

# 베타-이항모형을 이용한 과산포 공정용 $p$ 관리도의 개발

배봉수 · 서순근<sup>†</sup>

동아대학교 산업경영공학과

## Development of a $p$ Control Chart for Overdispersed Process with Beta-Binomial Model

Bae, Bong-Soo · Seo, Sun-Keun<sup>†</sup>

Dept. of Industrial & Management Systems Engineering, Dong-A University

### ABSTRACT

**Purpose:** Since traditional  $p$  chart is unable to deal with the variation of attribute data, this paper proposes a new attribute control chart for nonconforming proportions incorporating overdispersion with a beta-binomial model.

**Methods:** Statistical theories for control chart developed under the beta-binomial model and a new approach using this control chart are presented

**Results:** False alarm probabilities of  $p$  chart with the beta-binomial model are evaluated and demerits of  $p$  chart under overdispersion are discussed from three examples. Hence a concrete procedure for the proposed control chart is provided and illustrated with examples

**Conclusion:** The proposed chart is more useful than traditional  $p$  chart, individual chart to treat observed proportions nonconforming as variable data and Laney  $p'$  chart.

**Key Words:** beta-binomial distribution,  $I$  chart, Laney  $p'$  chart, overdispersion,  $p$  chart

● Received 8 April 2017, 1st revised 30 May 2017, accepted 14 June 2017

† Corresponding Author(skseo@dau.ac.kr)

© 2017, The Korean Society for Quality Management

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-Commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. 서 론

통계적 공정관리(SPC)의 대표적인 기법으로 관리도(control chart)를 들 수 있다. 전통적인 슈하트(Shewhart)의  $\bar{X}$ 와  $p$  및  $u$  관리도는 공정의 품질특성치가 각각 정규, 이항, 포아송 분포를 따를 때 중심선(center line)을 기준으로 관리한계(control limit)를 설정하여 타점되는 통계량이 관리상태를 벗어나는지에 따라 공정의 상태를 모니터링하는 역할을 한다(Wheeler 2001). 이 중에서 계량형인  $\bar{X}$  관리도가 가장 널리 쓰이고 있지만 최근 의료서비스(health care) 분야를 중심으로 계수형인  $p$ ,  $np$ ,  $u$ ,  $c$  관리도의 활용도도 증대되고 있다(Mohammed et al. 2008).

각 부분군의 부적합품률(불량률)을 타점하는  $p$  관리도는 부적합 개수가 이항분포를 따를 경우에 적용하는데, 이들 부분군의 부적합품률은 독립이고, 모수는 동일한 값을 가져야 한다. 이런 조건을 충족하지 않으면 예상된 산포보다 커지거나 적어지는 현상이 일어나 이를 감안하지 않는 관리도를 적용 시에 통계적 오류가 증가된다. 본 논문에서는 실제 산포가 작아지는 경우는 드물므로 산포가 커지는 과산포(overdispersion)인 경우를 대상으로 삼는다. 이런 과산포의 원인으로 공정에서 상관성이 존재하는 경우와 더불어 배치 간에 변동성이 존재하거나 매우 큰 표본크기로 인해 부적합품률이 변동하는 경우를 들 수 있는데, 본 논문에서는 후자의 상황만을 다룬다.

이럴 때 적용할 수 있는 부적합품률 관리도의 첫 번째 유형으로 부적합품률을 계량치로 취급하여 개별값 관리도에 적용하는 방법이다. Wheeler(2001)와 Mohammed et al.(2013)는 각 부분군의 부적합품률을 대상으로  $I-MR$ (국내외 표준에서는  $X-R_m$ 으로 표기됨) 관리도를 적용할 것을 주장하고 있다. 이런 유형의 관리도는 정규분포를 기반으로 삼고 있으므로 정규군사 여부에 대한 검토가 필요하며, 특히 이동범위(MR) 관리도는 산포의 변화를 검출하는데 비효율적이라고 알려져 있으므로(Woodall 1997), 본 논문에서 제안된 관리도와 비교 시에  $I$  관리도만( $I(p)$  관리도로 명명함)을 대상으로 삼는다. 그리고 Heimann(1996)는 각 부분군의 부적합품률로부터 표본분산을 구하고(군내와 군간 변동을 모두 포함한 총 분산),  $I(p)$  관리도의 이동범위의 평균 대신 이를 이용하여 관리도의 관리한계를 설정하는 방식을 제안하였다. 또한 그는  $p$  관리도와 대비하여 제안된 관리도의 채택 기준(표본크기와 총 분산과 군내 분산의 비)을 제시하고 있으나, 이들 기준이 실용적인 관점에서 설정되어 근거의 이론적 타당성에 대한 심층적인 검토가 필요하다.

두 번째 유형으로 부적합품률의 정규성을 제고하기 위해 Quarshie and Shindo(1996)와 Jones and Govindaraju(2000)는 각 부분군의 부적합품률을 역정현(arcsine) 변환하여 전통적인 슈하트 계량형의 관리도에 적용하는 방법을 추천하고 있다. 더욱이 Quarshie and Shindo(1996)는 이 방법보다 이항분포의 동질성을 검정하는 우도비로부터 도출된 통계량을 활용한 관리도가 더 우수하다고 보고하고 있으나, 타점 통계량이 보편적으로 적용되는 부적합률과는 판이하게 다른 형태를 취하고 있어 실용화할 수 있는 방법으로 보기는 힘들다.

상기와는 다른 새로운 유형으로 최근에 Laney(2002)는 군간 변동과 군내 변동을 분리 추정한 후 이를 결합한  $p'$  관리도를 제안하였다. 이 관리도에 대한 학술적 측면의 관심도는 높지 않지만, 관련 통계 및 품질관리 소프트웨어(Minitab(Minitab 사), CHARTrunner(PQ Systems 사), SPC software for Excel(QI Macros 사) 등)에 구현되어 현업에서 활용도가 높은 편이다.(Mohammed et al. 2013).

한편 부분군간 부적합품률의 변화를 수용할 수 있도록 베타-이항분포를 채택한 과산포 모형이 생의학 분야의 통계적 연구 등에서 널리 쓰이나 이를 직접적으로 관리도에 도입한 연구는 찾기 힘들다.

예외적으로 최근에 Raubenheimer and van der Merwe(2016)는  $p$ 의 사전분포로 베타분포를 채택하고 부분군의 실적 데이터를 베이시언(Bayesian) 접근법으로 반영하여 관리한계를 베타-이항분포에 의한 예측분포(predictive

distribution)로 설정하는 방안을 제시하였으며, 이 관리도의 특성을 시뮬레이션에 의해 조사하였다. 이 논문은 과산포를 수용하도록 베타-이항분포를 채택한 현 논문의 빈도론자(frequentist) 접근법과는 근본적으로 다른 베이시언 관리도에 관한 연구이다.

따라서 본 논문의 2절에서 상기에 언급된 대표적인 과산포용 부적합품률 관리도를 먼저 고찰한다. 3절에서는 베타-이항모형을 이용한  $p$  관리도를 제안하고 이의 설계방법을, 4절에서 구체적 적용 절차와 더불어 5절에서 이 관리도의 실제 활용과정을 예시한다. 마지막으로 6절에서 논문의 성과를 요약한다.

## 2. 과산포 공정용 부적합품률 관리도

이 절에서는 비교적 활용도가 높은 과산포용 부적합품률 관리도로, 전절에서 언급한 유형 중에서 첫 번째의  $I(p)$  관리도와 세 번째의  $p'$  관리도를 소개한다.

### 2.1 $I(p)$ 관리도

이 관리도는 각 부분군의 부적합품률을 크기 1의 계량치로 취급하며, 부분군의 개수가  $m$ , 각 부분군의 크기와 부적합 개수가  $n_i, r_i, i=1,2,\dots,m$ 일 때 이동범위(MR)의 평균( $\overline{MR}_p$ )으로부터 총 변동을 측정하여 다음과 같이 관리한계를 설정한다(Mohammed and Worthington 2013).

$$UCL/LCL_{MR} = \bar{p} \pm 2.66 \overline{MR}_p \quad (1)$$

$$\text{단, } \bar{p} = \frac{\sum_{i=1}^m r_i}{\sum_{i=1}^m n_i}, \quad \overline{MR}_p = \frac{\sum_{i=2}^m MR_{p_i}}{m-1}, \quad MR_{p_i} = |p_i - p_{i-1}|, \quad i = 2, \dots, m$$

식 (1)의 관리한계는 각 부분군의 크기가 상이할 경우를 반영하지 못하므로 Wheeler(2001)는  $n_i$ 의 변화범위가  $\bar{n} = \sum_{i=1}^m n_i/m$ 의 20% 범위 내에 있을 경우 식 (1)를 그대로, 그렇지 않을 경우  $\sqrt{\bar{n}/n_i}$ 를 식(1)의 우변 둘째 항에 곱할 것을 추천하고 있다.

한편 이 관리도는 비교적 단순한 방식으로 기존의 널리 쓰이는 개별값 계량치 관리도를 활용하는 장점이 있지만, 0~1 사이의 값을 가지는 부적합품률이 정규분포를 따른다고 취급하는 약점이 있다.

### 2.2 Laney $p'$ 관리도

Laney(2002)에 따르면 먼저 군내 변동을 군내 표준편차인  $\hat{\sigma}_{p_i} = \sqrt{\bar{p}(1-\bar{p})/n_i}$ 로 추정된 후에  $p_i$ 를  $z_i = (p_i - \bar{p})/\hat{\sigma}_{p_i}$ 로 표준화 한다.  $z_i$ 로부터 군간 변동에 대한  $\hat{\sigma}_{z_i}$ 를 다음과 같이 구하고, 이 둘을 결합하여 다음과 같이 관리한계를 설정한다. 여기서 부분군의 크기가 모두 동일하면  $I(p)$  관리도가 된다.

$$UCL/LCL_L = \bar{p} \pm 3\hat{\sigma}_z \hat{\sigma}_{p_i} \tag{2}$$

$$\text{단, } \hat{\sigma}_z = \overline{MR_z}/1.128, \overline{MR_z} = \frac{\sum_{i=2}^m MR_{z_i}}{m-1}, MR_{z_i} = |z_i - z_{i-1}|, i = 2, \dots, m$$

Laney의 방법론에는  $p$ 에 관해 어떤 분포도 가정하지 않으나,  $\sigma_z$ 을 추정할 때 불편상수로 1,128을 채택하고 있으므로 암묵적으로  $z_i$ 는 정규분포를 따른다고 가정하고 있다. 또한 이동범위는 전술한 바와 같이 산포의 변화를 검출하는데 비효율적으로 알려져 있으며, 또한 공정산포의 추정방법으로도 통계적 효율성이 떨어진다(Ryan 2011). 더불어 군내와 군간 표준편차를 곱하는 근거가 충분하지 않는 발견적인(heuristic) 접근법으로 볼 수 있다.

여기서  $\sigma_z > 1$ 이면 과산포로 볼 수 있는데, Minitab(2014)은 시뮬레이션 실험( $m$ 은 20~50,  $n$ 은 50~1000,  $p$ 는 0.01~0.2 범위)으로 구한 Jones and Govindaraju(2000)의 방법론을 적용하여 과산포 여부를 판정하고 있다. 즉, 각 부분군의  $n_i, r_i, i = 1, 2, \dots, m$ 로부터  $\sin^{-1} \sqrt{(r_i + 0.375)/(\bar{n} + 0.75)}$ 로 역정현 변환한 값을 정규확률지에 도시하고, 표준 정규 값이 -1과 1인 두 지점의 거리를 구하여 이 값과 이의 기대 거리( $1/\sqrt{\bar{n}}$ )의 비를 산출한다. 이 비가 식 (3)으로부터 구한 유의수준 5% 기준  $c_{0.05}$ 보다 크면 과산포로 판정한다.  $c_{0.05}$ 은 대강 1.5 근처의 값을 가진다.

$$\ln c_{0.05} = 0.185 + \frac{5.62}{m} + \frac{0.274}{n\bar{p}} \tag{3}$$

과산포를 판정하는 이 방법은 각 부분군의 부적합품률이 정규분포를 따른다는 근거가 부족한 사실에 기반하고 있으며,  $n_i$ 와  $p_i$ 의 변동이 반영되지 않는 단점이 있다. 또한 시뮬레이션 실험에 의해 도출된 결과이므로 유의수준 5%가 아닌 경우에는 활용하기 힘든 점도 있다.

이 절에 소개된 두 관리도는 공정산포(표준편차)를 두 연속값의 이동범위의 평균으로 추정하고 있는데, 두 연속된 이동범위는 독립이 아닌, 상관계수가 0.2239인 양의 상관관계를 가진다(Cryer and Ryan 1990). 또한 개별 값 관리도는 비정규분포를 따를 경우에 보다 심각한 문제를 발생시키므로(Ryan 2011), 이런 점까지 고려하면 이 절에서 언급된 두 관리도를 과산포용 부적합품률 관리도로 추천하기 힘들다.

### 3. 베타-이항모형 하의 과산포용 부적합품률 관리도

#### 3.1 베타-이항모형을 이용한 관리도의 설계

부적합품률이 일정하지 않고 부분군에 따라 변할 경우에 베타-이항모형을 적용하여 과산포를 나타낼 수 있다 (Kleinman 1973; Moore 1987; Ennis and Bi 1998).

$i$ 번째 부분군의 부적합 개수  $X_i$ 는 다음과 같이 이항분포( $B(n_i, p_i)$ )를,  $p_i$ 는 베타분포(beta( $a\pi, a(1-\pi)$ ))를 따른다고 가정한다(이 경우의 베타-이항분포를  $BB(n_i, a, \pi)$ 으로 표기함).

$$f_{X_i | p_i}(x) = \binom{n_i}{x} p_i^x (1 - p_i)^{n_i - x}, \quad x = 0, 1, \dots, n_i$$

$$f_{p_i}(p) = \frac{p^{a\pi - 1} (1 - p)^{a(1 - \pi) - 1}}{B(a\pi, a(1 - \pi))}, \quad 0 < p < 1$$

여기서  $B(\cdot, \cdot)$ 는 베타함수임.

이에 따라  $p_i$ 의 기댓값과 분산은 다음과 같이 구해진다.

$$E(p_i) = \pi$$

$$Var(p_i) = \frac{\pi(1 - \pi)}{a + 1} = \pi(1 - \pi)\phi \quad \text{단, } \phi = \frac{1}{a + 1}$$

베타분포의 두 모수 중  $\pi$ 는  $p_i$ 의 기댓값이 되며,  $a$ 는  $p_i$ 의 분산을 결정하는데 쓰인다. 즉,  $a = \infty (\phi = 0)$ 이면 이항 분포( $B(n_i, \pi)$ )를,  $a = 0 (\phi = \infty)$ 이면  $\Pr(X_i = 0) = 1 - \pi$ ,  $\Pr(X_i = n_i) = \pi$ 인 퇴화된(degenerate) 분포를 따른다. Figure 1은 표본크기( $n$ )가 20이고  $\pi$ 가 0.1일 때  $a$ 가 각각 0.01, 20, 100일 경우의 베타-이항분포 확률에 관한 막대그림표로, 전술한 특징을 확인할 수 있다.

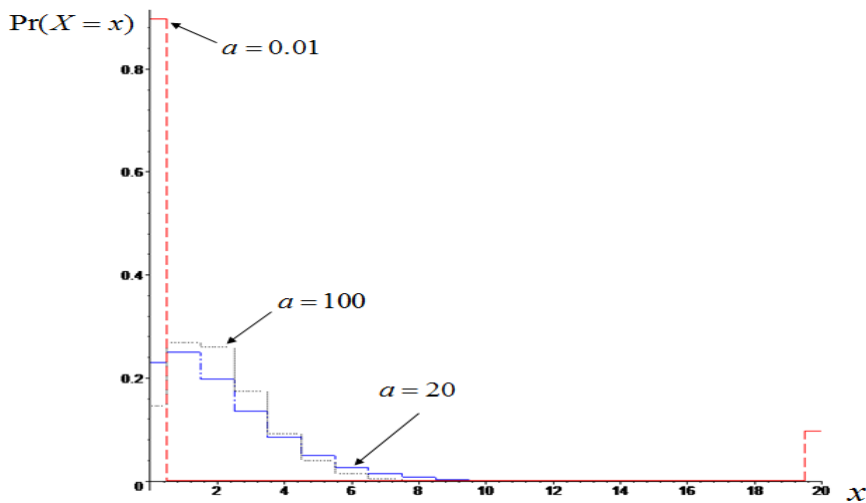


Figure 1. Beta-binomial distribution:  $n = 20, \pi = 0.1$

$X_i$ 의 주변분포는 다음과 같이 되므로

$$f_{X_i}(x) = \binom{n_i}{x_i} \frac{B(a\pi + x, a(1 - \pi) + n_i - x)}{B(a\pi, a(1 - \pi))}, \quad x = 0, 1, \dots, n_i \quad (4)$$

$X_i$ 의 기댓값과 분산은 다음과 같이 표현된다.

$$\begin{aligned} E(X_i) &= E_{p_i}[E(X_i | p_i)] = E_{p_i}(n_i p_i) = n_i \pi \\ \text{Var}(X_i) &= E_{p_i}[\text{Var}_{X_i | p_i}(X_i)] + \text{Var}_{p_i}[E_{X_i | p_i}(X_i)] \\ &= n_i \pi (1 - \pi) \left( 1 + \frac{n_i - 1}{a + 1} \right) \geq n_i \pi (1 - \pi) \end{aligned}$$

따라서  $\hat{p}_i = X_i/n_i$ 의 기댓값과 분산은 다음과 같이 주어지므로 불편성과 과산포를 충족한다.

$$\begin{aligned} E(\hat{p}_i) &= \pi \\ \text{Var}(\hat{p}_i) &= \sigma_{BB_i}^2 = \frac{\pi(1 - \pi)}{n_i} \left( 1 + \frac{n_i - 1}{a + 1} \right) \geq \frac{\pi(1 - \pi)}{n_i} \end{aligned} \quad (5)$$

여기서 베타-이항분포의 이항분포에 대한 분산 증가분은  $(n_i - 1)/(a + 1) = \phi(n_i - 1)$ 이므로  $a$ 와  $n_i$ 에 의존하며,  $a$ 가 커질수록 과산포가 감소한다.

이에 따라 베타-이항분포하의 관리도( $p(BB)$  관리도로 명명)의 관리한계는 식 (5)의  $\pi$ 에  $\bar{p}$ 를 대입한  $\hat{\sigma}_{BB_i}$ 로부터 다음과 같이 설정되며,  $LCL$ 이 0보다 작을 때는 0으로 설정된다.

$$UCL/LCL_{BB} = \bar{p} \pm 3\hat{\sigma}_{BB_i} \quad (6)$$

### 3.2 제안된 관리도의 특성

먼저 베타-이항분포를 따를 때 전통적인  $p$  관리도를 적용할 경우의 문제점을 검토해 보자.

$BB(230, a, 0.01)$ 를 따를 때  $B(230, 0.01)$ 를 따른다고 잘못 적용하여(표본크기( $n$ )인 230은 부적합 수  $X$ 가 이항분포를 따를 때  $\Pr(X \geq 1) \geq 0.9$ 이 되도록 설정한 값이며, 이를 기준으로  $n$ 인 100과 500을 추가한 분석 결과가 Figure 2와 같이 대체적 패턴이 유사하여 이 경우에 대해 대표적으로 예시함)  $p$  관리도의 두 관리한계 ( $UCL = 0.02968, LCL = 0$ )를 설정할 때, 참 분포가  $BB(230, a, 0.01)$ 일 경우에 관리한계를 벗어날 평균 부분군의 수(Average Run Length:  $ARL$ )을 구한 것이 Figure 2이다. 이를 보면 이항분포를 따를 경우의  $ARL$ 이 111.13(Table 1 참조)인데 반해, 참인 베타-이항분포에서  $a$ 가 작아질수록  $ARL$ 이 상당히 감소하고 있으므로 오경보 확률이 급격하게 증가됨을 알 수 있다.

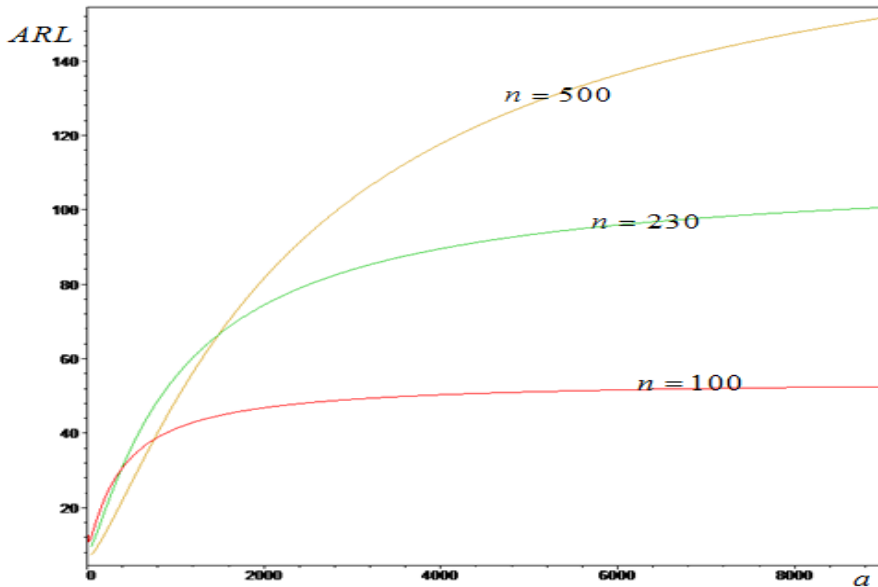


Figure 2. ARL of  $p$  control chart under beta-binomial distribution  
 :  $n = 230, p = \pi = 0.01$

Table 1에는 참 모형이 이항분포 또는 베타-이항분포를 따를 때  $n = 230, p$  또는  $\pi$ 가 0.01 하에 설계된  $p$  관리도 ( $LCL$ 은 모두 0이 됨)에서  $\pi$  또는  $p$ 에  $\delta$ 만큼 변화가 발생할 때  $UCL$ 을 벗어날 확률(%)이 정리되어 있다. 이를 보면 베타-이항분포를 따를 때 이항분포에 비해  $UCL$ 이 상당히 커지며, 두 분포에 관한  $a$ 에 따른 검출 확률이 일관된 패턴을 보이지는 않는다.

그리고 Table 2에는 참 모형이  $BB(230, a, 0.01)$ 인데, Figure 2와 같이  $B(230, 0.01)$ 로 상정하여 관리한계를 설정한 경우에  $p$ 의 변화에 따른  $UCL$ 를 벗어나는 확률로써  $p$  관리도의 수행도를 나타낸다. 여기서  $\delta = 0$ 인 경우의 해당 확률은 Figure 2와 동일한 경우이며, 참인 베타-이항모형을 이항분포로 가정하여  $p$  관리도를 적용할 경우에 경보 확률이 이항분포로 예상한 값이나 Table 1의 대응되는 값보다 상당히 큰 값을 가진다. 특히 이항분포로 상정하게 되면  $a$ 가 감소함에 따라(즉, 이항분포에서 멀어져 과산포가 커짐에 따라) 대체적으로 경보확률이 증가하는데,  $a$ 가 작은 경우에 이런 경향이 완화된거나 역전 현상도 발생된다. 또한 공정 변화의 양이 커짐에 따라 경보 확률이 증가되지만  $a$ 가 감소함에 따라 그 정도는 상대적으로 약화되며, 특히  $p$ 가 0.015의 이동할 경우는 예상된 값(Table 1의  $B(230, 0.01)$ 일 때)보다 작아지는 현상도 일부 발생된다.

Table 2와는 역으로 참 분포가 이항분포인데  $BB(230, a, 0.01)$ 를 따른다고 잘못 파악하여  $p(BB)$  관리도를 적용했을 때  $a$ 와  $\delta$ 에 따른  $UCL$ 을 벗어날 확률(%)이 Table 3에 정리되어 있다. 참인 이항모형을 베타-이항분포로 가정하여  $p(BB)$  관리도를 적용할 경우에,  $a$ 가 감소함에 따라  $UCL$ 이 커지므로 예상한 대로 경보확률이 작아지는 오류(즉, 검출력의 저하)가 발생하지만, 이항분포에 근접하는 경우인  $a$ 가 클 때(즉,  $a = 1,000$ ) 그 영향은 그리 크지 않다.

**Table 1.**  $\Pr(\hat{p} \geq UCL)$  (%) when  $UCL$  and  $LCL$  are correctly determined under binomial or beta-binomial model:  $n = 230, p = \pi = 0.01$

true model		$B(230, 0.01)$	$a$ of $BB(230, a, 0.1)$			
			1,000 (1.10850 <sup>1)</sup> )	100 (1.80757)	50 (2.34312)	20 (3.45033)
UCL		0.02968	0.03182	0.04558	0.05612	0.07791
shift ( $\delta$ )	+0	0.89825	0.68601	1.81037	2.43179	2.61304
	+0.002	2.20642	1.51054	2.63398	3.23839	3.26649
	+0.005	6.00704	3.82367	4.24477	4.66658	4.32699
	+0.01	18.00882	11.41137	8.01005	7.63576	6.30343
	+0.015	35.30948	23.64717	13.14225	11.32185	8.52962

Note: 1) Ratio of standard deviations under two models:  $\sigma(BB(230, a, 0.01))/\sigma(B(230, 0.01))$

**Table 2.**  $\Pr(\hat{p} \geq UCL | BB(n, a, \pi))$  (%) when  $UCL$  and  $LCL$  are falsely determined under binomial model:  $n = 230, p = \pi = 0.01$

true shift( $\delta$ )	$a$			
	1,000	100	50	20
+0	1.81427	7.93592	10.32607	11.20485
+0.002	3.60006	10.73894	13.07415	13.59134
+0.005	7.96509	15.59258	17.50043	17.21965
+0.01	19.83228	25.01538	25.45536	23.33026
+0.015	35.68968	35.37808	33.77057	29.42463

**Table 3.**  $\Pr(\hat{p} \geq UCL | B(n, p))$  (%) when  $UCL$  and  $LCL$  is falsely determined under beta-binomial model:  $n = 230, p = \pi = 0.01$

true shift( $\delta$ )	$a$			
	1,000	100	50	20
+0	0.24319	0.00253	0.00008	0.00000
+0.002	0.71122	0.01263	0.00055	0.00000
+0.005	2.38835	0.08104	0.00546	0.00000
+0.01	9.29402	0.71507	0.08452	0.00012
+0.015	22.02814	3.12918	0.56799	0.00242



### 4. 제안된 관리도의 적용 절차

3절에서 제안된 과산포 공정용 부적합품률 관리도를 적용하기 위해 다음과 같은 구체적 절차에 따라 수행하고자 한다.

- (i) 독립적인 베르누이 시행으로 간주할 수 있는 부적합품률 관리도를 적용할 수 있는 상황인지 파악한다.
- (ii) (i)의 적용상황일 때 배치간 부적합품률의 차이 또는 동질적으로 보기 힘든 매우 큰 표본크기를 가질 경우 적합도 검정에 의해 이항 또는 베타-이항모형을 따르는지를 검정한다. 여기서 베타-이항모형이 의심스러우면 Liang and McCullagh(1993)과 Garren et al.(2000) 등의 적합도 검정법을 적용한다. 특히 전자의 연구에서는  $n_i$ 와 잔차 $((r_i - n_i\bar{p})/\sqrt{n_i\bar{p}(1-\bar{p})})$ 를 도시하여, 잔차가  $n_i$ 에 따라 증가하면(식 (5) 참조) 베타-이항모형을 따르는 강한 증거임을 확인하는 도식적 방법을 제안하였다.
- (iii) 이항분포가 선택되면 전통적인  $p$  관리도를 적용하며, 베타-이항분포가 채택되면 이 분포의 모수  $a$ 를 추정한다.
- (iv) 베타-이항모형에 따른 관리한계를 설정하고 부분군의 부적합품률을 타점하여 관리상태 여부를 판정한다.

상기 절차를 따를 경우 과산포 여부에 관한 적합도 검정(goodness-of-fit test)과 베타-이항모형의 분포모수 추정 과정이 필요하다.

#### 4.1 적합도 검정

본 논문에서는 베타-이항분포 대한 이항분포의 검정법으로 Tarone(1979)의  $Z$  통계량을 채택한다. 이 검정법은 모수에 대한 사전추정이 요구되지 않으므로 베타-이항분포에서  $a = \infty$  ( $\phi = 0$ ) 일 때 최우추정법이 수렴하지 않을 위험을 회피할 수 있는 장점을 가지고 있으며, 높은 검정력과 계산상의 간편성을 가지고 있다(Ennis and Bi 1998; Capanu and Presnell 2008; Young-Xu and Chan 2008).

여기서 귀무가설은 다음과 같이 이항분포(베타-이항분포에서  $a = \infty$ )가 되며, 대립가설은 베타-이항분포가 된다.

$$\begin{aligned}
 H_0 : X_i &\sim BB(n_i, \infty, \pi) \\
 H_1 : X_i &\sim BB(n_i, a, \pi), \quad a \neq \infty
 \end{aligned}$$

이 때 검정통계량  $Z$ 는 각 부분군에서 얻은 표본정보(즉,  $m, n_i, r_i$ )에 의해 식 (7)이 되며 귀무가설 하에서 이 통계량은 점근적으로 표준정규분포를 따르는데,  $\phi > 0$ 이므로 단측검정을 적용한다.

$$Z = \frac{S - \sum_{i=1}^m n_i}{\sqrt{2 \sum_{i=1}^m n_i(n_i - 1)}} = \frac{\frac{N_0 R_1}{R_0} + \frac{N_0 Y_1}{Y_0} - N_1}{\sqrt{2N_1}} \tag{7}$$

$$\begin{aligned} \text{여기서, } S &= \frac{\sum_{i=1}^m (r_i - n_i \bar{p})^2}{\bar{p}(1 - \bar{p})} \\ R_0 &= \sum_{i=1}^m r_i, \quad Y_0 = \sum_{i=1}^m (n_i - r_i), \quad N_0 = \sum_{i=1}^m n_i \\ R_1 &= \sum_{i=1}^m r_i(r_i - 1), \quad Y_1 = \sum_{i=1}^m (n_i - r_i)(n_i - r_i - 1), \quad N_1 = \sum_{i=1}^m n_i(n_i - 1) \end{aligned}$$

### 4.2 모수추정

베타-이항분포의 두 모수  $\pi, a$ 의 추정법으로 최우추정법과 적률법이 널리 쓰인다(Kleinman 1973; Young-Xu and Chan 2008).

먼저 통계적으로 우수한 성질을 가진 최우추정량을 구하기 위한 대수우도함수는 다음과 같이 표현된다.

$$\begin{aligned} l(\pi, a) &= \sum_{i=1}^m \left[ \sum_{j1=0}^{r_i-1} \ln(a\pi_i + j1) + \sum_{j2=0}^{n_i-r_i-1} \ln(a(1-\pi_i) + j2) - \sum_{j3=0}^{n_i-1} \ln(a + j3) \right] \\ &= \sum_{i=1}^m [(\ln\Gamma(a\pi + r_i) - \ln\Gamma(a\pi)) + (\ln\Gamma(a(1-\pi) + n_i - r_i) - \ln\Gamma(a(1-\pi))) - (\ln\Gamma(a + n_i) - \ln\Gamma(a))] \\ &= \sum_{i=1}^m [d\lg(r_i, a\pi) + d\lg(n_i - r_i, a(1-\pi)) - d\lg(n_i, a)] \end{aligned} \tag{8}$$

여기서  $d\lg(y, b) = \ln\Gamma(b + y) - \ln\Gamma(b)$ 임.

대수우도함수를 두 모수  $\pi, a$ 에 대해 편미분하여 0으로 둔 다음의 두 방정식의 해가 최우추정값 ( $\hat{\pi}, \hat{a}$ )이 되는 데, 이를 구하기 위해서는 수치해법이 필요하다. 여기서  $0 < \pi < 1$ 이므로 수치해법을 적용할 때  $\ln(\pi/(1-\pi))$ 로 변환하여 해를 구한 후에 역변환하는 과정을 추천한다.

$$\frac{\partial l(\pi, a)}{\partial a} = \sum_{i=1}^m [\pi \cdot ddg(r_i, a\pi) + (1-\pi) \cdot ddg(n_i - r_i, a(1-\pi)) - ddg(n_i, a)] = 0 \tag{9}$$

$$\frac{\partial l(\pi, a)}{\partial \pi} = \sum_{i=1}^m [a \cdot ddg(r_i, a\pi) - a \cdot ddg(n_i - r_i, a(1-\pi))] = 0 \tag{10}$$

여기서  $ddg(y, b) = \frac{\partial d\lg(y, b)}{\partial b} = \psi(y + b) - \psi(b)$ , 단,  $\psi(\cdot)$ 는 di-gamma 함수임.

적률법은 부분군의 크기가 동일하지 않을 경우는 반복적인 계산 절차가 요구되어(Kleinman 1973) 최우추정법보다 계산상의 이점이 높지 않으므로, 부분군의 크기가 동일한 경우(모든  $i$ 에 대해  $n_i = n$ )에 한정하여 다음과 같이 적용할 수 있다.

$$\tilde{\pi} = \frac{R_0}{N_0} = \bar{p} \quad (11)$$

$$\tilde{a} = \max\left(\left[\frac{N_0}{N_0 - 1} \frac{\hat{Q}}{\bar{p}(1 - \bar{p})(m - 1)} - \frac{1}{N_0 - 1}\right]^{-1} - 1, 0\right) \quad (12)$$

$$\text{여기서 } \hat{Q} = \sum_{i=1}^m \left(\frac{r_i}{n_i} - \bar{p}\right)^2 \text{ 임.}$$

한편 식 (12)에서  $m-1$  대신에  $m$ 이 사용되기도 한다(Young-Xu and Chan 2008).

## 5. 제안된 관리도의 활용

이 절에서는 기존 연구에서 다룬 다음의 3가지 예제(동일한 부분군 크기일 경우 2종, 동일하지 않는 부분군 크기일 경우 1종)을 대상으로  $p$ ,  $I(p)$ ,  $p'$ ,  $p(BB)$  관리도를 적용한 결과를 비교하고 고찰한다. 본 논문에서 제안된 관리도 외는 Minitab(2014)을 이용하여 분석하였다.

### 5.1 예제 1

Jones and Govindaraju(2000)에서 다룬 부적합품 수에 대한 예제로 부분군의 크기가 동일(즉, 50개)하며, 부분군의 수는 30개이다. 전통적인  $p$  관리도의 관리한계를 설정하는데 필요한  $\bar{p}=0.2313$ 이며, 관리한계와 각 군의 부적합품을 타점한 관리도가 Figure 3(a)이다. 이를 보면 두 점이 관리한계를 벗어나고 있다.

본 논문에서 제안된  $p(BB)$  관리도를 적용하기 위해 먼저 식 (7)의 Tarone의 적합도 검정 통계량과  $p$ 값을 구하면 각각 7.226과 0으로  $p$  관리도를 적용할 수 없는 대상임을 확인하였다. 따라서 베타-이항 모형을 적용하기 위해 추정된 두 모수의 최우추정값은  $\hat{\pi}=0.2316$ ,  $\hat{a}=27.290$ 가 되며, 식 (6)에 의한 관리한계와 부분군의 부적합품을 타점한 관리도가 Figure 3(b)이다. 이를 보면 과산포(이항분포를 가정할 경우의 표준편차의 1.653배)를 반영한 관리한계(0과 0.5171)가 넓게 설정되어 이를 벗어나는 점이 하나도 없으며,  $LCL$ 은 0으로 설정된다.

같은 데이터에 대한 Figure 4의  $I(p)$  관리도와 Laney  $p'$  관리도의 관리한계는 전자의  $LCL$ 을 음수가 아닌 0으로 취하면 전술된 바와 같이 부분군의 크기가 동일하여 같아짐을 확인할 수 있으며,  $p(BB)$  관리도의  $UCL$ 도 거의 유사한 값(즉, 0.5265)을 가진다.  $p'$  관리도의 과산포 값은 170.1%로 유의수준 5%에서의 임계값 148.6%(식 (3))을 초과하여 과산포로 판정하고 있다.

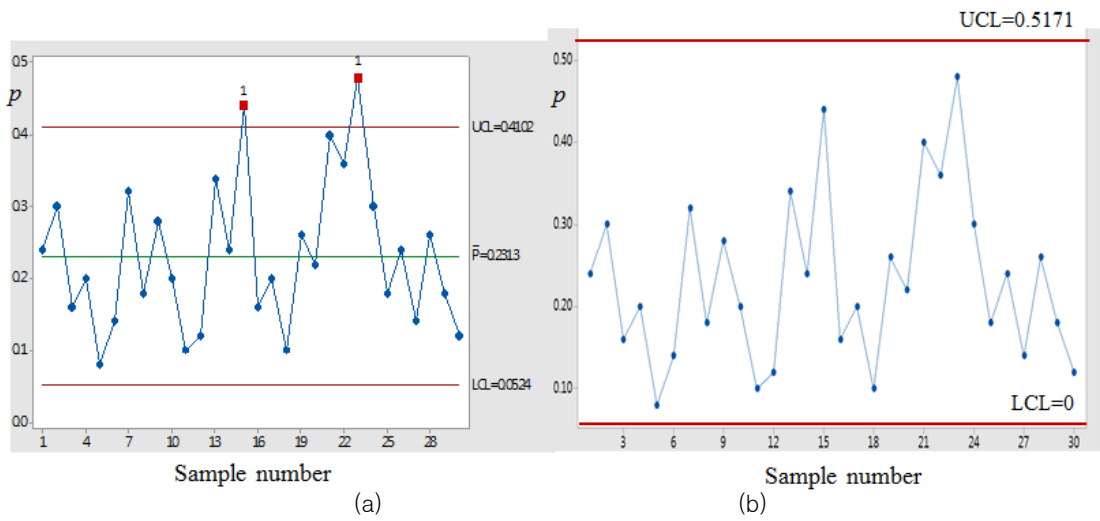


Figure 3.  $p$  control chart(a) and  $p(BB)$  control chart(b): Example 1

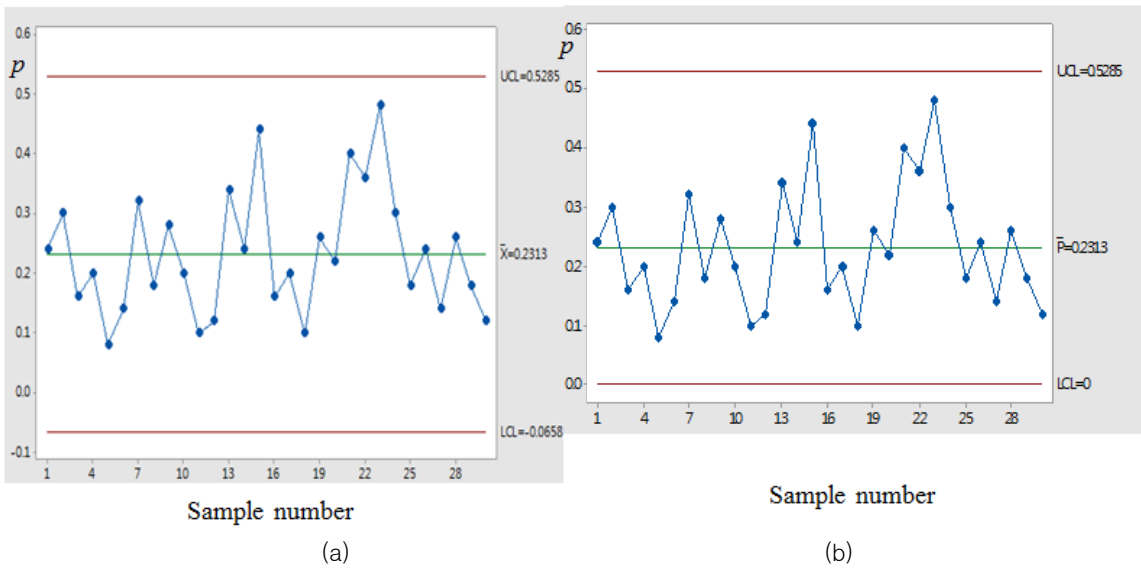


Figure 4.  $I(p)$  control chart (a) and  $p'$  control chart (b): Example 1

## 5.2 예제 2

Young-Xu and Chan(2011)에 인용된 데이터로 41 처리군의 경구용 항진균제의 부작용으로 임상시험을 중단한 환자 수에 관한 기록이며(여기서  $\bar{n}$ 는 73.22게임), 이 데이터는 Figure 5와 같이 4절에서 정의된 잔차가 부분군의 크기에 따라 증가하고 있으므로 베타-이항분포에 잘 적합된다고 볼 수 있다. 또한 식 (7)의 Tarone의 검정 통계량과  $p$ 값은 각각 7.484와 0으로  $p(BB)$  관리도를 적용해야 되는 상황임을 확인하였다.

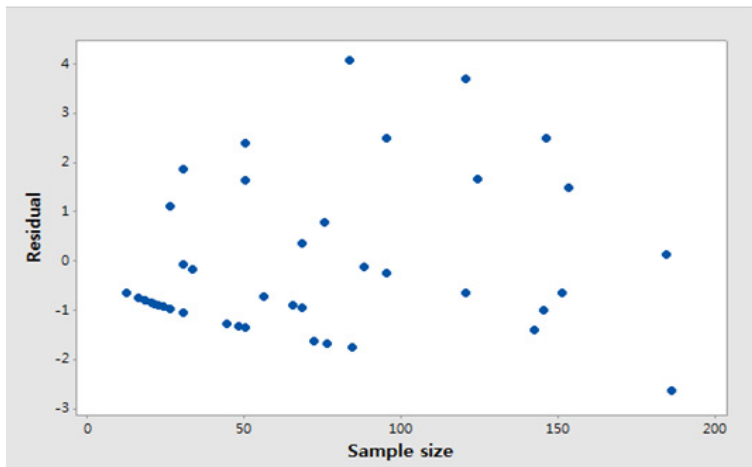
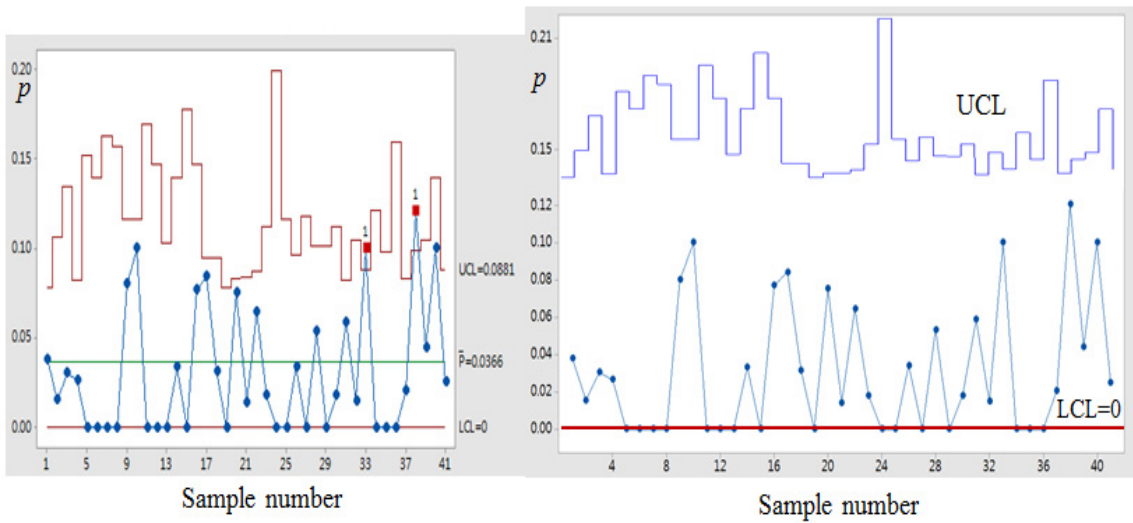


Figure 5. Residuals plotted against sample size: Example 2

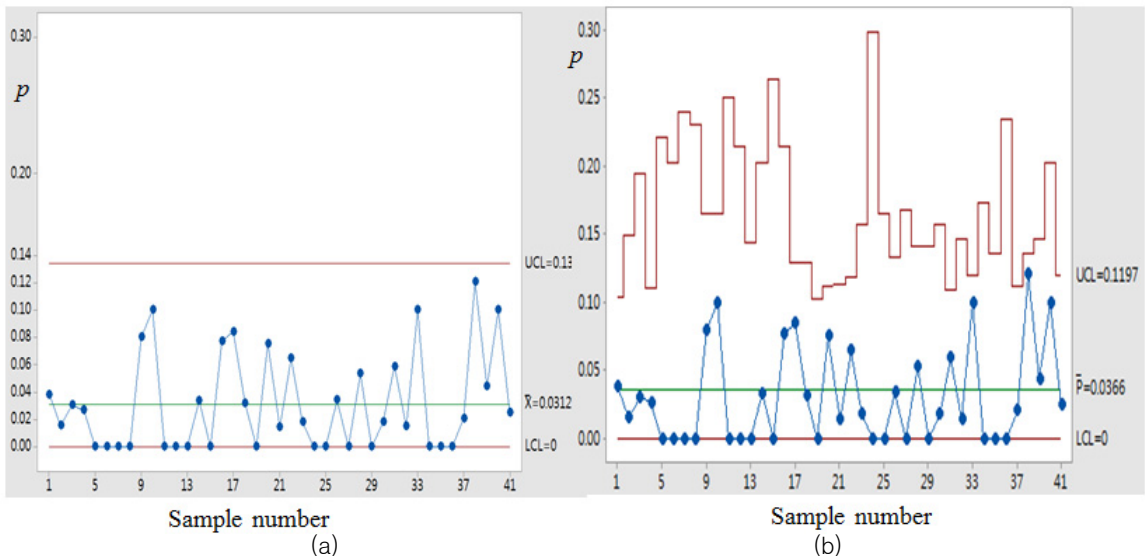
먼저  $\bar{p}=0.03664$ 로부터 구한 전통적인  $p$  관리도의 관리한계와 각 부분군의 부적합품률을 타점한 것이 Figure 6(a)로 관리한계를 벗어난 점이 2개로 이상상태로 판정된다.

과산포를 반영한  $p(BB)$  관리도 작성에 필요한 두 모수의 최우추정값은  $\hat{\pi}=0.03420$ ,  $\hat{a}=37.099$ 이고, 베타-이항분포를 적용한 부적합품률은 3.420%는 이항분포를 적용한 부적합품률보다 약 6.7% 작은 값이 도출되며, 표준편차는 이항분포의 1.702배( $\bar{n}$  기준)로 파악되었다. Figure 6(b)를 보면  $p(BB)$  관리도에서는 전통적인  $p$  관리도와 달리 관리한계를 벗어난 점이 하나도 없는 관리상태를 나타내고 있다.



(a) (b)  
**Figure 6.**  $p$  control chart(a) and  $p(BB)$  control chart(b): Example 2

동일한 데이터에 대해 작성된 Figure 7(a)의  $I(p)$  관리도는 다른 관리도와 달리  $UCL$ 에 부분군의 크기가 반영되지 않아( $LCL$ 은 0으로 설정함) 이를 감안하는 추가 조정과정이 요구된다. 한편 Laney 판정된다. Figure 7(b)의  $p'$  관리도에서는 부분군의 크기에 따라 두 관리한계가 변하며, 관리한계에 적시된 수치는  $\bar{n}$ 로 계산된 관리한계 값이고, 이 관리도는 과산포를 반영하여 있어 역시 관리한계를 벗어나는 점이 하나도 없다.  $p'$  관리도에서 과산포 값은 194.7%로 유의수준 5%의 임계값 152.8%(식 (3))을 초과하여 과산포로 판정된다. Figure 7(b)의  $p'$  관리도에서는 부분군의 크기에 따라 두 관리한계가 변하며, 관리한계에 적시된 수치는  $\bar{n}$ 로 계산된 관리한계 값이고, 이 관리도는 과산포를 반영하여 있어 역시 관리한계를 벗어나는 점이 하나도 없다.



(a) (b)  
**Figure 7.**  $I(p)$  control chart(a) and  $p'$  control chart(b): Example 2

### 5.3 예제 3

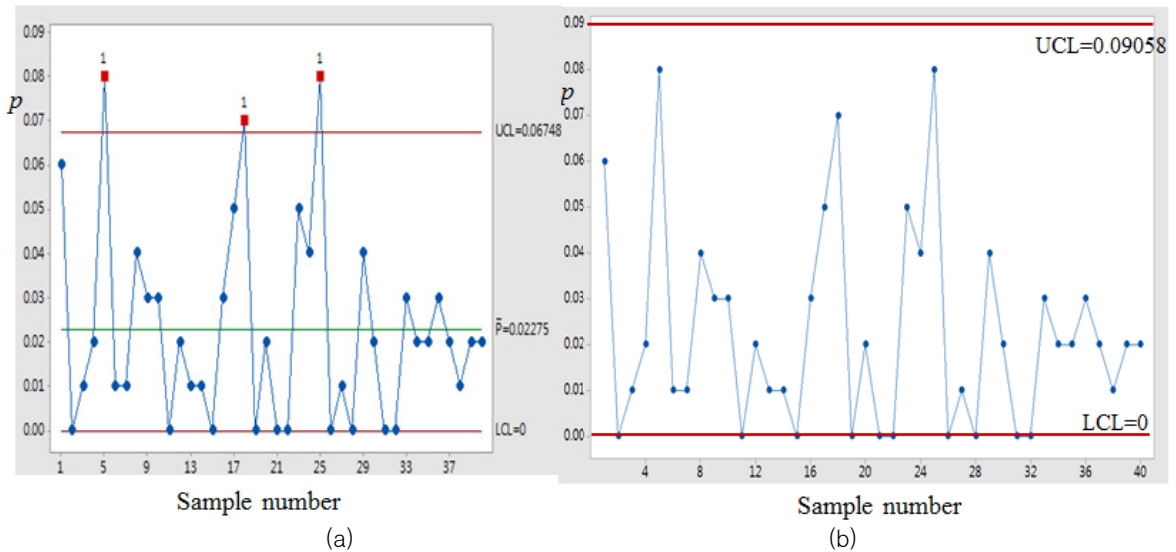
부분군의 크기가 100으로 동일하고  $\pi$ 가 0.2일 때 과산포를 반영한 분산 비(식(5)의 괄호 부분)의 값이 2가 되도록 설정하면  $a$ 는 98이 된다. 즉, 베타-이항모형이  $BB(100, 98, 0.02)$ 를 따를 경우의 확률 변량(부적합품 수)을 40개 부분군에 대해 발생시킨 특정 결과가 Table 4에 수록되어 있다.

**Table 4.** Simulated data( $r_i$ ) from  $BB(100, 98, 0.02)$ :  $n_i = 100, i = 1, 2, \dots, 40$

6, 0, 1, 2, 8, 1, 1, 4, 3, 3, 0, 2, 1, 1, 0, 3, 5, 7, 0, 2
0, 0, 5, 4, 8, 0, 1, 0, 4, 2, 0, 0, 3, 2, 2, 3, 2, 1, 2, 2

상기 데이터에서  $\bar{p}=0.02275$ 이므로 전통적인  $p$  관리도의 관리한계와 더불어 부적합품률을 타점한 관리도가 Figure 8(a)이며, 이를 보면 세 점이 관리한계를 벗어나고 있다. 이에 따라 과산포 여부를 판정하기 위해 구한 식 (7)의 Tarone의 적합도 검정 통계량과  $p$ 값은 각각 5.208과 0으로 전통적  $p$  관리도를 적용할 수 없는 대상임을 확인하였다.

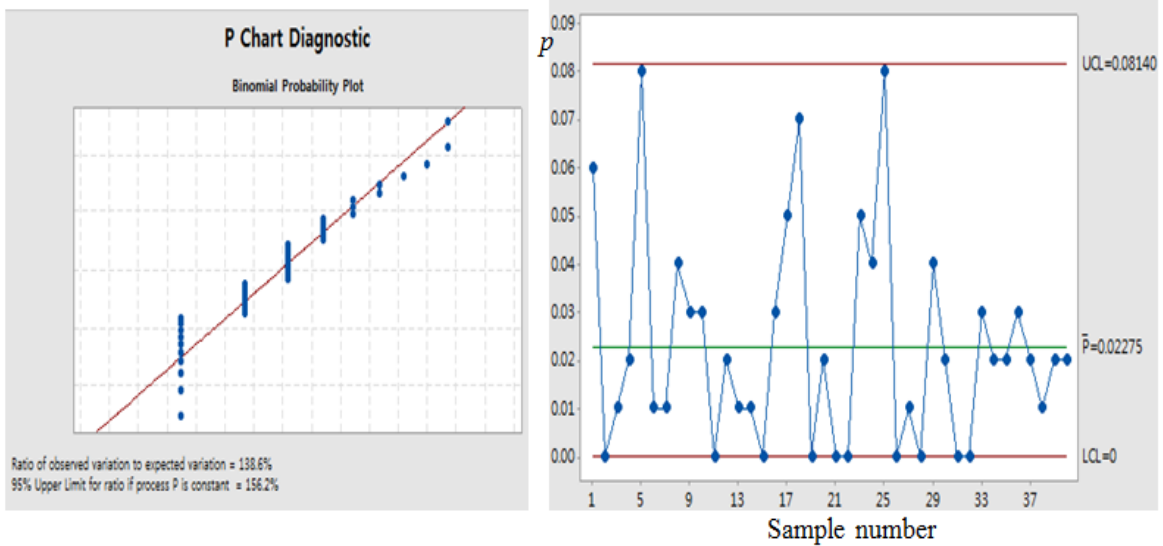
따라서 과산포용  $p(BB)$  관리도를 적용하기 위해 추정된 두 모수의 최우추정값은  $\hat{\pi}=0.02274, \hat{a}=75.117$ 이 되며, 식 (6)으로부터 구한 관리한계(0과 0.09058)와 부분군의 부적합품률을 타점한 관리도가 Figure 8(b)이다. 이를 보면 전통적인  $p$  관리도와 달리 과산포(이항분포를 가정할 경우의 표준편차의 1.517배)를 반영한 관리한계가 넓게 설정되어 이를 벗어나는 점이 하나도 없다.



**Figure 8.**  $p$  control chart(a) and  $p(BB)$  control chart(b): Example 3

부분군의 크기가 같으므로 동일한 관리한계를 가지는  $I(p)$  관리도는 제외하고 Laney  $p'$  관리도만 도시하였다. Figure 9(a)(Minitab 출력물)를 보면 과산포 값은 138.6%로 유의수준 5%에서의 임계값 156.2%(식 (3))보다 작아 Tarone의 적합도 검정과는 달리 과산포를 검출하지 못하고 있다. 따라서 이 데이터에 대해서는 Figure 9(b)에 도시

된  $p'$  관리도의 채택을 추천하지 않고 있으며, 또한  $p(BB)$  관리도와 달리 두 점이 관리한계에 근접해 있어 이상상태에 대한 경고신호로 볼 수 있다.



(a) Diagnostic results (Jones and Govindaraju method) (b) Control chart

Figure 9.  $p'$  control chart: Example 3

## 6. 결 론

각 부분군의 부적합품률을 타점하는  $p$  관리도는 부적합 개수가 이항분포를 따를 경우에 적용하는데, 모수인 각 부분군의 부적합품률이 일정하며 독립이어야 한다. 표본크기가 너무 크거나 부분군간 부적합품률에 변동이 있을 경우 이런 조건을 충족하지 않으므로 과산포(overdispersion)인 경우를 자주 접할 수 있다. 이럴 경우 전통적인  $p$  관리도를 적용하면 오경보 확률이 증가하므로  $p$  관리도를 대체할 수 있는 여러 관리도가 개발되어 있다. 하지만 이들 관리도는 부적합품률을 계량치로 취급하여 정규분포를 따른다고 가정하며, 특히 요즈음 현업에서 비교적 활용도가 높은 Laney(2002)의  $p'$  관리도는 구체적인 통계적 모형에 기반하지 않는 발견적인 방법에 속하여 이론적 근거가 부족한 편이다.

따라서 본 논문은 이항분포를 따르면서 부적합품률의 변동을 고려한 베타-이항모형 하에서 과산포를 반영한 관리한계를 설정하고 이를 체계적으로 활용할 수 있는 절차를 개발하였다. 또한 제한된 베타-이항분포 하에서 개발된 부적합품률 관리도의 유용성을 보여 주기 위해 세 가지 예제에 적용하였다. 이를 통해 제한된 관리도가 과산포일 경우  $p$  관리도의 대안이 될 수 있음을 예시하였다. 또한 비교적 활용도가 높은 두 종의 과산포용 관리도와도 예제를 통해 비교하였다.

앞으로 본 논문에서 제안한 과산포용 관리도에서 관리한계의 계수를 3(식 (6))으로 택하는 슈하트 관리도와 달리 다른 값으로 설정하는 관리도의 설계에 관한 연구(Table 1의 관리상한과 검출력 참조)와 더불어,  $p$  관리도 외의 다른 과산포용 계수형 관리도에 적용할 수 있는 구체적 방법론과 다양한 베타-이항분포 하에서 시뮬레이션을 통해 제안된  $p$  관리도의 수행도와 특성을 기존의 방법론과 정밀하게 비교·평가하는 추가 연구가 필요하다.



## REFERENCES

- Capanu, Marinela, and Presnell, Brett. 2008. "Misspecification Test for Binomial and Beta-Binomial Models." *Statistical in Medicine* 27:2536-54.
- Cryer, Jonathan D., and Ryan, Thomas P. 1990. "The estimation of sigma for an  $\bar{X}$  chart:  $MR/d_2$  or  $S/c_4$ ?" *Journal of Quality Technology* 22:187-92.
- Ennis, Daniel M., and Bi, Jian. 1998. "The Beta-Binomial Model: Accounting for Intertrial Variation in Replicated Difference and Preference Tests." *Journal of Sensory Studies* 13: 389-412.
- Garren, Steven T., Smith, Richard L., and Piegorsch, Walter W. 2000. "On a Likelihood-Based Goodness-of-Fit Test of the Beta-Binomial Model." *Biometrics* 56:947-949.
- Heimann, Peter A. 1996. "Attributes Control Charts with Large Sample Sizes." *Journal of Quality Technology* 28:451-9.
- Jones, G. and Govindaraju, K. 2000. "Graphical Method for Checking Attribute Control Chart Assumptions." *Quality Engineering* 13:19-26.
- Kleinman, Joel C. 1973. "Proportions with Extraneous Variance: Single and Independent Samples." *Journal of American Statistical Association* 68:46-54.
- Laney, David B. 2002. "Improved Control Charts for Attributes." *Quality Engineering* 14:531- 7.
- Liang, Kung-Yee, and Mccullagh, Peter. 1993. "Case Studies in Binary Dispersion." *Biometrics* 49:623-630.
- Minitab. 2014. "Minitab 17 Statistical Software." Accessed September 2, 2015. <http://www.minitab.com>.
- Mohammed, Mohammed A., and Worthington, Peter. 2013. "Why Traditional Statistical Process Control Charts for Attribute Data Should Be Viewed Alongside an xmr-Chart." *BMJ Quality and Safety* 22:263-9.
- Mohammed, Mohammed A., Panesar, Jagdeep S., Laney, David. B., and Wilson Richard. 2013. "Statistical Process Control Charts for Attribute Data Involving Very Large Sample Sizes: a Review of Problems and Solutions." *BMJ Quality and Safety* 22:362-68.
- Mohammed, Mohammed A., Worthington, Peter, and Woodall, William H. 2008. "Plotting Basic Control Charts: Tutorial Notes for Healthcare Practitioners." *BMJ Quality and Safety* 17: 137-145.
- Moore Dirk F. 1987. "Modeling the Extraneous Variance in the Presence of Extra-Binomial Variation." *Applied Statistics* 36:8-14.
- Quarshie, B. L., and Shindo, Hisakazu. 1996. "A Comparison of Operating Characteristics of the  $\hat{p}-V$  and the  $\bar{p}-R_p$  Charts." *Quality Engineering* 9:221-8.
- Ryan, Thomas P. 2011. *Statistical Methods for Quality Improvement*, 3rd ed. New Jersey(USA): John Wiley & Son.
- Raubenheimer, L., and van der Merwe, A. J. 2016. "Bayesian Process Control for the Phase II Shewart-Type  $p$ -Chart." *Quality Technology and Quantitative Management* 13:453-72.
- Tarone, Robert E. 1979. "Testing the Goodness of Fit of the Binomial Distribution." *Biometrika* 66:585-590.
- Wheeler, Donald J. 2001. *Understanding Statistical Process Control*, 3rd ed., Knoxville(USA): SPC Press.
- Woodall, William H. 1997. "Control Charts Based on Attribute Data: Bibliography and Review." *Journal of Quality Technology* 29:172-183.
- Young-Xu, Yinong, and Chan, K. Arnold. 2008. "Pooling Overdispersed Binomial Data to Estimate Event Rate." *BMC Medical Research Methodology* 8: Accessed November 22, 2016. doi:10.1186/1471-2288-8-58.