The Korean Communications in Statistics **Vol. 14** No. 2, 2007, pp. 365–376

Median Control Chart using the Bootstrap Method

Soo-Duck Lim[1] Hyo-Il Park[2] and Joong-Jae Cho[3]

Abstract

This research considers to propose the control charts using median for the location parameter. In order to decide the control limits, we apply several bootstrap methods through the approach obtaining the confidence interval except the standard bootstrap method. Then we illustrate our procedure using an example and compare the performance among the various bootstrap methods by obtaining the length between control limits through the simulation study. The standard bootstrap may be apt to yield shortest length while the bootstrap-*t* method, the longest one. Finally we comment briefly about some specific features as concluding remarks.

Keywords: Bootstrap method; control chart; median.

1. 서론

생산 현장에서 제품의 품질 향상과 품질 유지를 위하여 사용하는 위치모수에 대한 계량형 관리도 중의 하나인 $\bar{X}$-관리도의 특징은 다음과 같다.

(i) 생산과정의 분포가 정규분포라는 가정 하에서 작성됨.

(ii) 특이치 (outlier) 에 대하여 민감한 반응을 보임.

그러나 실제로는 정규분포보다 꼬리 부분이 두꺼운 (heavy-tailed) 속성을 지닌 분포를 갖는 자료가 생산현장에서 자주 발생한다는 것이 적지 않은 통계학자와 공학자들의 생각이다 (*cf.* Gunter, 1989). 따라서 특이치의 발견도 공정 분포가 비정규분포라는 점이 가장 큰 큰 원인이라고 여긴다. 모집단의 분포가 정규분포라는 가정이 용이하지 않은 경우에 통계적 추론 방법으로서 로버스트한 여러 가지 통계적 절차가 제시되었지만 비모수적 방법이 가정을 최소한으로 요구하며 사용하기 쉬워 정규성이 성립되는 경우에는 효율성이 떨어지는 단점에도 불구하고 많이 이용되고 있다. 평균을 대신할 수 있

1) Graduate Student, Department of Information and Statistics, Chungbuk National University, Chong-ju, Choong-book 361-763, Korea.
Correspondence : sdlim@dreanwiz.com
2) Professor, Department of Statistics, Chong-ju University, Chong-ju, Choong-book 360-764, Korea.
3) Professor, Department of Information and Statistics, Chungbuk National University, Chong-ju, Choong-book 361-763, Korea.

는 위치모수로는 중앙값 (median) 이 있으며 중앙값을 이용한 비모수적 검정절차와 추정이 많이 연구 제시 되어 있다. 이번 연구에서도 중앙값을 이용하여 위치모수에 대한 관리도를 제시하는 것이 연구 목표이다. 물론 중앙값을 이용한 관리도에 관한 연구는 처음이 아니며 기존에 여러 연구가 있으나 박효일 (2006)이 단순 표본 중앙값 또는 일반화한 중앙값 추정치를 이용한 관리도에 관하여 문헌을 탐색한 결과 기존의 중앙값 관리도는 다음과 같은 문제점들이 존재하기 때문에 실제 현장에서는 거의 이용되지 못하고 있는 것이다. Nelson (1982), Janacek과 Meikle (1997) 그리고 Khoo (2005)는 중앙값을 이용한 관리도를 제시하며 생산과정의 분포가 정규분포라는 가정 하에서 관리한계선을 결정하거나 또는 beta-binomial 분포를 이용하여 관리한계선을 결정하였다. 그러나 beta-binomial 분포 자체가 다루기 어려운 분포여서 이용에 제한이 있을 수밖에 없으며 Alloway와 Raghavachari (1991)는 Wilcoxon 부호순위검정통계량을 이용하여 중앙값에 대한 Hodges-Lehmann 추정치와 이와 관련된 구간추정을 이용하여 중앙값 관리도를 제시하였으나 3σ 관리한계선을 결정하기 위하여 표본의 크기는 최소 10이 되어야 하고 Wilcoxon 부호순위검정통계량의 분포가 이산형이므로 정확한 3σ 관리한계선을 정할 수도 없다. Grimshaw와 Alt (1997)는 분위수 함수를 이용하여 전통적인 관리도라는 개념 대신에 가설검정의 접근방법에 의한 공정의 변화를 확인하는 관리도로서 공정의 변화를 기록할 수 없다는 점과 공정 평균 또는 중앙값의 변화에 대한 판단보다는 분포함수의 변화를 감지하는 용도로서 적합도 검정 (goodness-of-fit test)에 사용하는 것이 더욱 적절한 것으로 판단된다. 더욱이 보다 더 불편한 사항은 검정절차를 이용하는 관리도 형태이므로 작업현장에 직접 적용하기가 용이하지 않다는 점이다.

표본 중앙값을 이용하여 관리도를 구축하는 경우 표본 중앙값의 분포함수와 극한 분산에는 확률밀도함수를 포함하고 있기 때문에 비모수적 방법을 이용하는 경우에는 미지의 확률밀도함수를 추정을 해야 한다. 그러나 확률밀도함수를 추정하여도 표본 중앙값의 분포함수를 이용한 관리한계선을 결정하기가 불가능하기 때문에 소표본인 경우에도 극한 분포인 정규근사를 이용하여 관리한계선을 결정하게 된다. 그러나 소표본인 경우에도 극한 분산의 추정치를 이용한다면 추정의 편의 또는 오차의 크기가 커질 것이라는 것은 명백하므로 각 표본의 크기에 대한 분산의 정확한 표현식이 있다면 이를 이용하여 분산의 추정치를 구한다면 극한 분산의 추정치 보다는 편의 또는 오차가 적을 것으로 기대할 수 있다. Maritz와 Jarrett (1978)는 변수변환을 이용하여 미지의 확률밀도함수를 추정하지 않으면서 표본중앙값에 대한 분산을 추정하였다. 박효일 (2006)은 Maritz와 Jarrett (1978)가 제시한 표본중앙값에 대한 분산의 추정치를 이용하여 중앙값 관리도를 제시하였다.

Efron (1979)이 붓스트랩 방법을 통계학 분야에 제시한 이래 붓스트랩 방법은 분산의 추정과 신뢰구간 그리고 가설검정 분야에서 많은 이론적인 발전과 응용이 이루어졌으나 놀랍게도 품질관리 분야에서는 거의 이용되지 않았다. 관리도 분야에 붓스트랩 방법의 이용은 Seppala 등 (1995)이 소표본 그리고 크게 왜곡 (skewed)된 분포인 경우 $\bar{X}$-관리도에 대한 관리한계선을 결정하는 데 붓스트랩 percentile 방법을 이용하였으며 Liu와 Tang (1996)은 의존적 자료에 대한 관리도를 제시하면서 역시 관리한계선을 결

정하는 데 있어 붓스트랩 방법을 사용하였다. Jones와 Woodall (1998)은 $\bar{X}$-관리도에 대한 관리한계선을 결정하는 데 사용되는 붓스트랩 방법에 대한 효용성을 평균 런 길이를 이용하여 모의실험을 통하여 연구하였다. 참고로 Shao와 Tu (1995)는 Maritz와 Jarrett (1978)의 표본 중앙값의 분산에 대한 추정치가 붓스트랩 추정치와 일치한다는 점을 밝혔다.

이번 연구에서는 다양한 붓스트랩 방법을 사용하여 중앙값 관리도에 대한 관리한계선을 결정하는 과정을 제시한다. 2장에서는 중앙값 관리도에서 Maritz와 Jarrett (1978)이 제시한 분산의 추정치를 이용하는 관리도(*cf.* 박효일, 2006)를 재검토하며 3장에서는 다양한 붓스트랩 방법을 이용한 중앙값 관리도에 대한 관리한계선을 결정하는 과정을 제시하고 4장에서는 Alloway와 Raghavachari (1991)가 이용한 자료를 사용하여 우리가 제시하는 다양한 붓스트랩 방법에 의하여 제시된 관리한계선과 2장에서 제시된 관리한계선 그리고 앞서 연구, 제시된 몇 가지 관리한계선과의 비교를 한 다음 끝으로 몇 가지 결언으로서 끝맺고자 한다.

2. 중앙값 관리도

중앙값이 θ인 분포함수 F를 갖는 모집단에서 추출된 표본 $X_1, \ldots, X_n$을 이용한 표본중앙값 $\widetilde{X}$은 다음과 같이 정의된다.

$$\widetilde{X} = \mathrm{med}\{X_1, \ldots, X_n\}.$$

만약 표본의 크기 n이 홀수이면 즉, $n = 2k + 1$이면

$$\widetilde{X} = X_{(k+1)}$$

이며 n이 짝수이면 즉, $n = 2k$이면

$$\widetilde{X} = \frac{X_{(k)} + X_{(k+1)}}{2}$$

라고 정의한다. 여기서 $X_{(i)}$는 표본 $X_1, \ldots, X_n$에서 i번째 순서통계량을 의미한다. 가장 단순한 중앙값 관리도를 구축하기 위하여 다음의 정리를 이용할 수 있다. 다음에서 f는 분포함수 F에 대한 확률밀도함수를 의미한다.

정리 2.1 $n \to \infty$함에 따라 다음의 결과를 얻는다.

$$\sqrt{n}\left(\widetilde{X} - \theta\right) \ \to^d \ N\left(0, \frac{1}{4f^2(\theta)}\right). \tag{2.1}$$

주의: n이 홀수와 짝수 두 경우 모두 근사분포는 (2.1)과 같다.

증명: n이 홀수인 경우에 Serfling (1980)의 정리A에 따르는 따름정리 B(p.77)에서 $p = 1/2$ 그리고 $\xi_p = \theta$를 취하면 정리 2.1을 얻을 수 있다. n이 짝수인 경우에 박효일 (2006)의 부록에서 확인할 수 있다. □

따라서 (2.1)을 이용한 근사적 관리한계선은 미지의 확률밀도함수 f를 추정하여 표준정규분포표로부터 상위 α^{th} 백분위수 z_α를 구하여 다음과 같이 결정할 수 있다.

$$LCL = \widetilde{X} - z_\alpha\sqrt{1/4n\widehat{f^2}(\widetilde{X})} \text{ 그리고 } UCL = \widetilde{X} + z_\alpha\sqrt{1/4n\widehat{f^2}(\widetilde{X})} \tag{2.2}$$

(2.2)에서 LCL과 UCL은 각각 관리 하한선과 상한선을 의미하며 $\widehat{f}$는 f의 추정치이다. 위의 관리한계선에서 사용한 $\widetilde{X}$의 분산 추정치는 극한분산에 대한 추정이다. Maritz와 Jarrett (1978)는 다음과 같이 각 n에 대한 $\widetilde{X}$의 정확한 분산의 표현식을 이용하여 분산을 추정하였다. 먼저 $n = 2k+1$, 즉 홀수인 경우에 대하여 생각한다. $\widetilde{X}$에 대한 분산, $V(\widetilde{X})$은 $V(\widetilde{X}) = E(\widetilde{X}^2) - \{E(\widetilde{X})\}^2$이므로 $\widetilde{X} = X_{(k+1)}$에 대한 분포함수를 이용하여 다음과 같이 적률을 먼저 생각한다. 아래의 F^{-1}는 분포함수 F에 대한 분위수 함수를 의미한다. 즉, 다시 말하면 임의의 $0 < p < 1$에 대하여

$$F^{-1}(p) = \inf\{t : F(t) \geq p\}.$$

따라서 $d = 1,\ 2$인 경우에 d차 적률은 다음과 같다.

$$\begin{aligned} E(\widetilde{X}^d) &= \frac{n!}{(k!)^2}\int_{-\infty}^{\infty} x^d[F(x)]^k[1-F(x)]^k f(x)dx \\ &= \frac{n!}{(k!)^2}\int_0^1 [F^{-1}(u)]^d u^k(1-u)^k du \\ &= \frac{n!}{(k!)^2}\sum_{j=1}^{n}\int_{(j-1)/n}^{j/n} [F^{-1}(u)]^d u^k(1-u)^k du. \end{aligned}$$

위에서 두 번째 등식은 $u = F(x)$를 이용하는 변수변환의 결과이다. $F^{-1}(u)$의 추정치 $F_n^{-1}(u)$는 다음과 같다. 여기서 F_n는 표본 $X_1, \ldots, X_n$에 의한 경험적 분포함수이며 F_n^{-1}는 이에 대응하는 경험적 분위수 함수이다. 따라서 만약 $(j-1)/n < u \leq j/n$이면

$$F_n^{-1}(u) = \inf\{x : F_n(x) \geq u\} = X_{(j)}$$

이므로 $E(\widetilde{X}^d)$에 대한 추정치 A_{dn}은 다음과 같다.

$$A_{dn} = \sum_{j=1}^{n} X_{(j)}^d \int_{(j-1)/n}^{j/n} u^k(1-u)^k du = \sum_{j=1}^{n} X_{(j)}^d W_j.$$

이러한 적률의 추정치를 이용하여 Maritz와 Jarrett (1978)은 $n = 2k+1$인 경우에 $\widetilde{X}$의 분산에 대한 추정치 $\widehat{\sigma^2}$를 다음과 같이 제시하였다.

$$\widehat{\sigma_n^2} = A_{2n} - A_{1n}^2.$$

$n = 2k$, 즉 짝수인 경우에는 $\widetilde{X} = (X_{(k)} + X_{(k+1)})/2$이므로

$$V\left(\widetilde{X}\right) = \frac{1}{4}E\left(X_{(k)}^2 + X_{(k+1)}^2\right) + \frac{1}{2}E\left(X_{(k)}X_{(k+1)}\right) - \frac{1}{4}\left\{E\left(X_{(k)} + X_{(k+1)}\right)\right\}^2.$$

이러한 관계를 이용하여 다음의 적률들을 구하여 $\widehat{\sigma_n^2}$를 얻었다.

$$B_{dn}=\sum_{j=1}^{n} X_{(j)}^{d} V_j,\ V_j = \frac{n!}{2[k!(k-1)!]} \int_{(j-1)/n}^{j/n} u^{k-1}(1-u)^{k-1} du,$$

$$C_n=\sum_{i=1}^{n}\sum_{j=i}^{n} X_{(i)}X_{(j)}V_{ij},\ V_{ij} = \frac{2k!}{[(k-1)!]^2} \int_{(i-1)/n}^{i/n} v^{k-1} \int_{\max(v,(i-1)/n)}^{j/n} (1-u)^{k-1} du dv,$$

$$\widehat{\sigma_n^2} = \frac{1}{2}(B_{2n}+C_n) - B_{1n}^2.$$

Maritz와 Jarrett (1978)은 $V(\widetilde{X})$가 유한한 값을 가질 때 $\widehat{\sigma_n^2}$는 $V(\widetilde{X})$의 일치성을 갖는 추정치라는 점을 $\widehat{\sigma_n^2} \rightarrow^p V(\widetilde{X})$을 통하여 간단히 보였다. 여기서 $\rightarrow^p$는 확률수렴을 의미한다. 따라서 순서통계량에 대한 정규근사 (*cf.* Wilks, 1961)와 Slutsky 정리를 이용하면

$$\frac{\widetilde{X}-\theta}{\widehat{\sigma_n}} \rightarrow^d N(0,1) \tag{2.3}$$

임을 간단히 보일 수 있다. 여기서 $\rightarrow^d$와 $N(0,1)$은 분포수렴과 표준정규분포를 의미한다.

박효일 (2006)은 (2.3)과 분산에 대한 추정치 $\widehat{\sigma_n^2}$를 이용하여 중앙값 관리도를 다음과 같이 구축하였다. 이를 위하여 표본의 크기가 n인 r개의 표본이 있다고 하자.

(i) r개의 표본 중앙값$(\widetilde{X}_1,\ldots,\widetilde{X}_r)$을 이용하여 다음과 같이 중심선$(CL)$을 얻는다.

$$\widetilde{\widetilde{X}} = \text{med}\left\{\widetilde{X}_1,\ldots,\widetilde{X}_r\right\}.$$

(ii) 관리한계선을 결정하기 위하여 r개의 $\widehat{\sigma_n^2}$를 이용하여 우리는 최종적으로 다음과 같은 σ_n^2에 대한 추정치 $\widehat{\widehat{\sigma_n^2}}$을 구할 수 있다.

$$\widehat{\widehat{\sigma_n^2}} = \text{med}\left\{\widehat{\sigma_{1n}^2},\ldots,\widehat{\sigma_{rn}^2}\right\}.$$

(iii) 정규근사식을 이용하여 다음과 같이 관리한계선을 정한다.

$$LCL = \widetilde{\widetilde{X}} - z_\alpha\widehat{\widehat{\sigma_n}} \text{ 그리고 } UCL = \widetilde{\widetilde{X}} + z_\alpha\widehat{\widehat{\sigma_n}}. \tag{2.4}$$

다음 장에서는 붓스트랩 방법을 이용한 관리한계선을 구하는 방법을 제시하고자한다. 참고로 위에서 구한 $\widehat{\sigma_n^2}$는 붓스트랩 추정치와 일치한다는 점을 Shao와 Tu (1995)가 밝혔다. 따라서 위에서 제시한 관리한계선 (2.4)는 붓스트랩 관리한계선이라고 할 수 있으며 이러한 붓스트랩 방법을 표준 붓스트랩 방법이라고 한다.

3. 붓스트랩 방법을 이용한 중앙값 관리도

붓스트랩 방법을 이용하여 관리한계선을 결정하는 절차를 제시하기 전에 먼저 붓스트랩 절차에 대하여 간단하게 문헌탐색하기로 한다. 붓스트랩 방법은 표본 $X_1, \ldots, X_n$을 모집단으로 간주하고 표본으로부터 복원 재추출하는 방법으로서 전체 재추출의 횟수는 n^n이나 된다. 따라서 전체를 재추출하기는 불가능하기 때문에 Monte-Carlo적인 방법을 택하게 된다. 그러므로 우리가 제시하는 모든 붓스트랩 방법에 의한 관리한계선의 결정은 정확한 붓스트랩 분포에 의한 것이 아니라 모두 Monte-Carlo적인 근사분포에 의하여 결정되는 것이다. 전반적인 Monte-Carlo적 붓스트랩 절차는 다음과 같은 순서로 수행 된다:

(i) 표본 $X_1, \ldots, X_n$로부터 붓스트랩 표본 $X_1^*, \ldots, X_n^*$를 복원 추출한다.

(ii) $X_1^*, \ldots, X_n^*$로부터 붓스트랩 표본 중앙값 $\widetilde{X}^*$를 구한다.

(iii) (i)과 (ii)의 과정을 B번 반복한다. 여기서 B의 크기는 일반적으로 1000이상이어야 한다. 따라서 우리는 다음과 같은 B개의 표본 중앙값을 가질 수 있다.

$$\widetilde{X}_1^*, \widetilde{X}_2^*, \ldots, \widetilde{X}_B^*.$$

위에서 구한 B개의 붓스트랩 표본 중앙값을 이용하여 다음에 소개하는 붓스트랩 방법들을 사용하여 붓스트랩 중앙값 관리도를 작성할 수 있다.

3.1. 붓스트랩 persentile 방법

모든 붓스트랩 방법 중에서 가장 쉬운 방법으로서 다음과 같은 절차로 이용된다. 먼저 $\widetilde{X}$를 이용하여 θ에 대한 구간추정을 하고자 한다면 $\widetilde{X}$의 분포 즉,

$$G_n(x) = \Pr\left\{\widetilde{X} \le x\right\}$$

이 필요할 것이다. 그러나 G_n은 붓스트랩 방법을 이용하여 다음과 같이 추정할 수 있다.

$$G_{boot}(x) = \Pr\left\{\widetilde{X}^* \le x\right\}.$$

따라서 θ에 대한 $(1-2\alpha) \times 100\%$붓스트랩 신뢰구간은 다음과 같이 추정된다.

$$\left(G_{boot}^{-1}(\alpha), G_{boot}^{-1}(1-\alpha)\right).$$

여기서 G_{boot}^{-1}은 붓스트랩 분포함수 G_{boot}의 분위수 함수이다. 위 신뢰구간을 $\widetilde{X}$에 대한 붓스트랩 추정치를 이용하여 표현한다면 다음과 같다.

$$\left(\widetilde{X}^*_{([\alpha B])}, \widetilde{X}^*_{([(1-\alpha)B])}\right).$$

$\widetilde{X}^*_{(i)}$는 붓스트랩 추정치 $\widetilde{X}^*_1, \widetilde{X}^*_2, \ldots, \widetilde{X}^*_B$에서 크기순으로 i번째 크기의 붓스트랩 추정치를 의미하며 $[c]$는 실수 c를 넘지 않는 정수를 의미한다. 따라서 만약 $1-2\alpha$가 우리가 원하는 관리한계선 간의 확률이라면

$$LCL = \widetilde{X}^*_{([\alpha B])} \text{ 이며 } UCL = \widetilde{X}^*_{([(1-\alpha)B])}$$

이 될 것이다.

3.2. bias corrected percentile 붓스트랩 방법

붓스트랩 percentile 방법은 이해하기 쉽고 간편하다는 장점이 있으나 편의가 내재되었을 가능성이 크다는 비판이 제기되었다. 따라서 내재된 편의를 해소하기 위하여 다음과 같은 보정절차를 거치는 붓스트랩 방법이 제시되었는데 이를 bias corrected percentile 붓스트랩 방법이라고 한다. 이에 대한 절차는 다음과 같다. 우선 $\widetilde{X}^*_1, \widetilde{X}^*_2, \ldots, \widetilde{X}^*_B$을 이용하여 다음의 확률을 구한다.

$$p_0 = \Pr\left\{\widetilde{X}^* \leq \widetilde{X}\right\}.$$

p_0를 이용하여 표준정규분포로부터 다음의 분위수 z_0를 구한 후

$$z_0 = \Phi^{-1}(p_0)$$

다음의 두 확률 P_L과 P_U를 구한다.

$$P_L = \Phi(2z_0 - z_{1-\alpha}) \text{ 그리고 } P_U = \Phi(2z_0 + z_{1-\alpha}).$$

두 확률 P_L과 P_U을 이용하여 또 다른 θ에 대한 $(1-2\alpha)\times 100\%$ 붓스트랩 신뢰구간을 구하면 다음과 같다.

$$\left(\widetilde{X}^*_{([P_L B])}, \widetilde{X}^*_{([P_U B])}\right).$$

붓스트랩 percentile 경우와 마찬가지로 위 신뢰구간을 중앙값 관리도의 관리한계선을 결정하는데 이용할 수 있다.

3.3. 붓스트랩-t 방법

2장에서 표본 중앙값 $\widetilde{X}$에 대한 일치성을 갖는 추정치 $\widehat{\sigma^2_n}$를 얻었다. $\widehat{\sigma^2_n}$를 이용하는 붓스트랩 방법을 소개하기로 한다. 먼저

$$H_n(x) = \Pr\left\{(\widetilde{X} - \theta)/\sqrt{\widehat{\sigma^2_n}} \leq x\right\}$$

에 대한 분포를 알 수 있다면 이를 이용하여 θ에 대한 구간추정을 할 수 있을 것이다. 따라서 붓스트랩 방법을 이용하여 다음과 H_n에 대한 붓스트랩 추정치를 얻는다.

$$H_{boot}(x) = \Pr\left\{(\widetilde{X}^* - \widetilde{X})/\sqrt{\widehat{\sigma^2_n}} \leq x\right\}.$$

주어진 α에 대하여 $H_{boot}^{-1}(\alpha)$를 H_{boot}에 대한 분위수이라고 하면 θ에 대한 $(1-2\alpha)\times 100\%$신뢰구간은 다음과 같다.

$$\left(\widetilde{X} - H_{boot}^{-1}(1-\alpha)\sqrt{\widehat{\sigma_n^2}}, \widetilde{X} - H_{boot}^{-1}(\alpha)\sqrt{\widehat{\sigma_n^2}}\right).$$

이러한 붓스트랩 방법을 붓스트랩-t 방법이라고 하며 위 신뢰구간을 이용하여 관리한계선을 결정할 수 있다. 붓스트랩-t는 단순하며 이해하기 쉽고 에지워스 전개를 통하여 신뢰구간의 포함확률과 명목수준 간의 차이가 second order accuracy를 달성함으로서 신뢰 한계선의 정확도가 높다고 알려져 있지만 사용하는 통계량에 대한 분산의 추정치가 있어야 한다는 단점이 있다. 이러한 단점을 보완하기 위하여 다음의 hybrid 붓스트랩 방법을 소개한다.

3.4. hybrid 붓스트랩 방법

앞서 소개한 붓스트랩-t 방법과 유사하지만 보다 적절한 분산의 추정치가 없는 경우에도 편리하게 이용할 수 있는 붓스트랩 방법으로서 hybrid 붓스트랩 방법이 있다. 우선 붓스트랩-t 방법에서와 같이 다음과 같은 붓스트랩 근사분포를 생각하자.

$$K_{boot}(x) = \Pr\left\{\sqrt{n}(\widetilde{X}^* - \widetilde{X}) \le x\right\}.$$

$K_{boot}(x)$은 $K_n(x) = \Pr\{\sqrt{n}(\widetilde{X} - \theta) \le x\}$의 붓스트랩 추정치이므로 다음이 성립한다.

$$\Pr\left\{\sqrt{n}(\widetilde{X} - \theta) \le K_{boot}^{-1}(1-\alpha)\right\} \approx 1-\alpha.$$

여기서 K_{boot}^{-1}는 K_{boot}의 분위수 함수이다. 따라서 θ에 대한 또 하나의 $(1-2\alpha)\times 100\%$ 붓스트랩 신뢰구간은 다음과 같다.

$$\left(\widetilde{X} - K_{boot}^{-1}(1-\alpha)/\sqrt{n}, \widetilde{X} - K_{boot}^{-1}(\alpha)/\sqrt{n}\right).$$

이러한 붓스트랩 방법을 hybrid 붓스트랩 방법이라고 하며 위 구간추정을 이용하여 또 다른 붓스트랩 관리한계선을 결정할 수 있다. hybrid 붓스트랩 신뢰구간의 구축과정에서 알 수 있듯이 hybrid 붓스트랩 방법은 붓스트랩 percentile과 붓스트랩-t의 방법들의 결합이라고 할 수 있으며 비교적 정확도가 높은 신뢰 한계선을 도출한다고 알려져 있어 많이 이용된다.

이상 소개한 다양한 붓스트랩 방법들은 내재된 분포 (underlying distribution)에 따라 선택하기 위하여 연구 개발된 것이 아니라 편의를 줄이고 정확도가 높은 신뢰 한계선을 얻기 위하여 연구 제시 되어 왔다. 따라서 자료의 여러 가지 상황에 따라 적절한 붓스트랩 방법 한 가지를 선택하여 중앙값 관리도를 구축할 수 있으며 다음과 같은 순서로 붓스트랩 방법에 의한 중앙값 관리도를 구축한다. 2장에서와 같이 각각의 표본의 크기가 n인 r개의 표본이 있다고 하자. 각 표본으로부터 구한 r개의 표본 중앙값, $\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_r$이 있으며 붓스트랩 방법들 중 한 가지를 선택하여 r개의 표본에 대하여 r개의 붓스트랩

신뢰구간 $(L_1^*, U_1^*), (L_2^*, U_2^*), \ldots, (L_r^*, U_r^*)$를 구할 수 있다. 이들을 이용하여 다음과 같이 중앙값 관리도를 작성할 수 있다.

$$\begin{aligned} CL &= \text{med}\left\{\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_r\right\}, \\ LCL &= \text{med}\left\{L_1^*, L_2^*, \ldots, L_r^*\right\}, \end{aligned}$$

그리고

$$UCL = \text{med}\left\{U_1^*, U_2^*, \ldots, U_r^*\right\}.$$

4. 예제, 모의실험 및 결언

Alloway와 Raghavachari(1991)가 사용한 Primer Thickness Data를 이용하여 2장과 3장에서 제시된 각 붓스트랩 방법에 대하여 3σ관리한계선을 구하여 표 4.1에 요약하였다. 표 4.1의 붓스트랩 방법 중에서 정규근사는 2장에서 소개된 Maritz와 Jarrett (1978)이 제시한 표본중앙값에 대한 분산의 추정치를 이용하는 방법을 의미하며 b-c percentile은 bias corrected 붓스트랩 방법에 의한 관리한계선을 정하는 절차를 의미한다.

표 4.1: 각종 붓스트랩 방법에 의한 관리한계선

붓스트랩 방법	median	LCL	UCL
정규근사	1.1225	0.995	1.245
표준		1.004	1.241
percentile		0.998	1.255
b-c percentile		1.010	1.283
붓스트랩-t		0.945	1.343
hybrid		0.963	1.250

위 표에서 먼저 주목하는 결과는 정규근사와 표준 붓스트랩 방법에 의한 관리한계선의 결과로서 비슷한 수치를 얻었다. 이는 Maritz와 Jarrett (1978)이 제시한 표본중앙값에 대한 분산의 추정치가 붓스트랩 추정치와 일치한다는 점을 확인한다는 의미가 있다. Alloway와 Raghavachari (1991)는 Hodges-Lehmann 추정치를 이용하여 다음과 같은 3σ관리한계선을 얻었다.

$$LCL = 0.950 \text{ 그리고 } UCL = 1.310.$$

따라서 표준 붓스트랩 방법을 이용한 중앙값 관리도는 Alloway와 Raghavachari (1991)가 제시한 관리한계선 보다는 훨씬 구간이 짧으며 오히려 정규분포를 가정하여 구한 $\bar{X}$-관리도 ($LCL = 1.014$ 그리고 $UCL = 1.230$)의 결과에 근접하는 수치를 얻었다. 나머지 방법들도 Alloway와 Raghavachari (1991)의 방법보다는 관리한계선의 길이가 훨씬 짧았다. 따라서 중앙값 관리도에 대한 관리한계선의 길이가 지나치게 길게 된다는 우려를 잠재울 수 있고 특히 통계량의 분포가 이산성으로 인한 정확한 3σ관리한계선을 정하지

못하는 단점도 극복되는 것이다. 박효일 (2006)의 논문내용 중에서 3σ관리한계선의 길이를 다음과 같이 발표하였으나

$$LCL = 0.8675 \text{ 그리고 } UCL = 1.375$$

이는 표본중앙값에 대한 분산의 추정계산에서 공식을 잘못 사용한 결과로서 이번 논문에서 바로 잡고자 한다. 따라서 우리가 제시하는 중앙값 관리도는 표본평균보다 계산이 간편한 표본중앙값을 이용하는 관리도로서 많은 이용이 기대된다 하겠다.

다음은 모의실험에 대한 결과로서 표준정규분포, 모수가 1인 지수분포 그리고 표준Cauchy 분포를 이용하여 표본의 크기가 10인 난수를 얻어 SAS/IML을 이용하여 결과를 얻어 표4.2–표4.4에 요약하였다. 일반적으로 붓스트랩-t 방법으로 얻어진 관리한계선의 길이가 긴 경향이 있으며 hybrid 붓스트랩 방법에 의한 길이가 비교적 짧다. 그러나 전반적으로 표준 붓스트랩 방법에 의한 관리한계선의 길이가 짧은 경향을 보이고 있으며 붓스트랩 percentile 방법과 bias corrected 붓스트랩 방법은 특히 지수분포인 경우에는 다른 방법들과는 달리 LCL의 값이 양수인 이유는 직접 점추정을 이용하기 때문이다.

붓스트랩 방법에 의한 관리도의 구축은 구간추정에 의하여 관리한계선을 결정하는 것이 일반적이다. 구간추정을 이용하여 관리한계선을 결정하는 경우는 Hodges-Lehmann 추정치를 이용하는 Alloway와 Raghavachari (1991)가 제시한 비모수적 관리도도 있다. 그런데 구간추정을 이용한 관리한계선을 결정하는 경우에는 중앙선의 중심으로부터 LCL과 UCL의 길이가 일반적으로 다르다. 일반적으로 비모수 방법을 이용하여 자료

표 4.2: 표준정규분포

붓스트랩 방법	LCL	UCL
표준	−1.133	1.113
percentile	−1.088	1.087
b-c percentile	−0.979	1.197
붓스트랩-t	−1.805	1.786
hybrid	−1.073	1.077

표 4.3: 지수분포

붓스트랩 방법	LCL	UCL
표준	−0.168	1.641
percentile	0.164	2.023
b-c percentile	0.177	2.235
붓스트랩-t	−0.561	2.575
hybrid	−0.533	1.210

표 4.4: 코시분포

붓스트랩 방법	LCL	UCL
표준	−2.447	2.430
percentile	−2.665	2.449
b-c percentile	−1.767	3.150
붓스트랩-t	−2.937	2.938
hybrid	−2.638	2.790

를 분석하는 이유는 코시분포와 같이 분포의 꼬리가 두꺼운 경우에도 해당하지만 분포가 좌우 대칭이 아닌 경우에도 많이 사용된다. 특히 우유의 포장 과정에 있어서와 같이 under-filling이 문제가 되는 경우에는 분명히 과정의 분포는 상당히 기울어진 (skewed) 상황임에는 틀림이 없다. 따라서 이와 같은 경우에 정규분포 가정 하에서 길이가 같은 관리한계선을 이용하는 관리도는 효율적이 못하는 것은 자명하므로 우리가 제시하는 붓스트랩 방법들 중 하나를 이용하여 구축한다면 훌륭한 대안이 될 수 있을 것이다. 그러나 모든 붓스트랩 방법이 모두 구간추정을 이용하여 관리한계선을 결정하는 것은 아니다. 표준 붓스트랩 방법은 붓스트랩 방법으로 분산을 추정하여 정규근사식을 이용하기 때문에 n의 크기가 중요하지만 나머지 붓스트랩 방법은 분포함수를 추정하기 때문에 표본의 크기와는 상관없다.

마지막으로 3장에서 소개한 여러 가지 붓스트랩 방법 중에서 어느 것을 선택할 지 여부는 각 종 음료수 중에서 한 가지를 선택하는 기호의 선택과 같을 수 있으며 이는 많은 경험을 요구하기도 할 것이며 전반적으로 붓스트랩 방법에 대한 확실한 지식이 요구되기도 할 것이다. 그러나 가장 중요한 사항은 새로운 방법이나 기술에 대한 적극적인 수용 자세가 무엇보다도 필요하므로 우리가 제시한 관리도의 이용을 통하여 기존의 $\bar{X}$-관리도와 비교될 수 있다면 연구자로서 보람도 클 것이다.

감사의 글

논문 내용의 개선과 향상을 위한 두 심사위원의 제안과 지적에 감사를 드립니다.

참고문헌

박효일, (2006). 중앙값을 이용한 관리도의 구축. *Journal of the Korean Data Analysis Society,* **8**, 1545–1555.

Alloway, J. A. Jr. and Raghavachari, M. (1991). Control chart based on the Hodges-Lehmann estimator. *Journal of Quality Technology,* **23,** 336–347.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics,* **7,** 1–26.

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

Gunter, B. H. (1989). The use and abuse of C_{pk}, Part 2, *Quality Progress,* **22,** 108–109.

Khoo, M. B. C. (2005). A control chart based on sample median for the detection of a permanent shift in the process mean. *Quality Engineering,* **17,** 243–257.

Janacek, G. J. and Meikle, S. E. (1997). Control charts based on medians. *Journal of the Royal Statistical Society,* Ser. D-The Statistician, **46,** 19–31.

Jones, L. A. and Woodall, W. H. (1998). The performance of bootstrap control charts. *Journal of Quality Technology,* **30,** 362–375.

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.

Maritz, J. S. and Jarrett, R. G. (1978). A note on estimating the variance of the sample median. *Journal of the American Statistical Association,* **73,** 194–196.

Nelson, L. S. (1982). Control chart for medians. *Journal of Quality Technology,* **14,** 226–227.

Seppala, T., Moskowitz, H., Plante, R. and Tang, J. (1995). Statistical process control via the subgroup bootstrap. *Journal of Quality Technology,* **27,** 139–153.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, New York.

Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer-Verlag, New York.

Singh, K. (1981). On the asymptotic accuracy of Efron's bootstrap. *Annals of Statistics,* **9,** 1187–1195.

Wilks, S. S. (1961). *Mathematical Statistics.* John Wiley & Sons, New York.

[Received February 2007, Accepted June 2007]