A NOVEL METHOD TO ASSESSING PROCESS VARIATION WITH A CONFIDENCE INTERVAL OF SAMPLE STANDARD DEVIATION

Boya Venkatesu[1], G. Siva[2], Christophe Chesneau[3*], V. Sai Sarada[4]

•

[1]School of Business, Woxsen University, Hyderabad, India.
[2]Department of Mathematics, VIT-AP University, Amaravati, India.
[3]LMNO, University of Caen-Normandie, Caen, France.
[4]Department of Research Labs, Asian Health Care Foundation, AIG Hospitals, India.
[1]venkateshboya50@gmail.com, [2]siva.g@vitap.ac.in, [3*]christophe.chesneau@gmail.com,
[4]saisaradav@gmail.com
Correspondence Email: christophe.chesneau@gmail.com

Abstract

One of the great tools of statistical process control for evaluating patterns of variation in the process is the control chart. In this article, we focus on developing a new approach to estimating process parameters in a control chart using confidence intervals (CIs) of sample standard deviations. They have the property of providing additional information about the quality of the estimate. A Monte Carlo simulation is used to investigate the accuracy of the obtained CIs. The performance of the classical and new approaches is compared in terms of standard error using simulated samples. The simulation results show that the new approach outperforms the classical approach.

Keywords: Control chart, point estimation, process variation, interval estimation, standard error.

I. Introduction

The control chart approach focuses on the estimation of process parameters. While the location parameter is estimated by the subgroup mean (average), there are various methods for estimating the process standard deviation. This problem is more important for non-Shewhart charts such as CUSUM or EWMA [7, 14]. In recent decades, there has been a growing interest in the use of better estimates than these classical methods [2]. In a sense, one class of estimates are robust estimates. Most classical estimation methods use point estimates, where a single value obtained from the sample data is used as an estimate. For various reasons, these estimates often do not represent the true parameters and hence the concept of interval estimation has emerged. Although Shewart's control chart is 8 decades old, researchers continue to identify gaps in both the theory and application of control charts. Some broad areas of research in this area can be such as economic design of control charts, control charts for dispersion, control charts for percentiles (median), control charts with asymmetric limits, mixed control charts and impact of parameter estimation on control charts [10]. Ronald Caulcutt [6] presented some statistical arguments about the design of control limits and suggested that standard methods should not be used without understanding the characteristics of the data. He argued that a delta chart should be used in

addition to x-bar and R-charts, where the delta chart takes into account the differences in the group means and thus operates on the medium-term variability of the process. Boyles [3, 4] addressed the problem of estimating the variation of a run chart (a control chart for individual values) using dynamic linear models (DLM) Estimation by using the Integrated Moving Average (1,1) model. This method was an improvement over the conventional moving range (MR-bar) method. He also found that the new estimate was robust. John Braun and Park [5] addressed the issue of calculating the sample standard deviation from individual data in the context of out-of-control and non-normal situations. They have proposed a method of pre- and post-control for outliers with control limits based on the median absolute deviation (MAD). This procedure shows better performance than the Boyles method. Tung-Lung Wu [15] has come up with the design of a distribution-free run-based control chart to detect location shifts (shift in process mean). This approach used the concept of "conditional longest run statistic" in addition to ARL. It is a truly distribution-free method and no historical data is required as no parameters are estimated.

In statistical process control (SPC), the control chart is an essential tool for monitoring the consistency and variability of a process over time [8, 9, 11]. To construct a control chart, the process variability must be quantified, which is typically done using either the range (R) or the standard deviation (σ) of the sample data. Boya et al. [1] developed a novel approach to estimate the process variation using the confidence interval of the sample range. Based on this previous work, the authors realized that the use of range to estimate control limits often involves the use of approximations or empirical rules (such as the average range chart), which may not be as accurate. Therefore, the focus of this study is to estimate process variation using confidence interval estimates of the sample standard deviation (σ), which is derived from all data points and provides a more accurate and consistent measure of process variation. The remainder of the article is organized as follows: In Sections 2 and 3, we consider interval estimates in the context of process mean and standard deviation. In Section 3, we develop a new approach to estimating process variation using interval estimation. In Section 4, we discuss the step-by-step procedure of the new method, and Section 5 consists of the simulation study.

II. Interval estimates of process mean in $N(\mu, \sigma^2)$

In order to describe our approach, it is necessary to present a precise mathematical framework. Let X be a measure of interest and θ be an associated parameter. We suppose that he has a point estimate $\hat{\theta}(x)$, where x is a random sample. If we replace $\hat{\theta}(x)$ by a statistic, which is a subset of the parameter space, then we get an interval estimate. The general format of an interval estimate is $\hat{C}(x) = \left[\hat{\theta}(x) - \delta(x), \hat{\theta}(x) + \delta(x)\right]$, where $\hat{\theta}(x)$ is the point estimate and $\delta(x)$ is known as the margin of error, such that $P_\theta\left\{\theta \in \hat{C}(x)\right\} = \gamma$, where $0 < \gamma < 1$ is called the level of significance, and the interval is called level of confidence (usually 95%) and $\hat{C}(x)$ is called Confidence Interval (CI).

In the case of normality, when the main measure can be modeled by a random variable, say $X \sim N(\mu, \sigma^2)$, we can find the $\frac{\alpha}{2}$ CI for the mean by using normal approximation. More precisely, let $Z \sim N(0,1)$ and Z_α be a value on Z, such that

$$P(|Z| > Z_\alpha) = \alpha. \tag{1}$$

Suppose μ_o denotes the true mean (unknown) and let $\bar{x}$ be the point estimate of μ. It follows that the condition $Z = |\frac{\bar{x}-\mu_0}{\sigma_0}| < Z_\alpha$ leads to the $\bar{x} - Z_\alpha \frac{\sigma_0}{\sqrt{n}} < \mu_0$ and $\bar{x} + Z_\alpha \frac{\sigma_0}{\sqrt{n}} > \mu_0$

If we take $\gamma = 1 - 2\alpha$ then $\alpha = \frac{1-\gamma}{2}$ so that the interval $\left[\bar{x} - Z_{(1-\gamma)/2}\frac{\sigma_0}{\sqrt{n}}, \bar{x} + Z_{(1-\gamma)/2}\frac{\sigma_0}{\sqrt{n}}\right]$ has a confidence level of γ. Hence the $100(1-\alpha)\%$ CI for the mean is given by

$$\left[\bar{x} - Z_{(1-\gamma)/2}, \bar{x} + Z_{(1-\gamma)/2}\right]. \tag{2}$$

The above CI is known as Z-interval because the margin of error is based on normal distribution. We can also construct the margin of error using student's t-distribution. Then the $100(1-\alpha)\%$ CI

for σ from $X \sim N(\mu, \sigma^2)$ is based on t- distribution, and it is given by

$$\left[\bar{x} - t_{(n-1),(1-\gamma)/2}\frac{s}{\sqrt{n}}, \bar{x} + t_{(n-1),(1-\gamma)/2}\frac{s}{\sqrt{n}}\right], \tag{3}$$

where the sample standard deviation s is an estimate of σ.

III. Interval estimates of process σ in $N(\mu, \sigma^2)$

In the context of control charts, the estimation of σ is done in two ways viz, a) Using sample range (R) b) Using sample standard deviation (s). With the method of ranges, we have: $\hat{\sigma}_R = \frac{\bar{R}}{d_{2,n}}$, where $\bar{R}$ is the average of ranges over m samples and $d_{2,n}$ is a scaling constant. Let us define $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ as the lower and upper $\frac{\alpha}{2}$ percentile points on the Chi-square distribution. Then the $100(1-\alpha)\%$ CI for σ from $X \sim N(\mu, \sigma^2)$ is based on the χ^2 distribution, and it is given by

$$\left[\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}}, \sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2}}}\right], \tag{4}$$

where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ as the lower and upper $\frac{\alpha}{2}$ percentile points on the Chi-square distribution. The $100(1-\alpha)\%$ CI for σ based on R is given by

$$\left[\frac{R}{d_{2,n}}G_1, \frac{R}{d_{2,n}}G_2\right], \tag{5}$$

where G_1, G_2 are constants and R denotes the sample range, and $d_{2,n}$ is a scaling constant. In this article, our focus is to study the method of triplet estimation of process variation (σ) when the quality characteristic follows the normal distribution $N(\mu, \sigma^2)$. Instead of using $\frac{\bar{R}}{d_{2,n}}$ as an estimate, we consider the $100(1-\alpha)\%$ CI of the sampling distribution of sample variance and thereby derive a new estimate as a linear combination of the triplet estimate.
Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from the process. From this sample process mean μ is estimated as

$$\bar{x}_i = \frac{1}{n}\sum_{j=1}^{n} x_{ij},$$

where x_{ij} denote the j-th observation in the i-th sample ($i = 1, 2, \ldots, k$; $j = 1, 2, \ldots, n$) and $\bar{x}_i$ is the sample mean of the i-th subgroup.
The corresponding sample standard deviation is denoted by

$$s_i = \sqrt{\frac{1}{n-1}\sum_{j=1}^{n}(x_{ij} - \bar{x}_i)^2},$$

which are the point estimates of the respective parameters, here s_i denotes i-th subgroup standard deviation. The overall (grand) mean of all the sample means is represented as $\bar{x} = \frac{1}{k}\sum_{i=1}^{k}\bar{x}_i$ and

$$s = \sqrt{\frac{1}{nk-1}\sum_{i=1}^{k}\sum_{j=1}^{n}(x_{ij} - \bar{x})^2}.$$

It is important to note that s is not an unbiased estimate of σ. In fact, s estimates the quantity $C_4\sigma$, where

$$C_4 = \sqrt{\frac{2}{n-1}}\left(\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\right)\sigma$$

is a constant whose values are tabulated [12, 13]. As such an unbiased estimate of σ is $\frac{s}{C_4}$.
For normally distributed data, the sampling distribution of the statistic $\frac{ns^2}{\sigma^2}$ follows the Chi-square distribution with n degrees of freedom. Then the $100(1-\alpha)\%$ CI for σ^2 is given by

$$\left[\frac{(n-1)}{\chi^2_{\alpha/2}}s^2, \frac{(n-1)}{\chi^2_{1-\alpha/2}}s^2\right].$$

The CI for σ is then defined as

$$\left[\sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}}\left(\frac{s}{c_4}\right), \sqrt{\frac{(n-1)}{\chi^2_{1-\alpha/2}}}\left(\frac{s}{c_4}\right)\right].$$

Hence, the triplet estimates of σ based on s can be written as $\eta = \{\eta_1, \eta_2, \eta_3\}$, where

$$\eta_1 = \sqrt{\frac{(n-1)}{\chi^2_{\alpha/2}}}\left(\frac{s}{c_4}\right), \quad \eta_2 = \frac{s}{c_4}, \quad \text{and} \quad \eta_3 = \sqrt{\frac{(n-1)}{\chi^2_{(1-\alpha)/2}}}\left(\frac{s}{c_4}\right). \tag{6}$$

It can be seen from Boya et al. [1] that the triplet estimates of σ based on R was stated as $\tau = [\theta_1, \theta_2, \theta_3]$, where $\theta_1 = \frac{R_i}{d_{2,n}G_1}, \theta_2 = \frac{R_i}{d_{2,n}}$, and $\theta_3 = \frac{R_i}{d_{2,n}G_2}$, with $d_{2,n}$, G_1 and G_2 being constants for a fixed sample size.

IV. The new estimate of σ using sample standard deviation

The new estimate of σ is a linear combination of the triplet components given in (6)

$$\hat{\sigma}_{s,CI} = l_1\eta_1 + l_2\eta_2 + l_3\eta_3, \tag{7}$$

such that $l_i \geq 0$ for $i = 1, 2, 3$ and $\sum_{i=1}^{3} l_i = 1$.
The subscript CI indicates that (6) is based on the CI.
Let us assume that the process is set up to have a certain variation σ' known either by hypothesis or by engineering specifications. Then the weights will be taken as

$$w_i = |\eta_i - \sigma'|^{-1}, \tag{8}$$

so that

$$l_i = \frac{w_i}{\sum_{i=1}^{3} w_i}, \quad \text{for } i = 1, 2, 3. \tag{9}$$

As the magnitude of the error rises up, the weights go downward. The outcome of this method $\hat{\sigma}_{s,CI}$ for each of the m samples, and this data can be used to study the empirical properties of the new estimate.
The stepwise method used in Boya et al. [1] for range-based estimation is now extended to the case of s-based estimation to arrive at the new estimate of σ.

1. Generate m random samples (subgroups), each of size n, from $N(\mu, \sigma^2)$.
2. Evaluate $\bar{x}$ and s for each subgroup.
3. Calculate η_1, η_2, η_3 of each subgroup using (6).
4. Calculate w_1, w_2, w_3 using (8).
5. Calculate l_1, l_2, l_3 using (9).
6. Find the new estimator $\hat{\sigma}_{s,CI}$ using (7).

This procedure leads to $\hat{\sigma}_{s,CI}$ for each of the m samples, and this value can be used to study the empirical properties of the new estimate.
In the subsequent section, a hypothetical exercise is presented to demonstrate the application of the new method.

V. Simulation study

We have conducted a study $M = 5000$ random samples of different sizes ($n = 10$, 15 and 20) are generated from the normal distribution $N(\mu, \sigma^2)$ with $\mu = 10.5$ and $\sigma = 1.5$. The summary of results is shown in Table 1.

Table 1: *Comparison of η_2 and $\hat{\sigma}_{s,CI}$ using triplet estimate*

n	M	η_2 (Mean $\pm$ S.E)	$\hat{\sigma}_{s,CI}$ (Mean $\pm$ S.E)
10	5000	1.4703 $\pm$ 0.0051	1.4907 $\pm$ 0.0034
15	5000	1.4932 $\pm$ 0.0050	1.4852 $\pm$ 0.0032
20	5000	1.4835 $\pm$ 0.0048	1.4802 $\pm$ 0.0030

From this table it can be seen that, regardless of the subgroup size n, the new estimate by the triplet method is closer to the hypothetical process variation of $\sigma = 1.5$. Furthermore, the standard error is much smaller than that of the classical estimate. Therefore, we propose that the new estimate of σ based on the CI of sample standard deviations gives a stable estimate with low bias compared to $\frac{s}{c_4}$.
The pattern of variation of in the classical and the new estimates are shown in Figure 1. We mention that the sample standard error from 5000 samples with $n = 10, 15$ and 20 are computed to assess the consistency of the new estimate.

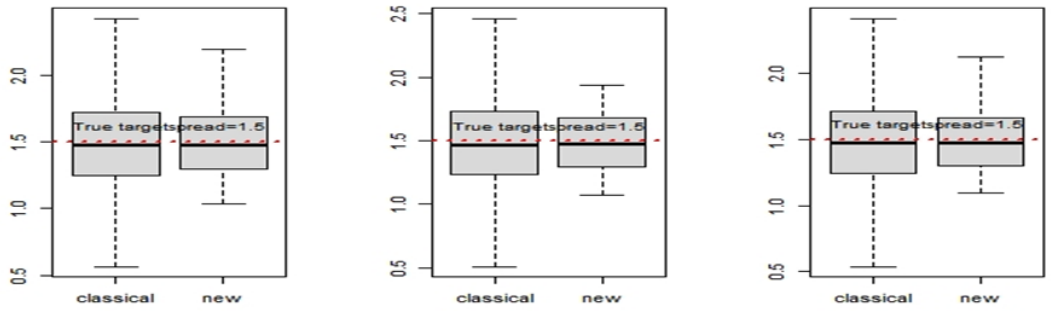

Figure 1: *The pattern of variation of in the classical and the new estimates*

From this figure, it can be seen that $\hat{\sigma}_{s,CI}$ of σ has lower variation around the center when compared to η_2.

VI. Conclusions

This paper introduces a robust and innovative method for estimating process variation using the CI of the sample standard deviation, aimed at improving the accuracy and stability of control charts in SPC. Traditional approaches typically rely on point estimates such as s or $\frac{s}{c_4}$, which often suffer from bias and variability, especially with small sample sizes. To address these limitations, the proposed method constructs a triplet estimator η_1, η_2, η_3 derived from the CI of the sample standard deviation. These components are then combined into a single weighted estimator $\hat{\sigma}_{s,CI}$ using an inverse-error weighting scheme, making the final estimate both data-driven and adaptively robust.

A comprehensive Monte Carlo simulation study with 5000 replications across various sample sizes ($n = 10, 15, 20$) was conducted to evaluate the performance of the proposed method. The results showed that the new estimator closely approximated the true process standard deviation

(σ = 1.5) while achieving a significantly lower standard error compared to the classical point estimator s or $\frac{s}{c_4}$. Visual analysis of the estimator's distribution further confirmed reduced variability and improved stability.

Overall, this interval-based approach is more precise and reliable than existing SPC tools, making it a valuable enhancement. It provides a statistically robust alternative to classical estimators and is particularly beneficial when working with moderate or small sample sizes. The method is flexible, simple to use, and based on solid statistical theory. This makes it very useful in quality control and industrial settings. Future research could involve extending this method to non-normal data distributions or integrating it with modern adaptive control chart frameworks.

References

[1] Boya Venkatesu, Abbaiah, R., and Sai Sarada, V. (2018). A New Method of Estimating the Process Variation Using Confidence Interval of Sample Range. *RESEARCH REVIEW International Journal of Multidisciplinary*, 3(08):204–207.

[2] Boya Venkatesu, Abbaiah, R., and Sai Sarada, V. (2019). On the Estimation of Operating Characteristic of Shewart Control Chart for Means Using Interval Estimates of Process Mean and Variation. *International Journal of Advanced Scientific Research and Management (IJASRM)*, 4(04):62–66.

[3] Boyles, R. A. (1979). Estimating Common-cause σ in the Presence of Special Causes. *Journal of Quality Technology*, 29:381–395.

[4] Boyles, R. A. (1997). Estimating Common-cause Sigma in the Presence of Special Causes. *Journal of Quality Technology*, 29(4):381–395.

[5] Braun, W. J. and Park, D. (2008). Estimaion of σ for Individuals Charts. *Journal of Quality Technology*, 40(3):332-344.

[6] Caulcutt, R. (1995). The Rights and Wrongs of Control Charts. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(3):279–288.

[7] Chakraborti, S. and Van de Wiel, M. A. (2008). A Nonparametric Control Chart Based on the Mann-Whitney Statistic. *Collections*, 156–172.

[8] Duncan, A. J. (1986). *Quality Control and Industrial Statistics* (5th ed.). Irwin, Homewood.

[9] Grant, E. L. (1964). *Statistical Quality Control* (3rd ed.). McGraw-Hill Book Company.

[10] Guo, B. and Wang, B. X. (2016). Control Charts for the Coefficient of Variation. *Statistical Papers*, 59(3):933–955.

[11] Mittag, H. J. and Rinne, H. (1993). *Statistical Methods of Quality Assurance*. Chapman and Hall, CRC Press.

[12] Montgomery, D. C. (2008). *Introduction to Statistical Quality Control* (4th ed.). John Wiley & Sons.

[13] Montgomery, D. C. and Runger, G. C. (2002). *Applied Statistics and Probability for Engineers* (3rd ed.). John Wiley & Sons.

[14] Park, C. and Reynolds, M. R. (1987). Nonparametric Procedures for Monitoring a Location Parameter Based on Linear Placement Statistics. *Sequential Analysis*, 6(4).

[15] Tung - Lung Wu. (2018). Distribution-Free Runs-Based Control Charts. *Journal of Computational Statistics and Data Analysis*, 1–14.