu}_R - \overline{\overline{x_{T,c}}} \right) \right]. \tag{9}$$

$$\hat{\mu}_a(\hat{\tau}) = \frac{1}{T - \hat{\tau}} \sum_{t=\hat{\tau}+1}^{T} \bar{x}_t = \overline{\overline{x_{T,\hat{\tau}}}}. \tag{10}$$

where, T is the subgroup number which causes the control chart issues a signal, index variable c is a candidate change point ranging from zero to subgroup T-1, $\hat{\Sigma}_R$ is the robust estimation of covariance matrix and $\hat{\tau}, \hat{\mu}_a(\hat{\tau})$ are MLE estimators of τ, μ_a. In the next section, we evaluate the performance of the proposed system using simulation.

3. Performance Evaluation

In order to evaluate the performance of the proposed system, a simulation study is conducted for different number of variables and contamination percentages. To formalize the idea of contaminated normal, we may imagine that a proportion $(1-\varepsilon)\%$ of the observations are generated by the normal model $N_p(\mu_1, \Sigma_1)$, while a proportion $\varepsilon\%$ is generated by $N_p(\mu_2, \Sigma_1)$. We first generate 100 random observations from a contaminated normal distribution. Using *S-Plus*, well-known software for robust statistics, we estimate the true parameters of the process, $\hat{\mu}_R, \hat{\Sigma}_R$, based on the robust estimators presented in section 3.1. Then, we generate 50 in-control random samples in subgroup of size *n* from a contaminated normal distribution. Starting with subgroup 51, observations are generated randomly with mean $\mu_a = \mu + (\delta, .., \delta)'$ and $\mu_a = \mu + (\delta, 0, .., 0)'$ and similar covariance matrix until the chart issues a signal. This procedure is repeated a total of 1000 times for each value of δ studied. The change point model is applied for each simulation run. Then, the frequency with which the model correctly identified the actual time of change is computed. All the aforementioned process is executed in *Matlab7* software. Table 1 and 2 illustrates the results for 3- and 5-variate normal distribution (p = 3, 5) with 6 and 10 percentage of contamination. The performance is calculated for two different rejection areas (α =0.025, 0.05). Table 3 presents the performance of the system when contamination is 6% and just one variable shift in the process.

Table1: $\varepsilon = 6\%$ ‘ $\tau = 50$ ‘ $\mu \to \mu_a = \mu + (\delta, .., \delta)'$

	$p=3$						$p=5$					
	$\alpha=0.025$			$\alpha=0.05$			$\alpha=0.025$			$\alpha=0.05$		
δ	n=4	n=5	n=6	n=4	n=5	n=6	n=6	n=7	n=8	n=6	n=7	n=8
0.75	51.3%	54%	51.3%	45.6%	47.4%	43.3%	89.4%	85.2%	76.6%	85.7%	74.3%	66.9%
1	69.3%	68.7%	66.7%	62.5%	60.8%	59.9%	98.8%	97.1%	95.5%	96.7%	94.3%	95.3%
1.25	81.9%	81.1%	76.4%	79%	75.6%	70.3%	99.9%	99.9%	99.7%	99.6%	99.6%	100%
1.5	90.3%	87.6%	86.9%	88.1%	83.6%	82.2%	100%	100%	100%	99.9%	99.8%	100%
1.75	94.8%	94.7%	93.1%	92.3%	91.3%	89%	100%	100%	100%	99.9%	100%	100%
2	97.7%	95.8%	96.9%	96.6%	94.1%	96%	100%	100%	100%	100%	100%	100%
2.25	99.1%	98.2%	99%	97.6%	97.8%	98.4%	100%	100%	100%	100%	100%	100%
2.5	99.3%	99.3%	99.5%	99.1%	98.9%	99.2%	100%	100%	100%	100%	100%	100%
2.75	99.5%	99.7%	99.9%	99.8%	99.8%	100%	100%	100%	100%	100%	100%	100%
3	99.8%	100%	99.8%	99.9%	99.9%	100%	100%	100%	100%	100%	100%	100%
3.25	100%	100%	100%	99.8%	100%	100%	100%	100%	100%	100%	100%	100%
3.5	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

Table2: $\varepsilon = 10\,\%$ ، $\tau = 50$ ، $\mu \rightarrow \mu_a = \mu + (\delta,..,\delta)'$

	$p=3$						$p=5$					
	$\alpha=0.025$			$\alpha=0.05$			$\alpha=0.025$			$\alpha=0.05$		
δ	n=4	n=5	n=6	n=4	n=5	n=6	n=6	n=7	n=8	n=6	n=7	n=8
0.75	50.1%	51.4%	49.6%	44.5%	39.5%	38.7%	80.4%	61%	35.7%	69.5%	38.8%	23.6%
1	71.2%	69.9%	67.4%	64.7%	60.6%	56.2%	89.8%	68.8%	50.2%	82.5%	56.2%	38%
1.25	84.5%	81.2%	75.7%	78.5%	74.3%	66.2%	95.6%	85.2%	76.8%	91.6%	76.6%	78.3%
1.5	89.9%	87.6%	85.2%	87.8%	82.1%	80.7%	99.4%	97.7%	98.9%	98.4%	97.8%	99%
1.75	95.5%	94.6%	92.4%	93.7%	91.3%	87.7%	100%	100%	100%	100%	99.8%	100%
2	97.9%	96.1%	95.6%	95.6%	95.7%	95.4%	100%	100%	100%	99.9%	100%	100%
2.25	99%	98.4%	99%	98.7%	98.2%	98.6%	100%	100%	100%	100%	100%	100%
2.5	99.1%	99.6%	99.6%	99.3%	99.1%	99.6%	100%	100%	100%	100%	100%	100%
2.75	99.9%	99.9%	99.8%	99.5%	99.8%	99.7%	100%	100%	100%	100%	100%	100%
3	99.9%	99.9%	100%	100%	100%	99.9%	100%	100%	100%	100%	100%	100%
3.25	99.8%	99.9%	100%	99.8%	99.9%	100%	100%	100%	100%	100%	100%	100%
3.5	99.9%	100%	100%	99.9%	100%	99.9%	100%	100%	100%	100%	100%	100%

Table3: $\varepsilon = 6\%$ ، $\tau = 50$ ، $\mu \rightarrow \mu_a = \mu + (\delta,0,..,0)'$

	$p=3$						$p=5$					
	$\alpha=0.025$			$\alpha=0.05$			$\alpha=0.025$			$\alpha=0.05$		
δ	n=4	n=5	n=6	n=4	n=5	n=6	n=6	n=7	n=8	n=6	n=7	n=8
0.75	32.9%	32.9%	34.5%	23%	29.4%	30.5%	35.5%	30.5%	26.1%	25.1%	19.7%	15.2%
1	47.8%	52.9%	50.6%	44.5%	44.3%	45.3%	57.2%	52%	41%	45.1%	35.7%	26.9%
1.25	64.6%	67.2%	62.8%	68.5%	59.9%	55%	74.2%	63.1%	47%	62.5%	45%	31.7%
1.5	73.9%	77.7%	72.8%	79%	68.7%	67.8%	80.9%	72%	53.5%	70.1%	57.2%	37.9%
1.75	83.9%	83.7%	83.2%	86.3%	78.7%	78.5%	87.2%	77.9%	60%	80.1%	65.9%	44.6%
2	89.9%	87.7%	86.2%	89.7%	85%	80%	91.3%	79.3%	69.3%	85.8%	70.6%	55.4%
2.25	92.7%	92.8%	91%	95.4%	91.3%	91.5%	94.8%	87.4%	77%	90.6%	78.2%	72.6%
2.5	95.2%	95.8%	95.7%	96.4%	92.8%	93.3%	96.6%	91.4%	87.4%	93.7%	89.2%	85.2%
2.75	97.2%	97.5%	96.5%	97.2%	95.7%	97.3%	98.2%	96.3%	94.3%	97.5%	92.1%	93.9%
3	98.8%	98.8%	98.5%	99%	97.4%	98.7%	99.4%	98.3%	98.2%	98.6%	97.4%	98.1%
3.25	98.9%	99.2%	99.6%	99.2%	98.2%	99.2%	99.6%	99.7%	99.6%	99.5%	99.1%	99.3%
3.5	99.8%	99.5%	99.7%	99.5%	99.4%	99.4%	99.7%	100%	99.9%	99.6%	99.6%	99.9%

4. Application of Proposed Quality Control System

Manufacturing shops are equipped with diverse types of machines and tools and are distributed around many geographically different locations. Because of such geographical diversity, the quality control of sub-assemble products (SAPs) and capability of tracing similar SAPs become more important for the decision makers in the holding company. However, despite the merit of traditional statistical process control methods, there exist some limitations in the control and coordination of such manufacturing systems. Agent-based systems are alternative technology to design real-time quality control system because of certain feature such as distribution, and collaboration. The approach presented in this paper facilitates identifying the change causes in the process and well suit to the agent-based quality control systems. In addition, to coordinate distributed manufacturing shops, quality-based approach is an effective solution in the case of assembly products. Therefore, utilizing the proposed model to obtain the reliability profile of each SAP at corresponding shop can be considered as a determinant factor in coordinating the distributed shops for satisfying the market demand.

5. Conclusions

In this paper, we proposed a quality control system for mean vector of a contaminated multivariate normal. In order to obtain the true parameters of the process and deal with contamination, robust estimators were applied. Afterward, we considered a control chart which is equipped with change point model to detect the shifts more effectively. The performance of the system was then evaluated using simulation. The outputs revealed the proficiency of the system even in the case of process variables augmentation. In addition, the application of the proposed quality control system for distributed manufacturing shops (DMSs) was discussed. Further research may consider more complexity in the quality system. Monitoring covariance matrix and mean vector simultaneously is interesting area to consider for a quality control system. The use of proposed approach for coordinating DMSs will lead to more practical outcomes.

References

1. Hawkins, D. M., Qiu, P., and Kang, C.W., 2003, The change point model for statistical process control, *Technometrics*, 35(4), 355-388.
2. Hawkins, D. M., and Zamba, K. D., 2005, A change point model for a shift in variance, *Journal of Quality Technology*, 37(1), 21-31.
3. Harrell, R.M., 2007, A multivariate magnitude robust control chart for mean shift detection and change point estimation, Master Thesis, *Air Force Inst of Tech Wright-Patterson AFB OH School of Engineering and Management*.
4. Johnson, R.A., and Wichern, D.W., 1998, Applied multivariate statistical analysis, *Prentice Hall*.
5. Mason, R.L., and Yong, J.C. 2002, Multivariate Statistical Process Control with Industrial Application, *Pliladellphia: SIAM*.
6. Maronna, R.A., Martin, R.D., and Yohai, V. J., 2006, Robust statistics: Theory and Methods, *John Wiley & Sons*.
7. Nedumaran, G., Pignatiello, J.J., and Calvin, J.A., 2000-01, Identifying the time of a step change with χ^2 control charts, *Quality Engineering*, 13(2), 153-159.
8. Rocke, D.M., 1992, $\overline{X}_Q$ and R_Q charts: robust control charts, *The Statistician*, 41, 97-104.
9. Samuel, T.R., Pignatiello, J.J., and Calvin, J.A., 1998, Identifying the time of a step change in a normal process variance, *Quality Engineering*, 1(3), 529 – 538.
10. Zamba, K.Z., and Hawkins, D.M., 2000, A multivariate change point model for statistical process control", *Technometrics*, 48(4), 539-549.