Robust Individuals Control Chart for Shifts in Process Mean and Variance

NG KOOI HUAT¹, HABSHAH MIDI²,
¹²Laboratory of Applied and Computational Statistics,
Institute for Mathematical Research, University Putra Malaysia,
43400 Serdang, Selangor, MALAYSIA
E-mails:¹ nkhuat76@yahoo.com, ²habshah@putra.upm.edu.my

Abstract: Statistical process control (SPC) involves ongoing checks to ensure whether the mean or variance of the process readings has changed. When a change does occur, the control chart should locate it quickly. If the operator can also be supplied information that aids in the search for the special cause, then critical off-line time can be saved. We examine a process-monitoring tool that not only provides speedy detection regardless of the magnitude of the process shift, but also provides useful change point statistics. A Likelihood ratio method is used to develop test statistics for step change shifts. Apart from traditional charts, robust control charts is a better alternative for a process with contaminated data as the effects of the outlying data values are eliminated. In which robustness against assignable causes of variation appears to be important. The proposed methodology is demonstrated with numerical example.

Key-Words: Change Point Estimation; Statistical Process Control; Outlier; Maximum-Type.

1. Introduction

In the area of mathematical statistics, the decision whether the observed series remained stationary or whether a change of a specific kind occurred is usually based upon hypotheses testing. The null hypothesis claims that the process is stationary while the alternative hypothesis claims that the process is non-stationary and the stationarity was violated in a specific way. The simplest situation is when at the beginning, a certain process (e.g. that of manufacturing process) is assumed to vary around a certain constant \( a_0 \) whereby it is assumed to be in control. However, it can happen that as a failure of the production device, e.g., the observed characteristics suddenly start to alter around another out-of-control constant \( a_1 \neq a_0 \). It can happen that, on account of that sudden failure, the variance \( \sigma^2 \) may change as well, yet it is possible that the variance remains the same. Furthermore, sometimes the variance can be presumed to be known due to researcher’s long experience with the production process.

2. Change in mean and/or variance

In some cases, it can happen that the change point may happen either in one parameter or in both (simultaneously). Then, the null hypothesis \( H \) against the alternative \( A \) can be written as follows;

\[
H : Y_1, \ldots, Y_n \sim N(a, \sigma^2) \\
A : \exists m \in \{2, \ldots, n-2\} \text{ such that } \\
Y_1, \ldots, Y_m \sim N(a_1, \sigma_1^2) \\
Y_{m+1}, \ldots, Y_n \sim N(a_2, \sigma_2^2)
\]

where \( (a_1, \sigma_1^2) \neq (a_2, \sigma_2^2) \).

Jaromir et al. (1999) proposed the maximum-type test statistics to test the hypothesis in (1). This test is formulated based on the maximum likelihood approach which have the form

\[
\max_{1 \leq k \leq n-1} \left\{ \left| Z_k \right| \right\} = \left| Z_k \right|
\]

and

\[
\max_{\left\{ \beta_n \right\} \leq k \leq \left\{ \left[ 1 - \beta_n \right] \right\}} \left\{ \left| Z_k \right| \right\}
\]
where

\[
\bar{Z}_k^2 = n \log \left( \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y}_n)^2 \right) - k \log \left( \frac{1}{k} \sum_{i=1}^k (Y_i - \bar{Y}_k)^2 \right) - (n-k) \log \left( \frac{1}{n-k} \sum_{i=k+1}^n (Y_i - \bar{Y}_k^0)^2 \right)
\]

and \( \bar{Y}_k = \frac{1}{k} \sum_{i=1}^k Y_i \) and

\[
\bar{Y}_k^0 = \frac{1}{n-k} \sum_{i=k+1}^n Y_i
\]

and \( \beta \) is a small positive constant less than one and \( \lfloor x \rfloor \) indicates the integer part of \( x \). The benefit of statistic (3) is that they are bounded in probability. The trimming off a \( 100 \beta \% \) portion of the sample (upper time points) implies that one assumes that the change did not occur during this time period. It is important to note that generally we take \( \beta \in [0.01, 0.1] \). The decision on “How much to trim off?” depends on the subjective decision of the statistician and his/her a priori knowledge of the problem. If the statistician decides to trim off only a very small portion of the time points or no time points (observations) at all, he/she pays for it by a loss of the power of his/her test as the critical values depend rather strongly on the value of \( \beta \).

For the decision about rejection of the null hypothesis \( H \), we need to know critical values of the suggested test statistics. It means to know their distribution under \( H \). Since the distribution of statistics (2) and (3) are not tractable, their critical values were obtained by simulation. Each type of the test statistic (2) and (3) with different sample sizes was repeated for 10000 times under \( H \) model and the corresponding desired percentile critical value was obtained from generated empirical distribution. One would reject \( H \) if test statistics of (2) or (3) is greater than the desired significance level \( \alpha \). The simulated critical values of statistic (2) and (3) are presented in Table 1-3.

For \( n \) large, Gombay and Horvath (1990) showed that the limit behavior of the studied probabilities is as follows

\[
P \left( \max_{1 \leq k \leq n-1} \left| \bar{Z}_k \right| > \frac{x + b_n}{a_n} \right) \approx 1 - \exp \left( -2e^{-x} \right),
\]

\( x \in \mathbb{R}^1 \),

\[
a_n = \sqrt{2 \log \log n} \quad \text{and} \quad b_n = 2 \log \log n + \log \log \log n.
\]

\[
P \left( \max_{(\beta n) \leq k \leq (1-\beta)n} \left| \bar{Z}_k \right| > x \right) \approx e^{-x^2/2} + e^{-x^2/2}x^2 \log \frac{1-\beta}{\beta}.
\]

### 3. Max-Type Critical Values

<table>
<thead>
<tr>
<th>( n )</th>
<th>Over-All Maximum-Type</th>
<th>Trimmed Maximum-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.087</td>
<td>3.076</td>
</tr>
<tr>
<td>20</td>
<td>3.174</td>
<td>3.173</td>
</tr>
<tr>
<td>30</td>
<td>3.301</td>
<td>3.263</td>
</tr>
<tr>
<td>40</td>
<td>3.289</td>
<td>3.271</td>
</tr>
<tr>
<td>50</td>
<td>3.322</td>
<td>3.327</td>
</tr>
<tr>
<td>100</td>
<td>3.375</td>
<td>3.387</td>
</tr>
</tbody>
</table>

Table 1: Simulated 2.5% critical values of the over-all maximum-type test statistic (2) and the corresponding trimmed maximum-type statistics (3) for different trimming portions \( \beta \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>Over-All Maximum-Type</th>
<th>Trimmed Maximum-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2.841</td>
<td>2.817</td>
</tr>
<tr>
<td>20</td>
<td>2.942</td>
<td>2.956</td>
</tr>
<tr>
<td>30</td>
<td>3.046</td>
<td>3.021</td>
</tr>
<tr>
<td>40</td>
<td>3.031</td>
<td>3.040</td>
</tr>
<tr>
<td>50</td>
<td>3.090</td>
<td>3.076</td>
</tr>
<tr>
<td>100</td>
<td>3.142</td>
<td>3.142</td>
</tr>
</tbody>
</table>

Table 2: Simulated 5% critical values of the over-all maximum-type test statistic (2) and the corresponding trimmed maximum-type statistics (3) for different trimming portions \( \beta \).
and test the significance of these shifts, namely proposed by De Mast and Roes (2004). In this paper, we consider the max-type change point detection method [4]. The desired critical values for testing the significance of location shifts are displayed in Tables 1-3. De Mast and Roes (2004) proposed a robust procedure to estimate $\mu_1$ and $\mu_2$ to accommodate the presence of possible outliers. According to De Mast and Roes (2004) robust procedure, $\hat{\mu}_1[r]$ and $\hat{\mu}_2[r]$ be the solution of
\[
\sum_{i=1}^{r} \psi \left( \frac{(y_i - \mu)}{c_0 s[r]} \right) = 0
\]
(11)
where
\[
s_0[r] = \text{median} \{|y_i - m|\}_{i=1,...,n}
\]
with $m = m_1[r]$ if $1 \leq i \leq r$ and $m = m_2[r]$ if $r + 1 \leq i \leq n$ for which $m_1[r]$ and $m_2[r]$ are the medians of $y_1,...,y_r$ and $y_{r+1},...,y_n$, respectively. $\psi$ is an odd function and $c$ is a tuning constant. It can be observed that $\hat{\mu}_1[r]$ and $\hat{\mu}_2[r]$ are the M-estimates for the location based on the scale initial estimate denoted as $s_0[r]$. The the final scale estimate is given by
\[
\sqrt{n cs_0[r]} \left[ \frac{\sum_{i=1}^{r} \psi'((y_i - \hat{\mu}_1[r])/c_0 s_0[r])}{\sum_{i=r+1}^{n} \psi'((y_i - \hat{\mu}_1[r])/c_0 s_0[r])} \right]^{1/2}
\]
(12)
with $\psi'$ indicating the derivative of $\psi$. The scale estimate obtained by employing the asymptotic variance of the M-estimators for location in order to estimate the standard deviation of the error. This kind of scale estimator in Eq. (12) is called an A-estimators by Lax (1985). There are many psi function to choose from, and in this study, De Mast and Roes (2004) proposed to use bisquare function which is given by
\[
\psi(u) = \begin{cases} 
  u(1-u^2)^2, & |u| \leq 1, \\
  0, & |u| > 1.
\end{cases}
\]
(13)

4. Methodology

De Mast and Roes (2004) developed the subsequent control charting procedure that is illustrated as follow.

1. Determine the locations of possible shifts and test the significance of these shifts. Upon completion of this step, the original data is divided into intervals in which the mean of the measurements is presumed constant.

2. Estimate (using robust estimators) the means of the intervals between successive shifts. As well, estimate (using robust estimators) the variance of the in-control measurements.

3. On the basis of these estimates, estimate a pair of control limits for each interval. Points are identified as outliers if they fall beyond these control limits.

Let the in-control process be denoted as $y_i$, $i = 1,...,n$, with the following null model:
\[
y_i = \mu + \epsilon_i, \quad \text{for} \quad i = 1,...,n
\]
(9)
with $\epsilon_i$ i.i.d. $N(0,\sigma^2)$.

Assuming that a single shift has occurred. Then model (9) is written as follows:

\[
\begin{cases} 
  y_i = \mu_1 + \epsilon_i & \text{for} \quad i = 1,...,\tau \\
  y_i = \mu_2 + \epsilon_i & \text{for} \quad i = \tau + 1,...,n
\end{cases}
\]
(10)
with $\epsilon_i$ i.i.d. $N(0,\sigma^2)$;

The main aim is to estimate the location shift, $\tau$. There are two change point detection methods to determine the location of shifts $\tau$ and test the significance of these shifts, namely proposed by De Mast and Roes (2004). In this paper, we consider the max-type change point detection method [4]. The desired critical values for testing the significance of location shifts are displayed in Tables 1-3. De Mast and Roes (2004) proposed a robust procedure to estimate $\mu_1$ and $\mu_2$ to accommodate the presence of possible outliers. According to De Mast and Roes (2004) robust procedure, $\hat{\mu}_1[r]$ and $\hat{\mu}_2[r]$ be the solution of
\[
\sum_{i=1}^{r} \psi \left( \frac{(y_i - \mu)}{c_0 s[r]} \right) = 0
\]
(11)
where
\[
s_0[r] = \text{median} \{|y_i - m|\}_{i=1,...,n}
\]
with $m = m_1[r]$ if $1 \leq i \leq r$ and $m = m_2[r]$ if $r + 1 \leq i \leq n$ for which $m_1[r]$ and $m_2[r]$ are the medians of $y_1,...,y_r$ and $y_{r+1},...,y_n$, respectively. $\psi$ is an odd function and $c$ is a tuning constant. It can be observed that $\hat{\mu}_1[r]$ and $\hat{\mu}_2[r]$ are the M-estimates for the location based on the scale initial estimate denoted as $s_0[r]$. The the final scale estimate is given by
\[
\sqrt{n cs_0[r]} \left[ \frac{\sum_{i=1}^{r} \psi'((y_i - \hat{\mu}_1[r])/c_0 s_0[r])}{\sum_{i=r+1}^{n} \psi'((y_i - \hat{\mu}_1[r])/c_0 s_0[r])} \right]^{1/2}
\]
(12)
with $\psi'$ indicating the derivative of $\psi$. The scale estimate obtained by employing the asymptotic variance of the M-estimators for location in order to estimate the standard deviation of the error. This kind of scale estimator in Eq. (12) is called an A-estimators by Lax (1985). There are many psi function to choose from, and in this study, De Mast and Roes (2004) proposed to use bisquare function which is given by
\[
\psi(u) = \begin{cases} 
  u(1-u^2)^2, & |u| \leq 1, \\
  0, & |u| > 1.
\end{cases}
\]
(13)

<table>
<thead>
<tr>
<th>$n$</th>
<th>Over-All Maximum-Type</th>
<th>Trimmed Maximum-Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta = 0.01$</td>
<td>$\beta = 0.05$</td>
</tr>
<tr>
<td>10</td>
<td>2.562</td>
<td>2.539</td>
</tr>
<tr>
<td>20</td>
<td>2.684</td>
<td>2.673</td>
</tr>
<tr>
<td>30</td>
<td>2.765</td>
<td>2.763</td>
</tr>
<tr>
<td>40</td>
<td>2.774</td>
<td>2.793</td>
</tr>
<tr>
<td>50</td>
<td>2.817</td>
<td>2.815</td>
</tr>
<tr>
<td>100</td>
<td>2.889</td>
<td>2.897</td>
</tr>
</tbody>
</table>

Table 3: Simulated 10% critical values of the overall maximum-type test statistic (2) and the corresponding trimmed maximum-type statistics (3) for different trimmed portions $\beta$. 


Proceedings of the 9th WSEAS International Conference on APPLICATIONS of COMPUTER ENGINEERING
Hence, in other words, observations further away from the mean will be more and more downweighted, and observations beyond \( c \approx 9 \) for the tuning constant seem to perform well.

De Mast and Roes (2004) pointed out that the value \( c = 9 \) for the tuning constant seem to perform well.

It is important to note that once the shift has been detected, the data are split into two groups, justifying whether more shifts can be detected. De Mast and Roes (2004) recommended that process will be continued until the size of the groups becomes smaller than 4 or any other selected minimum value.

Let \( \hat{\mu}_1, \ldots, \hat{\mu}_k \) be the estimated means of the groups in between the detected shifts \( \hat{\tau}_2, \ldots, \hat{\tau}_k \), the standard deviation of the error is estimated by:

\[
\hat{\sigma} = \sqrt{\frac{1}{n-k} \sum_{i=r_{k+1}}^{n} \left( y_i - \hat{\mu}_k \right)^2 / (c s_0^2) + \ldots + \frac{1}{n-k} \sum_{i=r_{k+1}}^{n} \left( y_i - \hat{\mu}_k \right)^2 / (c s_0^2)}
\]

The Median Absolute Deviation (MAD) is used for the initial scale, \( s_0 \) which is defined as \( s_0 = \text{median} \{ y_j - m \} \) where \( m \) denoting the median of the corresponding subgroup. The factor \( \sqrt{n} \) in the numerator of Eq. (14) is replaced by \( \sqrt{n^2(n-k)} \) for the sake of accounting the loss of degrees of freedom in the estimation of \( \hat{\mu}_1, \ldots, \hat{\mu}_k \).

For each identified group, \( j = 1, \ldots, k \), De Mast and Roes (2004) proposed the following control limits:

\[
\begin{align*}
UCL_j &= \hat{\mu}_j + h \sqrt{\frac{\hat{\tau}_{j+1} - \hat{\tau}_j - 1}{\hat{\tau}_{j+1} - \hat{\tau}_j}} \hat{\sigma} \\
LCL_j &= \hat{\mu}_j - h \sqrt{\frac{\hat{\tau}_{j+1} - \hat{\tau}_j - 1}{\hat{\tau}_{j+1} - \hat{\tau}_j}} \hat{\sigma}
\end{align*}
\]

We define \( \hat{\tau}_1 = 0 \) and \( \hat{\tau}_{k+1} = n \). The conventional value of \( h = 3 \) is employed and the factor \( \sqrt{\hat{\tau}_{j+1} - \hat{\tau}_j - 1} / (\hat{\tau}_{j+1} - \hat{\tau}_j) \) is introduced to explain the connection between the \( y_i \) and the control limits. Hereinafter, we refer this control chart as DMR chart.

5. Numerical Example: Hybrid Microcircuits Data

In this section, we will apply the proposed procedure to analyze a data taken from Camil and Ron (1998) for verification purposes. Raw materials used in manufacturing of hybrid microcircuits consist of components, dyes, pastes and ceramic substrates. The substrates plate undergo a process of printing and firing during which layers of conductors, dielectric, resistors and platinum or gold are added to the plates. Subsequent production steps consist of laser trimming, mounting and reflow soldering or chip bonding. The last manufacturing stage is the packaging and sealing of complete modules.

Five dimensions of substrate plates are considered here, with labels (a, b, c), (W, L). The first three are determined by the laser inscribing process. The last two are outer physical dimensions. These outer dimensions are measured on a different instrument than the first three, using a different sample substrates.

In this study, we will consider only the first dimension ‘a’ which measured the substrate plates. The Average Moving Range (AMR) chart and the proposed robust control charts for these individual length measurement of ‘a’ are presented in Fig. (1) and Fig. (2) respectively. The numerical results are performed using S-PLUS language.

The plot of Fig. (1) indicates that there are two signals, which are observations number 19 and 31. The two signals would suggest that the length measurements is not completely stable in the sample, which is in itself a valuable suggestion of assignable cause. It can be seen that the AMR chart detects the larger outliers. It does not, however, provide indications about the presence of the shifts. The standard deviation of the in-control process is estimated to be 2.125, which seems too large. The large estimate can be explained by the fact that it includes the additional variation that caused by the shifts and by the fact that the AMR chart is not robust. The AMR chart does not provide clear information about their number or the time points on which they occur. Furthermore, the detected shifts are not incorporated into the analysis, and, as a consequence, the chart is less sensitive in detecting the remaining assignable causes.

On the other hand, the robust control chart reveals the presence of several assignable causes, three isolated disturbances occurred which are indicated by observation number 19, 20 and 31 as well as one shifts in which it happen in between observation number 18 and 19. Observation number 22 which is located very close to border line of second pair of lower control
limit would suggest that the point might be a possible outlier. On the contrary, this point seems still within the control limit of the AMR chart. Removing the outliers and correcting the remaining measurements for the outliers, the process engineer is left with measurement that can be described considerably well by a normal distribution with standard deviation 1.962. This value is smaller than the the standard deviation of the AMR chart due to the fact that the effects of outlying observations have been downweighted.

Once the special cause has been identified, the necessary action can then be taken to rectify or improve the process. Fig. (2) gives important indications of their characteristic and the time points on which they occurred. Based on the chart, it also implies that the measurement length has shifted between observation 18 and observation 19.

In addition to these shifts, the length measurement appeared to be in statistical control, indicating that no assignable causes should be sought on the basis of these measurements.

It seems that the robust control chart is preferred as it provides the most revealing description of this dataset.

---

**Fig. 1: AMR Control Chart for Hybrid Microcircuits.**

![AMR Control Chart for Hybrid Microcircuits](image1)

**Fig. 2: Proposed Robust Control Chart for Hybrid Microcircuits.**

![Proposed Robust Control Chart for Hybrid Microcircuits](image2)
6. Conclusion

The classical Shewhart control charts for individual measurement are widely used in the monitoring processes. However, these charts require the assumption that the process variables are normally distributed and thus, are very sensitive to the presence of occasional outliers.

The robust control chart offers some significant advantages over existing control chart. This paper has presented a simple alternative robust univariate variable control chart which using Max-Type change point detection formulation for monitoring the process change in variance and mean. The robust control charting method efficiently monitor contaminated data processes and process shift. The proposed robust control chart is based on the median absolute deviation from the sample median (MAD) which is used for estimating the process standard deviation, $\sigma$. Traditional charts are not a preference for process monitoring where contamination may arise.

References:


