# 行政院國家科學委員會專題研究計畫　成果報告

## 在小子群時的一個新的管制圖

計畫類別：個別型計畫
計畫編號：NSC92-2118-M-032-009-
執行期間：92年08月01日至93年07月31日
執行單位：淡江大學統計學系(所)

計畫主持人：蔡宗儒

計畫參與人員：王曉齡,林宏嘉

報告類型：精簡報告

處理方式：本計畫可公開查詢

中　華　民　國 93年9月17日

# An Alternative Control Chart Approach Based on Small Number of Subgroups

Tzong-Ru Tsai
Department of Statistics, Tamkang University, Tamsui, Taipei, Taiwan

## Abstract

This paper presents an approach for constructing control limits that can enable the user to begin monitoring the process means at an earlier stage than the standard approaches. The proposed control limits can be constructed easily and are close to any specific percentile of run length distribution of the true limits, even when only a few initial subgroups are available. Performances of the proposed approach are studied by Monte Carlo simulation. The simulation results show that the proposed control limits perform similarly to the true limits even when the limits are estimated using data from only a few initial subgroups.

**Keywords:** $\overline{X}$ Control Chart, Prospective Control Limits, False Alarm Probability, Run Length Distribution.

## 1 Background and Motivation

Suppose that a quality characteristic is normally distributed with mean $\mu$ and standard deviation $\sigma$. When $\mu$ and $\sigma$ are both known, their values can be used to construct the center line and control limits of the Shewhart $\overline{X}$ chart. However, in practice, these parameters are usually unknown. Hence, the $\overline{X}$ control chart can be set up by collecting $m$ initial subgroups of size $n$ and using these values to obtain the estimates of $\mu$ and $\sigma$.

Let $X_{ij}$ be the $j$th observation in the $i$th subgroup, $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n$. Let $\overline{X}_i$ and $S_i$ denote the sample mean and sample standard deviation, respectively, of the $i$th subgroup. The usual approach for constructing the $\overline{X}$ control chart is then to set the center line (CL) at $\overline{\overline{X}}$ and the upper control limit (UCL) and lower control limit (LCL) at

$$\begin{aligned} UCL &= \overline{\overline{X}} + 3\frac{\overline{S}}{c_4\sqrt{n}} \\ LCL &= \overline{\overline{X}} - 3\frac{\overline{S}}{c_4\sqrt{n}}, \end{aligned} \tag{1}$$

where $\overline{\overline{X}} = \frac{1}{m}\sum_{i=1}^{m}\overline{X}_i = \frac{1}{m}\sum_{i=1}^{m}\sum_{j=1}^{n}X_{ij}$,

$\overline{S} = \frac{1}{m}\sum_{i=1}^{m}S_i$, and $c_4$ is a constant that depends only on the subgroup size $n$. Tables of $c_4$ are given in most quality control books such as Montgomery (2001, p.761) and Ryan (2000, p.540).

The $\overline{X}$ control chart constructed using the $m$ initial subgroups is first used to test retrospectively whether the process was in control. When the in-control process is established, the control chart is used to monitor the process continuously and the subgroup mean are plotted on the chart one-at-a-time as they are obtained. Many researchers have recommended the number of subgroups $m$ and the sample size $n$ required in order to estimate $\mu$ and $\sigma$. Many of these recommendations suggest that 20 to 30 subgroups of size 4 or 5 are adequate (*e.g.*, see Montgomery (2001, p.208) and Ryan (2000, p.84)). These recommendations are apparently based on empirical experience. Hillier (1969) and Quesenberry (1993) showed that these recommendations result in larger probabilities of false alarm initially than the nominal value 0.0027, and are not sufficient to ensure that the estimated control limits (1) are close enough to the true limits. In order to study the effect of $m$ on the in-control run length (RL) distribution, Quesenberry (1993) used Monte Carlo simulation to estimate the probability of a signal within $x$ subgroups (*i.e.*, $Pr(RL \leq x)$) for several values of $x$ when the control limits were obtained from $m$ subgroups. Quesenberry (1993) suggested that a reasonable number of initial subgroups $m$ should be at least 100 for $n = 5$ so that the estimated control limits can be close to the true limits.

Let $B_i$ denote the event that the $i$th future subgroup mean $\overline{X}_i$ $(i > m)$ either exceeds the UCL or is less than the LCL, where UCL and LCL are defined in (1). Quesenberry (1993) showed that the events $B_i$ and $B_j$ are positively correlated for $i, j > m$, and that this correlation is a function only of $m$ and $n$. Quesenberry (1993) also found that this dependence results in the average run length (ARL) and the standard deviation of run length (SDRL) to increase from their nominal value after short runs. To find the minimum value of $m$ for which the estimated limits perform similarly to the true limits, Quesenberry (1993) conducted a simulation study to investigate the ARL performance of the $\overline{X}$ control chart for different values of $m$, with $n = 5$. The simulation results showed that both ARL and SDRL exceed their nominal values for $m < 50$ and get reasonably close to their nominal values when $m \geq 100$. Hence, Quesenberry (1993) concluded that the number of subgroups should be at least 100 for sample size $n = 5$. For other sample sizes, Quesenberry (1993) suggested that $m$ should be taken at least $\frac{400}{n-1}$ based on the speculation that the degrees of freedom of the variance estimator should be approximately 400 for this to work.

In practice, process engineers often need to start on-line monitoring of the process before $m = \frac{400}{n-1}$ subgroups have been observed and studied. Nedumaran and Pignatiello (2001) proposed an approach to construct $\overline{X}$ control chart when only a few subgroups are available. In their approach, prospective control limits are constructed for a specific number $(k)$ of future subgroups so that the probability of a signal within the $k$ future subgroups will match the corresponding percentile of the RL distribution of the true limits. Let $\overline{X}_i$ be the sample mean of a future subgroup, $i = m+1, m+2, \ldots, m+k$. Let $\overline{V}$ be the average variance of the $m$ initial in-control subgroups, that is,

$$\overline{V} = \frac{1}{m}\sum_{i=1}^{m} S_i^2 = \frac{1}{m(n-1)}\sum_{i=1}^{m}\sum_{j=1}^{n}(X_{ij}-\overline{X}_i)^2. \quad (2)$$

Then, we can show that $E[\overline{X}_i - \overline{\overline{X}}] = 0$ and $Var[\overline{X}_i - \overline{\overline{X}}] = \left(\frac{m+1}{mn}\right)\sigma^2$. Let $Z_i = \frac{\overline{X}_i - \overline{\overline{X}}}{\sigma\sqrt{\frac{m+1}{mn}}}$, $i = m+1, m+2, \ldots, m+k$. It is easy to show that $Corr(Z_i, Z_j) = \frac{1}{m+1}$, for $i \neq j$ and $i, j > m$. Thus, $(Z_{m+1}, Z_{m+2}, \ldots, Z_{m+k})$ has a positively equicorrelated multivariate normal distribution. Let $T_i = \frac{\overline{X}_i - \overline{\overline{X}}}{\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}}}$, $i = m+1, m+2, \ldots, m+k$. Then, $(T_{m+1}, T_{m+2}, \ldots, T_{m+k})$ has a positively equicorrelated multivariate $t$-distribution with correlation $\frac{1}{m+1}$. Let $\gamma$ denote the desired probability of a signal with $k$ future subgroups such that $Pr(\widehat{LCL} \leq \overline{X}_i \leq \widehat{UCL}, \quad i = m+1, m+2, \ldots, m+k) = 1-\gamma$. Nedumaran and Pignatiello (2001) showed that the control limits $\widehat{UCL}$ and $\widehat{LCL}$ for the next $k$ subgroups are given by

$$\widehat{UCL} = \overline{\overline{X}} + h'_{\gamma,m,k,\nu}\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}}$$
$$\widehat{LCL} = \overline{\overline{X}} - h'_{\gamma,m,k,\nu}\sqrt{\frac{m+1}{mn}}\sqrt{\overline{V}},$$

where $\nu = m(n-1)$ is the degrees of freedom associated with the variance estimator $\overline{V}$ and $h'_{\gamma,m,k,\nu}$ is the critical value such that $Pr\left[\max_{m+1\leq i\leq m+k}|T_i| \leq h'_{\gamma,m,k,\nu}\right] = 1-\gamma$. Nedumaran and Pignatiello (2001) suggested that after the first $k$ future subgroups have been observed and plotted on the control chart, new prospective control limits are constructed for the next $k$ future subgroups. That is, $\overline{\overline{X}}$ and $\overline{V}$ are re-computed from $(m+k)$ initial subgroups. Prospective control limits can be constructed repeatedly, for $k$ future subgroups at a time, until $(m+bk)$ initial subgroups are obtained and studied, where $b \geq 1$ is a given positive integer and $(m+(b+1)k) \geq \frac{400}{n-1}$. The simulation results in Nedumaran and Pignatiello (2001) showed that the control limits constructed using their proposed method perform similarly to the true limits even when estimated from a small number of subgroups.

Although this approach accounts for the correlation between events $B_i$ and $B_j$ within the $k$ subgroups and performs well, it is inconvenient to obtain the critical values, $h'_{\gamma,m,k,\nu}$, for quality practitioners. In this paper, we propose an approach for constructing prospective control limits that can avoid dealing with the correlation between $B_i$ and $B_j$, and only need to use the tables of student's $t$-distribution to obtain the critical values.

# 2 Proposed Method

Assume that $X_{ij}$, $i = 1, 2, \ldots, m$, and $j = 1, 2, \ldots, n$, represent $m$ samples of size $n$ from an in-control process having a normal distribution with mean $\mu$ and variance $\sigma^2$, and assume that $k$ is the specified number of future subgroups inspected at each stage. Let $a$ be the largest integer that is less than or equal to $\frac{m}{k}$, and let $c = m - ka$. Then, the $m$ initial subgroups can be divided into $k$ collections, each has $a$ subgroups. For example, suppose that there are $m = 13$ initial subgroups from an in-control process. They are subgroups #1, #2, …, #10, and #13. Let the number of future subgroups inspected at each stage be $k = 3$. Then, we have 3 collections, each has $a = 4$ subgroups. That is, the first collection consists of subgroups #2, #3, #4 and #5, the second collection consists of subgroups #6, #7, #8 and #9, and the third collection consists of subgroups #10, #11, #12 and #13.

Let

$$Y_i = \overline{X}_{m+i} - \overline{\overline{X}}_i, \tag{3}$$

where $\overline{X}_{m+i}$ is the sample mean of the $i$th future subgroups, $\overline{\overline{X}}_i = \frac{1}{a}\sum_{j=c+(i-1)a+1}^{c+ia}\overline{X}_j$ is the grand mean of the $i$th collection, and $\overline{X}_j$ is the sample mean of the $j$th initial subgroup, $i = 1, 2, \ldots, k$. Then $E(Y_i) = E(\overline{X}_{m+i}) - E(\overline{\overline{X}}_i) = 0$ and $Var(Y_i) = Var(\overline{X}_{m+i}) + Var(\overline{\overline{X}}_i) = \frac{\sigma^2}{n}\left(1+\frac{1}{a}\right)$, for $i = 1, 2, \ldots, k$. It is easy to show that $Corr(Y_i, Y_\ell) = Corr(\overline{X}_{m+i} - \overline{\overline{X}}_i, \overline{X}_{m+\ell} - \overline{\overline{X}}_\ell) = 0$, for $i \neq \ell$ and $i, \ell = 1, 2, \ldots, k$. Hence, $Y_1, Y_2, \ldots, Y_k$ are independent and identically normally distributed with mean 0 and variance $\frac{\sigma^2}{n}\left(1+\frac{1}{a}\right)$. In general, $\sigma^2$ is unknown and it can be estimated by $\overline{V}$ which is defined in (2). Thus,

$$\frac{Y_i}{\sqrt{\frac{\overline{V}}{n}\left(1+\frac{1}{a}\right)}} = \frac{\overline{X}_{m+i} - \overline{\overline{X}}_i}{\sqrt{\frac{\overline{V}}{n}\left(1+\frac{1}{a}\right)}}$$

has a student's $t$-distribution with degrees of freedom $\nu = m(n-1)$, for $i = 1, 2, \ldots, k$. It can be shown that $Cov\left(\frac{Y_i}{\sqrt{\frac{\overline{V}}{n}(1+\frac{1}{a})}}, \frac{Y_j}{\sqrt{\frac{\overline{V}}{n}(1+\frac{1}{a})}}\right) = 0$, for $i \neq j$, $i, j = 1, 2, \ldots, k$. Thus, $\frac{Y_1}{\sqrt{\frac{\overline{V}}{n}(1+\frac{1}{a})}}, \frac{Y_2}{\sqrt{\frac{\overline{V}}{n}(1+\frac{1}{a})}}, \ldots, \frac{Y_k}{\sqrt{\frac{\overline{V}}{n}(1+\frac{1}{a})}}$ form an uncorrelated sequence.

Let $\gamma$ be the desired probability of a signal within $k$ future subgroups based on the proposed approach and satisfy $Pr(\widetilde{LCL} \leq Y_i \leq \widetilde{UCL},\ i = 1, 2, \ldots, k) = 1 - \gamma$. Let $\delta_i = Pr(\widetilde{LCL} \leq Y_i \leq \widetilde{UCL})$, $i = 1, 2, \ldots, k$. In practice, for the convenience of computation, we can treat $1 - \gamma \approx \delta_1 \cdot \delta_2 \cdots \delta_k$, and then choose $\delta_i = \delta$, for $i = 1, 2, \ldots, k$, such that $\gamma = Pr(RL \leq k) = 1 - (1-\alpha)^k$ is fulfilled, where $\alpha$ is the false alarm probability for one subgroup. Hence, we have $\delta = 1 - \alpha$. Thus, a feasible control chart can be constructed by plotting $Y_1, Y_2, \ldots, Y_k$ on a chart for monitoring the $k$ future subgroups with a center line $\widetilde{CL} = 0$ and control limits

$$\begin{aligned} \widetilde{UCL} &= 0 + t_{\frac{\alpha}{2}(\nu)}\sqrt{\frac{\overline{V}}{n}\left(1+\frac{1}{a}\right)} \\ \widetilde{LCL} &= 0 - t_{\frac{\alpha}{2}(\nu)}\sqrt{\frac{\overline{V}}{n}\left(1+\frac{1}{a}\right)}, \end{aligned} \quad (4)$$

where $t_{\alpha(\nu)}$ denotes the critical value that computed from a student's $t$-distribution with degrees of freedom $\nu = m(n-1)$ such that $Pr(T > t_{\alpha(\nu)}) = \alpha$. After the first $k$ future subgroups have been observed and studied, new prospective control limits should be constructed for the next $k$ future subgroups, and the process parameters can be re-estimated from $m + k$ initial subgroups. This procedure can be done repeatedly until the number of subgroups recommended by Quesenberry (1993) are accumulated. That is, $(m + (d+1)k) \geq \frac{400}{(n-1)}$, where $d$ is a positive integer. When the recommended number is fulfilled, a standard $\overline{X}$ control chart with control limits (1) based on the $(m + (d+1)k)$ retrospective subgroups can be constructed to monitor the future subgroups. Besides, in practice, it is a great convenience to choose $m$ and $k$ so that $\frac{m}{k}$ is also an integer.

As the consideration of Nedumaran and Pignatiello (2001), it should be noted that the paper do not address the issue of controlling the overall probability of a false alarm for all $\frac{400}{n-1}$ subgroups. The quantity $\gamma$ is only applied to each set of $k$ future subgroups plotted on the control chart. Moreover, the proposed method handles only for the correlation between events $B_i$ and $B_j$ within the $k$ subgroups and it does not account for the situation of correlation between sets of $k$ future subgroups. Some numerical results of the new method and its application can be addressed from Tsai (2004).

# 3 Conclusions

A new method is provided in this study to construct prospective control limits for $\overline{X}$ control chart and only need to used the tables of the student's $t$-distribution to obtain the critical values. The simulation results indicate that the proposed control limits perform similarly to the true limits, even when only a small number of subgroups are available.

# Acknowledgements

The work of the first author was partially supported by the National Science Council of R.O.C. grant NSC 92-2118-M-032-009.

# References

Hillier, F. S. (1969). $\overline{X}$ and $R$-Chart control limits based on a small number of subgroups, *Journal of Quality Technology*, **1**, 17-26.

Montgomery, D. C. (2001). *Introduction to Statistical Quality Control*, 4th edition, John Wiley & Sons, New York.

Nedumaran, G. and Pignatiello, J. J. (2001). On estimating $\overline{X}$ control chart limits, *Journal of Quality Technology*, **33**, 206-212.

Quesenberry, C. P.(1993). The effect of sample size on estimated limits for $\overline{X}$ and $X$ control charts, *Journal of Quality Technology*, **25**, 237-247.

Ryan, T. P. (2000). *Statistical Methods for Quality Improvement*, 2nd edition, John Wiley & Sons, New York.

Tzong-Ru Tsai, Shuo-Jye Wu and Hung-Chia Lin (2004). An Alternatative Control Chart Approach Based on Small Number of Subgroups, will appear at *International Journal of Information and Mangement Sciences*, Vol.15, No.4.