Bootstrap control limits of process control charts for correlative process data

Hideo Suzuki

Institute of Policy and Planning Sciences
University of Tsukuba, Japan

Abstract

This research explores the application of the bootstrap methods to the construction of control limits for the x charts and the EWMA charts based on single observations with stationary autoregressive processes. The subsample means-based control chars in the presence autocorrelation are also considered. We use a technique for inferring confidence intervals using bootstrap, the percentile method. Simulation studies are conducted to compare the performance of the bootstrap method and that of standard method for constructing control charts under several conditions.

1. Introduction

Statistical process control (SPC) techniques are widely used in industry for process monitoring and quality improvement. Various statistical control charts have been developed to monitor the process mean and variance. Traditional SPC charts, including the Shewhart, the exponential weighted moving average (EWMA) and the cumulative sum (CUSUM) charts, are based on a fundamental assumption that process data are statistically independent and normally distributed when the process are in control. In practice, however, process data are not always independent from each other; in continuous process industries such as the chemical industry, most process data are autocorrelated. In such conditions, the traditional SPC methods may not be appropriate for monitoring, controlling and improving process quality. Various authors have considered the effect of data correlation on control charts (Alwan and Radson (1992), Montegomery and Mastrangelo(1991), Wardell, Moskowitz and Plante(1992), and Zhang(1998)).

The bootstrap is a powerful computer-based method, and is useful if complicated models seem necessary, since mathematical complication is no impediment to a bootstrap analysis of accuracy. The concept of bootstrapping was theoretically by Efron (1979 and 1987). Some authors discussed the use of bootstrap with quality control (Franklin and Wasserman (1992), and Seppla, Moskowitz, Plante and Tang (1995)). Seppla, Moskowitz, Plante and Tang (1995) proposed a bootstrap approach for assessing process control limits of $\bar{x}$ and s^2 charts. The bootstrap may also be useful in designing control charts for correlated process data. This research explores the application of the bootstrap methods to the construction of control limits for the x charts and the EWMA charts based on single observation with stationary autoregressive processes. The subsample means-based control ($\bar{x}$) chars in the presence autocorrelation are also considered. We use a technique for inferring confidence intervals using bootstrap; the percentile method, proposed by Efron (1987). Simulation studies are conducted to compare the performance of the bootstrap method and that of the standard methods for designing control charts under several conditions.

2. Process Models and Standard Control Charts

2.1 Models

In this study, we assume that processes follow a first-order stationary autoregressive model, AR(1). That is,

$$x_t = (1 - \phi_1)\mu + \phi_1 x_{t-1} + \epsilon_t \tag{1}$$

where x_t is the process data at time t, μ is the mean of the process, ϕ_1 is the autoregressive parameter, $|\phi_1| \leq 1$, and ϵ_t are independent and identically distributed normal random variables with mean 0 and variance σ_ϵ^2. Without loss of generality, we assume $\mu = 0$ and $\sigma_\epsilon = 1$, and thus $E(x_t) = 0$ and $V(x_t) = 1/(1 - \phi_1^2)$.

2.2 Standard methods

x charts

The x charts are merely a plot of the each single process data. In practice, parameters for the process mean and variance are usually not known and must be estimated. Such estimates are usually based on past samples taken when the process assumed to be in-control. Therefore, the upper and lower control limits of x chart for the stationary process, which are similar to the traditional x chart for an independent process, are given by

$$\hat{\mu} \pm L_x \hat{\sigma}_x, \tag{2}$$

where $\hat{\mu}$ and $\hat{\sigma}_x$ are an estimate of the process mean and of the process standard deviation in equation (1) respectively. L_x is usually equal to 3, which means that the significance level of the three sigma control limits is supposed to be 0.27%.

EWMA charts

The exponentially weighted moving average (EWMA) charts are designed to detect small shifts in the mean more quickly than the x charts by giving exponential weight to past data. For the EWMA charts, the static that is plotted can be thought as a forcast which is a weighted sum of the current data and previous periods' forecast. If we call the forecast at time t z_t, then we can write

$$z_t = \lambda x_t + (1 - \lambda) z_t, \tag{3}$$

where λ $(0 \leq \lambda \leq 1)$ is a smoothing parameter which determines the weight given to past data. When λ is large, relatively little weight is given to older data. As λ becomes smaller, more weight is given to the older weight. The upper and lower control limits of the EWMA chart are

$$\hat{\mu} \pm L_z \hat{\sigma}_x \sqrt{\frac{\lambda}{2 - \lambda}}, \tag{4}$$

where L_z is a multiplier usually assumed to be 3.

$\bar{x}$ charts

Most studies for autocorrelation on SPC methods have been confined to control charts based on single observations. We study the property of the subsample-based control charts, the $\bar{x}$ charts, in the presence of autocorrelation. In the case we consider, sequential samples are periodically drawn from the process and sample means are plotted through time on the $\bar{x}$ chart. The upper and lower control limits for the $\bar{x}$ chart are given by

$$\hat{\mu} \pm L_{\bar{x}} \frac{\hat{\sigma}_x}{\sqrt{n}}, \tag{5}$$

where $L_{\bar{x}}$ is also usually equal to 3.

3. Bootstrap Control limits

3.1 Bootstrap

We use a model-based bootstrapping in the time series analysis. The idea is to fit a suitable process model to the process data, to construct residual from the fitted model, and then to generate new series by incorporating random samples from residual into the fitted model. Now, we have the observed process data, $x_1, x_2, \ldots, x_N$, which are assumed to be in-control and follow the AR(1) model.

The balanced bootstrap algorithm for assessing process control limits with single process data is given as follows:

1. Estimate the autoregressive parameter ϕ_1 by solving the Yule Walker equation.
2. Calculate original residuals $e_i = x_i - (1 - \hat{\phi}_1)\hat{\mu} - \hat{\phi}_1 x_{i-1}$, $i = 2, \ldots, N$, where $\hat{\phi}_1$ is an estimate of the autoregressive parameter, $\hat{\mu}$ is the sample mean.
3. Obtain B by finding an integer, A, such that $B = A * (N-1) \geq 2000$ (minimum number of resembles required to obtain accurate percentile estimates).
4. Replicate the original $N - 1$ residuals A times for a total of B ressiduals (note that balance is achieved since each residual occurs A times):

$$\underbrace{(e_2, \ldots, e_N), \ldots, (e_2, \ldots, e_N)}_{\text{A times}}.$$

5. Permute the B residuals randomly, then $e_1^*, \ldots, e_B^*$.
6. The bootstrap samples x_i^*, $b = 1, \cdots, B$ are obatin by

$$\begin{aligned} x_1^* &= z_1^* = x_1, \\ x_i^* &= (1 - \hat{\phi}_1)\hat{\mu} + \hat{\phi}_1 x_{i-1}^* + e_i^*, \\ z_i^* &= \lambda x_i^* + (1 - \lambda) z_i^*, \qquad i = 1, \ldots, B. \end{aligned} \tag{6}$$

3.2 Percentile control limits for single process data

Let $\hat{G}_{(x,B)}$ be the cumulative distribution function of x^*, that is,

$$\hat{G}_{(x^*,B)}(t) = \frac{1}{B} \sum_{i=1}^{B} I\{x_i^* \leq t\},$$

where I is an indicator function. The $\hat{G}^{-1}_{(x^*,B)}(\alpha)$ is the 100·αth empirical percentile of the x_i^* values, that is, the $B \cdot \alpha$th value in the ordered list of the B replications of x^*. We can write the $(1-\alpha)$ percentile control limits of x chart as

$$\text{UCL} = \hat{G}^{-1}_{(x^*,B)}(1-\alpha/2), \tag{7}$$

$$\text{LCL} = \hat{G}^{-1}_{(x^*,B)}(\alpha/2), \tag{8}$$

where the α of the traditional three sigma control limits is 0.9974. If $B \cdot \alpha/2$ is not an integer, we use the following procedure; assuming that $\alpha/2 \leq 0.5$, let $y = [(B+1)\alpha/2]$, which is the largest integer $\leq (B+1)\alpha/2$, then the empirical $\alpha/2$ and $1-\alpha/2$ quantiles are defined by yth largest and $(B+1-y)$th largest values of x_i^*, respectively.

Likewise the $1-\alpha$ percentile control limits of EWMA chart is given by

$$\text{UCL} = \hat{G}^{-1}_{(z^*,B)}(1-\alpha/2), \tag{9}$$

$$\text{LCL} = \hat{G}^{-1}_{(z^*,B)}(\alpha/2), \tag{10}$$

where the $\hat{G}^{-1}_{(z^*,B)}(\alpha)$ is the 100·αth empirical percentile of the z_i^* values.

3.3 Percentile control limits for subgroup mean charts

Suppose n sequential observations of the process are taken every n periods, as follows:

$$\underbrace{x_1, x_2, \ldots, x_n}_{\text{Subgroup } 1}, \underbrace{x_{n+1}, x_{n+2}, \ldots, x_{2n}}_{\text{Subgroup } 2}, \;\ldots\ldots,\; \underbrace{x_{(k-1)n+1}, x_{(k-1)n+2}, \ldots, x_N}_{\text{Subgroup } k},$$

where each x_i follows the stationary autoregressive process, as shown in equation (1), and $kn = N$. The subsample mean m_j is simply defined as a sum or aggregation of n successive x_i subsequently divided by the constant n. The jth subsample mean m_j is gven by

$$m_j = \frac{1}{n}(x_{(j-1)n+1} + x_{(j-1)n+2} + \cdots + x_{jn}), \;\; j = 1, \ldots, k. \tag{11}$$

Then, the bootstrap procedure for assessing control limits from a series of k subgroup samples of size n, such that $N = nk$, is described as follows:

1. Estimate the autoregressive parameter, calculate the original residuals and obtain B. (This step is the same as Step 1, Step 2 and Step 3 in the above bootstrap procedures for control limits with single process data.)
2. Replicate the original $N-1$ residuals An times for a total of Bn residuals (note that balance is achieved since each residual occurs An times):

$$\underbrace{(e_2, \ldots, e_N), \ldots, (e_2, \ldots, e_N)}_{An \text{ times}}.$$

3. Permute the Bn residuals randomly, then $e_1^*, \ldots, e_{Bn}^*$.
4. The bootstrap samples x_i^*, $b = 1, \cdots, B$ are obatin by

$$\begin{aligned} x_1^* &= z_1^* = x_1, \\ x_i^* &= (1-\hat{\phi}_1)\hat{\mu} + \hat{\phi}_1 x_{i-1}^* + e_i^*, \quad i = 1, \ldots, Bn. \end{aligned} \tag{12}$$

5. Compute the bootstrap subgroup sample menas m_j^*, that is,

$$m_j^* = \frac{1}{n}(x^*_{(j-1)n+1} + \cdots + x^*_{jn}), \quad j = 1, \ldots, B, \tag{13}$$

$$\underbrace{x_1^*, x_2^*, \ldots, x_n^*}_{\text{Subgroup 1}}, \underbrace{x_{n+1}^*, x_{n+2}^*, \ldots, x_{2n}}_{\text{Subgroup 2}}, \ldots\ldots, \underbrace{x^*_{(B-1)n+1}, x^*_{(B-1)n+2}, \ldots, x^*_{Bn}}_{\text{Subgroup } B}.$$

Finally, we can obtain the $1-\alpha$ percentile control limits of subgroup samples menan charts,

$$\text{UCL} = \hat{G}^{-1}_{(m^*,B)}(1-\alpha/2), \tag{14}$$

$$\text{LCL} = \hat{G}^{-1}_{(m^*,B)}(\alpha/2), \tag{15}$$

where the $\hat{G}^{-1}_{(m^*,B)}(\alpha)$ is the $100 \cdot \alpha$th empirical percentile of the m_j^* values.

4. Comparison of Control Limits for the Standard methods and Control the Bootstrap Control Method

For various AR(1) processes, the control limits estimates for the standard methods and the bootstrap method are compared. Table 1 shows the mean of 1000 estimated upper and lower control limits for the x charts and the EWMA charts, and the figures in parentheses for this table represent the bias, the difference between the estimated mean of the control limits and the true values. The number of original samples N is 200, the significance level is 99.74% control limits (0.13% for lower and 99.87% for upper), and the three sigma control limits are used for the standard methods. The parameter of the EWMA charts are $\lambda = 0.1$ and 0.3. Table 2 shows the results for the subgroup mean charts, in which the sample size is $n = 5$, and the number of subgroups is $k = 20$.

Reference

1. Alwan, L.C., Radson, D., "Time-series Investigation of Subsample Mean Charts," *IEE Transaction*, vol.24, 66-80, 1992.
2. Efron, B., "Bootstrap Methods: Another Look at the Jakknife," *Annals of Statistics*, vol.7, 1-26, 1979.
3. Efron, B., "Better Bootstrap Confidence Intervals," *Journal of the American Statistical Association*, Vol.82, 171-185, 1987.
4. Franklin, L.A. and Wasseerman, G.S. (1992), "Bootstrap Lower Confidence Limits for Capability Indices," *Journal of Quality Technology*, vol.24, 196-210.
5. Montgomery, D.C., and Mastrango, C.M., "Some Statistical Process Control Methods for Autocorrelated data," *Journal of Quality Technology*, vol.23, 179-193, 1991.
6. Seppala, T., Moskowitz, H, Plante, R. and Tang, J (1995), "Statistical Process Control via the Subgroup Bootstrap," *Journal of Quality Technology*, vol.27, 139-153, 1995.
7. Wardell, D.G., Moskowitz, H., and Plante, R., "Control Charts in the Presence of Data Correlation," *Management Science*, vol.38, 1084-1105, 1992.
8. Zhang, N.F., "A statistical Control Chart for Stationary Process data," *Technometrics*, vol.40, 24-38, 1998.

Table 1. Comparison of control limits for x control charts and the EWMA charts in the case of AR(1) processes (the number of original samples $N = 200$)

(a) x charts

ϕ	9.87 Percentile			0.13 Percentile		
	Standard	Bootstrap	True	Standard	Bootstrap	True
.25	3.09 (0.00)	3.04 (-0.05)	3.09	-3.07 (0.02)	-3.03 (0.06)	-3.09
.50	3.43 (-0.03)	3.52 (0.06)	3.46	-3.44 (0.02)	-3.55 (-0.09)	-3.46
.75	4.42 (-0.12)	4.58 (0.04)	4.54	-4.51 (-0.03)	-4.60 (-0.06)	-4.54
.95	8.36 (-1.25)	8.13 (-1.48)	9.61	-8.54 (1.06)	-8.33 (1.27)	-9.61
-.25	3.10 (0.01)	3.08 (-0.01)	3.09	-3.10 (-0.01)	-3.08 (0.01)	-3.09
-.50	3.47 (-0.01)	3.58 (0.12)	3.46	-3.47 (-0.01)	-3.58 (-0.12)	-3.46
-.75	4.51 (-0.03)	4.73 (0.19)	4.54	-4.51 (0.03)	-4.72 (-0.18)	-4.54
-.95	9.28 (-0.33)	9.33 (-0.28)	9.61	-9.29 (0.32)	-9.34 (0.27)	-9.61

(b) EWMA charts ($\lambda = 0.1$)

ϕ	9.87 Percentile			0.13 Percentile		
	Standard	Bootstrap	True	Standard	Bootstrap	True
.25	0.71 (-0.19)	1.00 (0.10)	0.90	-0.71 (0.19)	-1.00 (-0.10)	-0.90
.50	0.79 (-0.49)	1.32 (0.04)	1.28	-0.79 (0.50)	-1.32 (-0.03)	-1.29
.75	1.01 (-1.35)	2.22 (-0.14)	2.36	-1.02 (1.34)	-2.21 (0.15)	-2.36
.95	1.90 (-6.00)	5.65 (-2.25)	7.90	-1.94 (5.96)	-5.71 (2.19)	-7.90
-.25	0.71 (0.14)	0.73 (0.16)	0.57	-0.71 (-0.14)	-0.71 (-0.14)	-0.57
-.50	0.79 (0.30)	0.65 (0.16)	0.49	-0.80 (-0.31)	-0.65 (-0.16)	-0.49
-.75	1.04 (0.58)	0.63 (0.17)	0.46	-1.03 (-0.57)	-0.63 (-0.17)	-0.46
-.95	2.09 (1.48)	0.73 (0.12)	0.61	-2.10 (-1.48)	-0.74 (-0.12)	-0.62

(c) EWMA charts ($\lambda = 0.3$)

ϕ	9.87 Percentile			0.13 Percentile		
	Standard	Bootstrap	True	Standard	Bootstrap	True
.25	1.30 (-0.25)	1.62 (0.07)	1.55	-1.29 (0.27)	-1.62 (-0.06)	-1.56
.50	1.45 (-0.67)	2.16 (0.04)	2.12	-1.44 (0.67)	-2.15 (-0.04)	-2.11
.75	1.87 (-1.53)	3.39 (-0.01)	3.40	-1.86 (1.55)	-3.35 (-0.44)	-3.41
.95	3.63 (-5.41)	7.31 (-1.73)	9.04	-3.41 (5.64)	-7.05 (2.00)	-9.05
-.25	1.30 (0.21)	1.15 (0.06)	1.09	-1.30 (-0.21)	-1.15 (-0.06)	-1.09
-.50	1.46 (0.45)	1.06 (0.05)	1.01	-1.45 (-0.36)	-1.06 (0.03)	-1.01
-.75	1.88 (0.81)	1.11 (0.04)	1.07	-1.88 (-0.81)	-1.11 (-0.04)	-1.07
-.95	3.88 (2.07)	1.78 (-0.03)	1.81	-3.88 (-2.06)	-1.78 (0.04)	-1.82

Table 2. Comparison of control limits for subsample mean control charts with $(n = 5, k = 20)$ in the case of AR(1) processes

ϕ	99.87 Percentile			0.13 Percentile		
	Standard	Bootstrap	True	Standard	Bootstrap	True
.25	1.37 (-0.03)	1.69 (0.29)	1.40	-1.38 (0.02)	-1.71 (-0.31)	-1.40
.50	1.52 (-0.39)	2.30 (0.39)	1.91	-1.52 (0.39)	-2.30 (-0.39)	-1.91
.75	1.95 (0.55)	3.59 (0.52)	3.07	-1.93 (1.15)	-3.55 (-0.47)	-3.08
.95	3.34 (-4.31)	6.71 (-0.94)	7.65	-3.29 (4.33)	-6.70 (0.92)	-7.62
-.25	1.39 (0.46)	1.18 (0.25)	0.93	-1.38 (-0.38)	-1.17 (-0.24)	-0.93
-.50	1.55 (0.71)	1.05 (0.21)	0.84	-1.54 (-0.70)	-1.04 (-0.20)	-0.84
-.75	2.01 (1.15)	1.08 (0.22)	0.86	-2.01 (-1.15)	-1.08 (-0.22)	-0.86
-.95	3.96 (1.07)	1.80 (-1.09)	1.61	-3.97 (-2.36)	-1.81 (-0.20)	-1.61