行政院國家科學委員會專題研究計畫 成果報告

對稱型分位數管制圖
研究成果報告(精簡版)

計 畫 類 別 ： 個別型
計 畫 編 號 ： NSC 97-2119-M-009-007-
執 行 期 間 ： 97 年 08 月 01 日至 98 年 10 月 31 日
執 行 單 位 ： 國立交通大學統計學研究所

計 畫 主 持 人 ： 陳鄰安

計畫參與人員： 碩士班研究生-兼任助理人員：游雅芳
碩士班研究生-兼任助理人員：侯智飛
博士班研究生-兼任助理人員：謝宛茹

報 告 附 件 ： 國外研究心得報告

處 理 方 式 ： 本計畫可公開查詢

中 華 民 國 98 年 12 月 03 日

Reserach Project Report: Symmetric Quantile Control Chart

Reporter: Lin-An Chen (Institute of Statistics, National Chiao Tung University)

Project number: NSC 97-2119-M-009-007

Report type: Brief Version

Excution period: 2008/08/01 - 2009/10/31

Contents:

Typeset by $\mathcal{AMS}$-TEX

1. Introduction

In statistical process control, one understanding is that it is not enough to control just the distribution mean, nor it is enough to control just the distribution standard deviation. Repco (1986), Van Nuland (1992), Chao and Cheng (1996) and Spiring and Cheng (1998) developed combination control charts to control the normal mean and standard deviation simultaneously. A good review of the existing works for control chart of this problem can be found in Cheng and Thaga (2006). These works basically proposed tests dealing with hypothesis of the normal parameters as

$$H_0 : \mu = \mu_0, \sigma = \sigma_0$$

in a simultaneous surveillance of location and scale parameters.

It is known that the performance of normal-based control charts is seriously degraded if the underlying distribution is different from normal while manufacturing process with non-normal characteristic variable is, however, very common (see, for examples, Cheng and Thaga (2006), Schiling and Nelson (1976) and Kanji and Arif (2000)).

2. Research Purpose

With full investigation of parametric control chart studies, our research interest is constructing the nonparametric control charts since it makes no assumptions concerning the type of controlling variable.

3. Literature Review

Some nonparametric control charts are suggested. For examples, Janacek and Meikle (1997) considered the median chart, Liu and Tang (1996) considered the boostrap control chart and Grimshaw and Alt (1997) considered using quantile function to construct the control chart. Very few of the existed nonparametric control charts that are designed to simultaneously control distribution parameters where the quantile based control chart by Grimshaw and Alt (1997) is an exception.

With considering nonparametric control chart, it is interesting to study if there is an alternative quantile technique that its produced quantile control chart may gain benifit of better efficiency in some sense than the empirical quantiles based control chart of Grimshaw and Alt (1997). In an attempt to improve the efficiency of the trimmed mean for estimating the location parameter, Kim (1992) and Chen and Chiang (1996) introduced the symmetric quantile to construct an alternative trimmed mean. Chen and Chiang (1996) observed that this trimmed mean can has asymptotic variances very close to the Cramer-Rao lower bounds for several distributions, including heavy tail ones. Can this interesting result be carried over to the construction of control chart. Our aim in this research is to construct an alternative control chart by symmetric quantiles and show that it does gain better efficiency than the classical version constructed by empirical quantiles.

4. Methodologies

For percentages $\alpha_1, ..., \alpha_k$ with $\alpha_1 < \alpha_2 < ... < \alpha_k$, let us consider the population quantile vector

$$Q(\alpha_1, ..., \alpha_k) = \begin{pmatrix} F^{-1}(\alpha_1) \\ F^{-1}(\alpha_2) \\ \vdots \\ F^{-1}(\alpha_k) \end{pmatrix}$$

for monitoring that can be estimated by the corresponding empirical quantile vector

$$Q_e(\alpha_1, ..., \alpha_k) = \begin{pmatrix} F_n^{-1}(\alpha_1) \\ F_n^{-1}(\alpha_2) \\ \vdots \\ F_n^{-1}(\alpha_k) \end{pmatrix}$$

where F_n^{-1} is the empirical quantile function. Grimshaw and Alt (1997) proposed to apply $Q_e(\alpha_1, ..., \alpha_k)$ to monitor the population quantile vector $Q(\alpha_1, ..., \alpha_k)$. The asymptotic property of $Q_e(\alpha_1, ..., \alpha_k)$ relies on the empirical quantiles $F_n^{-1}(\alpha_j)$'s.

Suppose that we have a training sample $y_{ij}, i = 1, ..., n, j = 1, ..., m$ that represents an in-control data set of m samples of size n from distribution F so that estimate of $Q(\alpha_1, ..., \alpha_k)$ and its covariance matrix Σ are available. Generally we let $Q_{ej}(\alpha_1, ..., \alpha_k)$ and $\hat{\Sigma}_j$ be estimates, respectively, based on sample $y_{ij}, i = 1, ..., n$ and define $Q_0(\alpha_1, ..., \alpha_k) = \frac{1}{m}\sum_{j=1}^{m} Q_{ej}(\alpha_1, ..., \alpha_k)$ and $\Sigma_0 = \frac{1}{m}\sum_{j=1}^{m}\hat{\Sigma}_j$. Treated estimates Q_0 and Σ_0 as true values of $Q(\alpha_1, ..., \alpha_k)$ and Σ, the control statistic and upper control limit proposed by Grimshaw and Alt (1997) are

Control statistic $T_e = n(Q_e(\alpha_1, ..., \alpha_k) - Q_0(\alpha_1, ..., \alpha_k))'\Sigma_0^{-1}(Q_e(\alpha_1, ..., \alpha_k)$
$- Q_0(\alpha_1, ..., \alpha_k))$

$UCL_e = \chi^2_\alpha$

Unlike that the empirical quantile is constructed based on the cumulative distribution function, the so-called symmetric quantile of Chen and Chiang (1996) is formulated based on a folded distribution function. The folded cumulative function about a constant μ, known or unknwon, is

$$G_s(a) = P(|y - \mu| \leq a), a \geq 0.$$

Then, the γ symmetric quantile pair is defined as

$$\{F_s^{(-)}(\gamma), F_s^{(+)}(\gamma)\} = \{\mu - G_s^{-1}(\gamma), \mu + G_s^{-1}(\gamma)\}$$

where $G_s^{-1}(\gamma) = \inf\{a : G_s(a) \geq \gamma\}$. If distribution function F is continuous, the γ symmetric quantile pair satisfies $\gamma = P(F_s^{(-)}(\gamma) \leq y \leq F_s^{(+)}(\gamma))$. If we further assume that F is symmetric at μ, it can be seen that

$$\{F_s^{(-)}(\gamma), F_s^{(+)}(\gamma)\} = \{F^{-1}(\frac{1-\gamma}{2}), F^{-1}(\frac{1+\gamma}{2})\},$$

the population classical quantiles and the symmetric quantiles are identical. This leads to the fact that sample type symmetric quantiles can play the role of the empirical quantiles to estimate the population quantiles $F^{-1}(\alpha)'s$.

For the random sample $y_1, ..., y_n$ from distribution F. let $\hat{\mu}$ be an estimate of μ. We may define the sample type γ symmetric quantile pair as

$$\{F_{sn}^{(-)}(\gamma), F_{sn}^{(+)}(\gamma)\} = \{\hat{\mu} - G_{sn}^{-1}(\gamma), \hat{\mu} + G_{sn}^{-1}(\gamma)\}$$

with sample folded distribution function $G_{sn}(a) = \frac{1}{n}\sum_{i=1}^{n} I(|y_i - \hat{\mu}| \le a)$ and

$$G_{sn}^{-1}(\gamma) = \inf\{a : G_{sn}(a) \ge \gamma\}.$$

4. Results and Discussion

Considering a number ℓ decreasing percentages $\gamma_1 > \gamma_2 > ... > \gamma_\ell$, we define its corresponding 2ℓ symmetric quantile vector and population symmetric quantiles, respectively, as

$$Q_{sn}(\gamma_1, ..., \gamma_\ell) = \begin{pmatrix} F_{sn}^{(-)}(\gamma_1) \\ F_{sn}^{(-)}(\gamma_2) \\ \vdots \\ F_{sn}^{(-)}(\gamma_\ell) \\ F_{sn}^{(+)}(\gamma_\ell) \\ F_{sn}^{(+)}(\gamma_{\ell-1}) \\ \vdots \\ F_{sn}^{(+)}(\gamma_1) \end{pmatrix} \text{ and } Q_s(\gamma_1, ..., \gamma_\ell) = \begin{pmatrix} F_s^{(-)}(\gamma_1) \\ F_s^{(-)}(\gamma_2) \\ \vdots \\ F_s^{(-)}(\gamma_\ell) \\ F_s^{(+)}(\gamma_\ell) \\ F_s^{(+)}(\gamma_{\ell-1}) \\ \vdots \\ F_s^{(+)}(\gamma_1) \end{pmatrix}.$$

From Chen and Chiang (1996), we may see that $n^{1/2}(Q_{sn}(\gamma_1, ..., \gamma_\ell) - Q_s(\gamma_1, ..., \gamma_\ell))$ is asymptotically normal $N_{2\ell}(0_{2\ell}, \Sigma_s)$ for some matrix Σ_s that will be given explicitly latter. This further implies that the following

$$n(Q_{sn}(\gamma_1, ..., \gamma_\ell) - Q_s(\gamma_1, ..., \gamma_\ell))'\Sigma_s^{-1}(Q_{sn}(\gamma_1, ..., \gamma_\ell) - Q_s(\gamma_1, ..., \gamma_\ell)) \to \chi^2(2\ell)$$

holds asymptotically in distribution.

Again, from the training sample $y_{ij}, i = 1, ..., n, j = 1, ..., m$ that represents an in-control data set of m samples of size

n, we let $Q_{s0}(\gamma_1, ..., \gamma_\ell) = \frac{1}{m}\sum_{j=1}^{m} Q_{sn,j}(\gamma_1, ..., \gamma_\ell)$ and $\Sigma_{s0} = \frac{1}{m}\sum_{j=1}^{m} \hat{\Sigma}_{s,j}$ where $Q_{sn,j}(\gamma_1, ..., \gamma_\ell)$ and $\hat{\Sigma}_{s,j}$ are estimates of, respectively, $Q_s(\gamma_1, ..., \gamma_\ell)$ and Σ_s. Let us denote these two estimates by Q_{s0} and Σ_{s0}. Based on these estimates, we proposed control statistic and upper control limit as

Control statistic $T_s = n(Q_{sn}(\gamma_1, ..., \gamma_\ell) - Q_{s0}(\gamma_1, ..., \gamma_\ell))'\Sigma_{s0}^{-1}$
$(Q_{sn}(\gamma_1, ..., \gamma_\ell) - Q_{s0}(\gamma_1, ..., \gamma_\ell))$
$UCL = \chi_\alpha^2(2\ell)$

We observe that the estimation of population quantile vector by empirical quantiles is more efficient when the quantile percentage is 0.6. However, it is impressed that it gains more precision to use symmetric quantile to construct the quantile vector estimator when percentage is equal or more than 0.8. In fact, the case that when the underlying distribution is the Laplace one the estimator of quantile vector constructed by symmetric quantiles totally dominate the one by empirical quantiles.

The average run length (ARL), representing the average number of samples taken before an action signal is given, is the most popular technique in evaluating a control chart or comparaison of alternative control charts. In the case of process being in-control, both $ARL's$ by symmetric quantiles and by empirical quantiles are the expected number 200 for setting $\alpha = 0.05$. Surprisingly $ARL's$ of symmetric quantiles are all smaller than the corresponding $ARL's$ of empirical quantiles. This indicates that the symmetric quantiles based control chart can detect the distributional shift with smaller number of samples.

In application, it is more interesting in comparing these two control charts when the coverage probability of the charts is fixed at 0.9973. In this consideration, $ARL's$ are expected to be 370 when the process is in-control. We then com pute $ARL's$ for two quantile control charts with significance level 0.0027. We see that in this setting of coverage interval the symmetric quantiles based control chart is still more efficient than the empirical quantiles based control chart in dectection of distributional shift.

6. references

Chao, M. T. and Cheng, S. W. (1996). Semicircle control chart for variables data. *Quality Engineering.* **8**, 441-446.

Chen, L.-A. and Chiang, Y. C. (1996). Symmetric type quantile and trimmed means for location and linear regression model. *Journal of Nonparametric Statistics.* 7, 171-185.

Cheng, S. W. and Thaga, K. (2005). Multivariate Max-CUSUM chart. *Quality Technology and Quantitative Management International.* **2**, 191-206.

Grimshaw, S. D. and Alt, F. B. (1997). Control charts for quantile function values. *Journal of Quality Technology,* **29**, 1-7.

Janacek, G. J. and Meikle, S. E. (1997). Control charts based on medians. *The Statistician.* 46, 19-31.

Kanji, G. K. and Arif, O. H. (2000). Median ranki control chart by quantile approach. *Journal of Applied Statistics.* **27**, 757-770.

Kim, S. J. (1992). The metrically trimmed means as a robust estimator of location, *Annals of Statistics.* 20, 1534-1547.

Liu, R. Y. and Tang, J. (1996). Control charts for dependent and independent measurements based on bootstrap methods. *Journal of the American Statistical Association*, **91**, 1694-1700.

Repco, J. (21986). Process capability plot. *The Proceedings of the 330th EQQC Conference.* 373-381.

Ruppert, D. and Carroll, R.J. (1980). Trimmed least squares estimation in the linear model. *Journal of American Statistical Association* **75**, 828-838.

Shiling, E. G. and Nelson, P. R. (1976). The effect of non-normality on the control limits of $\bar{X}$ control chart. *Journal of Quality Technology*, **8**, 183-187.

Spiring, F. A. and Cheng, S. W. (1998). An alternative variables control chart: The univariate and multivariate case. *Statistica Sinica.* **8**, 273-287.

Van Nuland, Y. (1992). ISO 9002 and the circle technique. *Quality Engineering.* **5**, 269-291.

6. Self Evaluation

This study presents an alternative nonparametric control chart. This symmetric quantile based control chart inherits the adavantage of efficient estimation of population quantiles in resulting efficient estimators of population control limits. The average run length study also show that this new control chart is appealing in application. In the future research, this new control chart is desired to be expanded to the multivariate distribution.

出國移地研究結束報告

申請人：陳鄰安

單位：交通大學統計所

國科會計畫名稱：近似容忍區間 95-2118-M-009-007-

執行期限：2006/08/01—2007/10/31

出國期間：2007 年 8 月 7 日至 2007 年 9 月 4 日

研究地點：School of Public Health, U. of Texas.

研究成果：本次研究我們完成了在存分析相關分配的近似容忍區間之理論分析並建立相當有價值的資料分析。同時也完成兩項相關議題：

(a) 參考區間(Reference interval)的無母數估計。

傳統上無母數方法估計參考區間是用經驗分位數(Empirical quantile)來估計的。我們驗證由對稱分位數(Symmetric quantile, Chen and Chiang (1996, J. of Nonparametric Statistics))具有較小變異之優點。

(b) 參考區間之診斷

一般醫院採用參考區間來診斷病人，參考區間是醫師用來診斷病人有病沒病的重要根據。一般醫院採用學術界、衛生單位或大醫院所建立的參考區間之前必須用統計方法診斷其是否對本地區居民眾有代表性。傳統方法皆非常粗糙。本計劃將建立由概似函數方法的一種統計診斷方法。