

## Cluster Sampling in Sampling Inspection: Bayes Estimation<sup>1)</sup>

Minwoong Shin<sup>2)</sup> and Juyoung Lee<sup>3)</sup>

### Abstract

We propose a sample design which minimize Bayes risk for cluster sampling in sampling inspection.

We treat a pilot sample and an additional sample size as random variable. In addition, we compute an appropriate cluster size for handling over-dispersion.

### 1. 서 론

샘플링검사에서는 계수규준형 1회 샘플링검사와 계수규준형 2회 샘플링검사가 많이 이용되고 있다. 그런데 동일한 크기의 표본으로 샘플링검사를 할 때에 다음과 같이 1회보다는 2회 샘플링검사가 유리하다.

계수규준형 1회 샘플링검사는  $n$ 개의 시료를 1회 검사하여 로트의 합격 여부를 판정한다. 계수규준형 2회 샘플링검사는  $n_1$ 개의 시료(예비표본)를 검사한 후 로트의 합격 여부를 판정하지만, 판정이 나지 않은 경우는  $n_2$ 개의 두번째 시료(추가표본)를 추출하여 검사한다. 여기  $n_1 + n_2 = n$ 이다. 박 과 박(1995, p320)에 의하면 2회 샘플링검사는 1회 샘플링검사와 비교하여 볼 때 검사량의 절감을 기할 수 있다는 것과, 2회에 걸쳐서 로트를 판정하는 데서 오는 심리적 효과를 기대할 수 있다는 장점을 가지고 있다.

이 논문은 샘플링 검사(sampling inspection)에서 부-로트(sub-lot)를 집락으로 간주하여 일단 집락 추출을 하는 문제를 다룬다. 예컨대, 집락이 생산라인(line)으로 부터 출하되는 성분들의 상자(box)들로 구성되었다고 가정한다. 즉, 하나의 집락은 한 라인으로부터의 성분들로부터 이루어진다. 여기서, 모든 라인들이 근사적으로 같은 비율로 불량품을 생산한다고 가정한다. 산림조사에서 구획(plot)을 만들어 조사 할 때에 적절한 집락의 크기를 선택하는 문제도 유사하다. 박 과 박(1995)에 의하면 모집단은 여러개의 집락으로 만들어 집락표본 추출을 할 때에 집락간 분산이 작아 질수록 집락 샘플링이 유리하고 집락간 분산이 커지면 랜덤 샘플링보다 효율이 떨어짐을 지적하였다. 그러나, 샘플링 비용을 고려하면 집락 샘플링이 유리한 경우가 많고, 집락수가 증가하면 운반, 포장비용이 증가한다. 이 연구의 목적은 모집단을 집락화 하는 과정에 있어서 집락의 크기에 따라 과대 산포(over-dispersion) 문제가 일어나므로 먼저 집락의 크기를 결정하는 문제를 다루고 주어진 오차 범위 내에서 모수를 포함한 기대되는 포함확률(expected coverage probability)

1) This paper is supported by " Korea university professor association research fund, 1997"

2) Professor, Department of Statistics, HanKuk University of Foreign Studies, Kyonggi, 449-791, Korea

3) Ph.D. candidate, Department. of Statistics, Inha University, 402-751, Korea.

이 적어도  $1-\alpha$ 가 되도록 표본의 크기  $n$ 을 정한 후, 예비표본의 크기가  $n_1$ , 추가표본의 크기가  $n_2$ 라 할 때,  $n_1 + n_2 = n$ 인 조건하에서,  $n_1$ 과  $n_2$ 를 모두 변수로 간주하여 베イズ 위험을 최소로 하는  $n_1$ 과  $n_2$ 를 결정하는데 있다.

지금까지 연구된 바로는 Miller와 Freund(1977)가 이항분포에서 모수  $p$ 를 추정하는데 있어서 주어진 예비표본( $n_1$ )으로부터  $p$ 의 추정치  $\hat{p}_1$ 을 구한 후, 이 추정치를 이용하여 추가표본의 크기( $n_2$ )를 결정하였다. Cohen과 Sackrowitz(1984a)는 지수족 분포의 평균에 대한 이-단 베이지안 추정(double sample Bayes estimation)을 하였다. 그 과정은 예비표본의 크기가  $n_1$ 인 자료  $X_1$ 로부터  $n_2(X_1)$ 을 결정한다. 그들은 더욱이  $n_1$ 과  $n_2$ 를 확률변수로 보아 위험(risk)를 최소로 하는 값을 구하였다.

Adcock(1987)은 다항표본추출에 대한 표본의 크기를 구하는데 베이지안 접근법을 사용하였다. Adcock(1988)은 표본의 크기를 구하는데 이항 모비율에 대한 사전 베타분포의 여러 가지 기준을 세워 베이지안 접근법을 사용하였다. Adcock(1992)은 이항과 다항표본추출에서 표본의 크기를 결정하는데 여러 가지의 베이지안 접근법을 비교하였다. Pham-Gia와 Turkan(1992)은 이항모수가 베타 사전분포를 할 때 사후분산을 구하여 표본크기를 결정하는데 이용하였다. Josep(1995)은 베이지안 접근방법으로 사후밀도 함수(HPD)의 관점에서 표본의 크기를 결정하였다. 신과 이(1997)는 집락 표본 추출에서 이-단계로 표본을 추출하는 문제를 다루었다.

본 논문은 2절에서는 과대 산포가 일어나지 않도록 부-로트의 크기  $M$ 을 정하는 문제를 다루고 3절에서는 주어진 오차의 범위 내에서 모수를 포함한 기대되는 확률이 적어도  $1-\alpha$ 가 되도록 표본의 크기  $n$ 을 정한다. 4절에서는 예비표본의 크기와 추가표본의 크기를 적절히 정하여 베イズ 위험이 최소가 되도록 결정한다.

## 2. 집락 부-로트의 크기 $M$ 의 결정

과대 산포는 어떤 변수의 분산이 명목(nominal)분산을 초과하는 것을 의미한다. 과대 산포는 실제로 드물지 않게 일어난다. 오히려 과대 산포가 흔히 일어나고 명목산포가 예외적으로 일어난다고 할수 있다. 과대 산포는 여러 원인으로 일어난다. 가장 단순하고, 일반적인 원인은 모집단에 있어서 집락화(clustering)에 있다. 그리고, 이러한 과대 산포를 무시했을 때는 모수 추정치에 대한 분산을 과소 추정하고 이로 인하여 가설 검정에서 낮은 검정력을 갖게 된다. 가족, 가구, 마을은 모집단에서 자연스럽게 일어나는 집락들이다.

모집단에서  $m$ 개의 원소를 추출하는데 단순 랜덤 추출로  $m$ 를 추출하거나, 모집단을 크기가  $M$ 인 집락들로 집락화 하여  $m$ 개의 원소를 추출 할 수 있다. 즉, 집락의 크기가  $M$ 이라 하고  $m$ 개의 원소들을  $m/M$ 개의 집락들로 부터 추출한다.  $i$ 번째 집락에서 양의 응답(positive respondents)의 수,  $Z_i$ 가 모수  $M$ 와  $p_i$ 를 갖는 이항분포를 하고,  $p_i \sim \text{beta}(\alpha, \beta)$ ,  $\alpha > 0, \beta > 0$ 이라고 가정한다. 그러면, 총 양의 응답 수는

$$Y = Z_1 + Z_2 + \cdots + Z_{m/M}$$

이다.

$E(p_i) = p$  그리고,  $Var(p_i) = \rho p(1-p)$  라 하면,  $Y$ 의 무조건 평균과 분산은

$$E(Y) = mp \quad (2.1)$$

$$Var(Y) = mp(1-p)\{1 + (M-1)\rho\} \quad (2.2)$$

이다. 산포모수  $1 + (M-1)\rho$ 은 집락크기와 집락으로부터 집락으로의  $p$ 의 변이(variability)에 의존한다. 여기서, Moor(1987)에 의하면  $\rho = 1/(\alpha + \beta + 1)$ 이다.

부-로트의 크기  $M$ 에 따라서 과대 산포의 크기는 결정되므로 적절한  $M$ 을 정해야 한다. 우리는 이미 지금까지 생성된 제품에서 각 라인으로부터 생산된 제품들의 불량률  $p_i$  ( $i$ 번째 라인으로부터 생산된  $i$ 번째 집락의 불량률)가 모수  $\alpha$ 와  $\beta$ 를 갖는 베타분포를 한다는 것을 알고 있다고 가정한다. 이제 새로운 공법에 의하여 생산되는 제품의 불량률을 알기 위하여 표본 설계를 하는 과정에서 부-로트의 크기  $M$ 을 과거의 자료를 이용하여 구한다. 특히 추가 변동(extra-variation)이 이항 변동(binomial variation)에 비하여 무시할수 있도록  $M$ 을 정한다.

### 3. 집락(부-로트)의 수 $n$ 의 결정

$X_{ij}$ 는  $i$ 번째 부-로트의  $j$ 번째 원소이다. 확률변수  $X_{ij}$ ,  $i=1, 2, \dots, j=1, 2, \dots, M$   $p$ 를 갖는 베르누이 분포를 한다고 하자. 그런데, 모집단을 집락화 하는 과정에서 집락(부-로트)의 크기  $M$ 을 적절히 정함으로서 과대 산포가 일어나지 않게한다. 부로트의 크기  $M$ 이 정해진 다음에는 오차의 범위 내에서 기대되는 포함확률이 적어도  $1-\alpha$ 가 되도록 부-로트의 수  $n$ 을 정한다.

$f(p)$ 는  $p$ 의 사전분포,  $n$ 개 집락의 자료는  $x = (x_1, x_2, \dots, x_n)$ , 그리고  $x_i = (x_{i1}, x_{i2}, \dots, x_{iM})$  이라고 하자.  $x$ 의 사전사후 주변분포는

$$f(x) = \int f(x|p)f(p)dp \quad (3.1)$$

이고, 자료  $x$ 가 주어졌을 때  $p$ 의 사후 분포  $f(p|x)$ 는

$$f(p|x, n) = \frac{f(x|p)f(p)}{\int f(x|p)f(p)dp} \quad (3.2)$$

이다. 우리는 HPD 구간 길이  $l$ 이 주어졌을 대에 기대포함확률이 적어도  $1-\alpha$ 인 최소의 표본크기  $n$ 을 구한다.

우리는 Joseph et al (1995)와 유사하게 다음식을 만족 시키는 최소의  $n$ 을 구한다.

$$\sum \Pr \{ p \in (a(x, nM), a(x, nM) + l) \} p(x, nM) \geq 1 - \alpha \tag{3.3}$$

여기서

$$\Pr \{ p \in (a(x, nM), a(x, nM) + l) \} \propto \int_{a(x, nM)}^{a(x, nM) + l} p^x (1-p)^{n-x} f(p) dp$$

이다.  $a(x, nM)$ 은 HPD의 하한이고,  $p(x, n)$ 은 자료의 사전사후 확률함수이다. 만일  $f(p)$ 가 모수  $(\alpha, \beta)$ 를 갖는 베타분포라고 하면, 즉

$$f(p) = \frac{1}{B(\alpha, \beta)} p^{\alpha-1} (1-p)^{\beta-1}, \quad 0 < p < 1$$

이다. 여기서  $B(\alpha, \beta)$ 는 모수  $(\alpha, \beta)$ 를 갖는 베타함수이다. 그러면

$$f(p | x, nM, \alpha, \beta) = \frac{1}{B(x + \alpha, nM - x + \beta)} p^{x + \alpha - 1} (1-p)^{nM - x + \beta - 1}, \quad 0 < p < 1 \tag{3.4}$$

는 데이터  $x$ 가 주어졌을 때 사후분포이고

$$p(x, n) = \binom{nM}{x} B(x + \alpha, nM - x + \beta - 1) / B(\alpha, \beta) \tag{3.5}$$

는  $x$ 에 대한 사전 사후분포이다. 그러면, (3.5)에서

$$\sum_{x=0}^{nM} \binom{nM}{x} / B(\alpha, \beta) \int_{a(x, nM)}^{a(x, nM) + l} p^{x + \alpha - 1} (1-p)^{nM - x + \beta - 1} dp \geq 1 - \alpha \tag{3.6}$$

를 만족시키는 최소의  $n$ 을 구한다.

실제로  $n$ 값을 구하는 방법은 다음과 같다. 먼저 HPD구간에 대응하는 상한과 하한을 구하여야 한다.  $l$ 의 값이 주어져 있으므로 하한  $(a)$ 와 상한  $(a + l)$ 을 다음 방정식

$$f(a | x, nM, \alpha, \beta) - f(a + l | x, nM, \alpha, \beta) = 0$$

을 풀므로써 구할수 있다. 여기서,  $f$ 는 (3.4)식에 나타나 있다. 이 방정식은 Newton-Rapson 반복법(iteration)으로 구할수 있다. 그리고, 적분은 Abramowitz와 Stegun(1965)의 “Hand book of Mathematical Functions”에 의하여 구한다.

#### 4. 베이즈 위험을 최소로 하는 예비표본과 추가표본의 결정

확률변수  $X_{ij}, i=1, 2, \dots, j=1, 2, \dots, M$ 이 모수  $p$ 를 갖는 베르누이 분포를 하므로 자연모수는

$$\theta = \log(p/(1-p)) \quad (4.1)$$

이다. 그리고,

$$M(\theta) = \log(1 + e^\theta) \quad (4.2)$$

라고 하자.

이제, 집락추출에서  $c_1$ 은 부차단위에 대한 조사비용이고,  $c_2$ 는 집락에 대한 포장비, 이동비용이다. 손실함수는 행동  $a = (r, n_1, n_2)$ 에 대하여

$$L(p, a) = (r - p)^2 + c_1 M(n_1 + n_2) + c_2 \sqrt{n_1 + n_2} \quad (4.3)$$

이라고 하자. 여기서,  $r$ 은  $p$ 의 추정치이고,  $n = n_1 + n_2$ 이다.

그리고,

$$Y_1 = \sum_{i=1}^{n_1} \sum_{j=1}^M X_{ij}, \quad Y_2 = \sum_{i=n_1+1}^{n_2} \sum_{j=1}^M X_{ij}$$

이고,

$$Z_1 = E_p[p(1-p) | Y_1], \quad Z_2 = E_p[p(1-p) | Y_1, Y_2] \quad (4.4)$$

이라고 놓자.

표본의 크기  $n$ 을 정한 후에 손실함수 (4.3)에 대한 베이즈 위험이 최소가 되도록 하는 예비표본의 크기(부-로트의 수)  $n_1$ 과 추가표본의 크기  $n_2$ 를 정한다. 여기서,  $n_1 + n_2 = n$ 이며,  $n_1$ 은  $n_1$ 개의 집락들의 예비표본의 크기이고,  $n_2$ 는  $n_2$ 개의 집락들인 추가 표본의 크기이다. 그러면, 베이즈 위험은

$$E_{Y_1, Y_2} E_p [E(p) | Y_1, Y_2] - p | Y_1, Y_2]^2 \\ + E_{Y_1, Y_2} E_p [c_1 M(n_1 + n_2) + c_2 \sqrt{n_1 + n_2}]$$

$$\begin{aligned}
 &= E[ \text{Var}(p \mid Y_1, Y_2) \mid Y_1, Y_2 ]^2 \\
 &\quad + E_{Y_1, Y_2} E_p [ c_1 M(n_1 + n_2) + c_2 \sqrt{n_1 + n_2} ] \\
 &= \sum_{Y_1} E_{Y_2} \{ [ \text{Var}(p) \mid Y_1, Y_2 ] \mid Y_1 \} \binom{n_1 M}{y_1} p^{y_1} (1-p)^{n_1 M - y_1} \\
 &\quad + c_1 \sum (n_1 + n_2) M \binom{n_1 M}{y_1} p^{y_1} (1-p)^{n_1 M - y_1}
 \end{aligned} \tag{4.5}$$

이다.  
 그런데,

$$\begin{aligned}
 \text{Var}(p \mid Y_1, Y_2) &= E_p (p(1-p) \mid Y_1, Y_2) / (\alpha + \beta + n_1 M + n_2 M) \\
 &= Z_1 / (\alpha + \beta + n_1 M + n_2 M)
 \end{aligned}$$

이므로 (4.6)식을 (4.5)식에 대입하여 (4.7)식이 나온다. 즉,

$$\begin{aligned}
 \sum [ \frac{Z_1}{n_1 M + n_2 M + \alpha + \beta} + c_1 (n_1 + n_2) M \\
 + c_2 \sqrt{n_1 + n_2} ] \binom{n_1 M}{y_1} p^{y_1} (1-p)^{n_1 M - y_1}
 \end{aligned} \tag{4.7}$$

이다.

그런데 집락들 사이의 운반, 포장비인  $c_2 \sqrt{n_1 + n_2}$ 을 무시할 수 있는 경우를 생각한다. 그러면  $n_1 + n_2 = n$ 인 조건 아래에서 다음 식을 최소로 하는  $n_1$ 을 구한다.  
 그런데, (4.7)식을 최소화하기 위하여 고정된  $n_1$ 과  $y_1$ 에 대하여 (4.7)식을 먼저 최소화시키는 문제를 생각한다. 그리고, (4.7)식을  $n_2$ 에 관하여 미분하여

$$n_2 = (Z_1 / c_1)^{1/2} - n_1 - (\alpha - \beta)$$

를 얻고, 이 값을 (4.7)식에 대입하여 (4.8)을 얻는다.

$$\begin{aligned}
 \sum_B \left[ \left( \frac{Z_1}{n_1 M + \alpha + \beta} \right) + n_1 M c_1 \right] \binom{n_1 M}{y_1} p^{y_1} (1-p)^{n_1 M - y_1} \\
 + \sum_{B'} [ 2(Z_1 c_1)^{1/2} - (\alpha + \beta) c_1 ] \binom{n_1 M}{y_1} p^{y_1} (1-p)^{n_1 M - y_1}
 \end{aligned} \tag{4.8}$$

여기서,  $B = \{y_1 : Z_1 \leq c_1(n_1 M + n_2 M)\}$ 이고,  $B' = \{y_1 : Z_1 > c_1(n_1 M + n_2 M)\}$ 이다.  $n_1$ 을 구하는 방법은 다음과 같다.  $n_1 = 1$ 부터  $n_1 = 10$ 까지의 베イズ 위험을 구하기 위하여  $n_1 = 1$ 일 때에 각각의  $y_1$ 에 대하여  $Z_1$ 을 구한다. 이  $Z_1$ 를 (4.8)식에 대입하여 컴퓨터 프로그램에 의하여

베이지 위험을 구한다. 그리고, 이와같이 계속하여  $n_2 = 2$  에서  $n_1 = 10$  까지의 베이지 위험을 구한다.

## 5. 모의실험

우리는 과대 산포를 무시할수 있는 부-로트의 크기  $M$ 을 먼저 정한다. Pham-Gia와 Tukkan(1992)는 실제 공업문제에 적용하여 다음과 같이 표본의 크기  $n$ 을 정하였다. 그는 생산제품의 불량률  $p$ 를 추정하는데 과거자료에 의하여 제조과정에서 불량률이 대략적으로 모수  $\alpha = 13$  과  $\beta = 57$ 를 갖는 베타 분포를 한다는 것을 알고 있다.

우리도 Pham-Gia와 Tukkan(1992)과 유사하게 모비율  $p$ 가  $\alpha = 13, \beta = 57$ 를 갖는 베타분포를 한다고 가정하면  $\rho = 1/(\alpha + \beta + 1) = 0.0145$ 이다. 과산포가 대략 5% 이상 넘지 않도록 하려면  $\{1 + (M-1)\rho\} < 1.05$ 가 되어야 하므로  $M = 5$ 이다. 이제, 임의의 자료  $x$ 에 대하여  $l = 0.17$ 보다 작고  $1 - \alpha = 0.95$ 가 되는 이-표준편차 신뢰구간을 얻기 위하여 식(3.7)를 만족시키는  $n$ 을 구하면  $n = 10$ 이다. 다음으로  $n = 10$ 일때,  $n_1 + n_2 = n$ 인 조건하에서 식(4.7)의 베이지 위험을 최소로 하는  $n_1$ 과  $n_2$ 를 구한다. 그런데,

$$Z_1 = E_p(p(1-p) | Y_1) = (y_1 M + \alpha) \frac{n_1 M - y_1 M + \beta}{(n_1 M + \alpha + \beta + 1)(n_1 M + \alpha + \beta)} \quad (5.1)$$

이다. 이  $Z_1$ 을 이용하여 식(4.7)의 베이지 위험을 최소로 하는  $n_1$ 과  $n_2$ 를 구한다. 그러나, 모수에 대한 사전정보가 없을 때에는 모비율  $p$ 는 모수  $\alpha = 1$ 과  $\beta = 1$ 를 갖는 베타분포를 한다.

그리고, 표1에서는 사전분포가 일양분포를 한다고 가정했을 때의 베이지 위험과 사전분포가 베타분포를 한다고 가정 했을 때의 베이지 위험을 비교 하였고, 또한 예비표본  $n_1$ 과 추가표본  $n_2$ 의 크기의 변화에 따른 베이지 위험을 비교하였다. 그림1에서는 예비표본  $n_1$ 의 크기에 따른 베이지 위험을 나타 내었다.

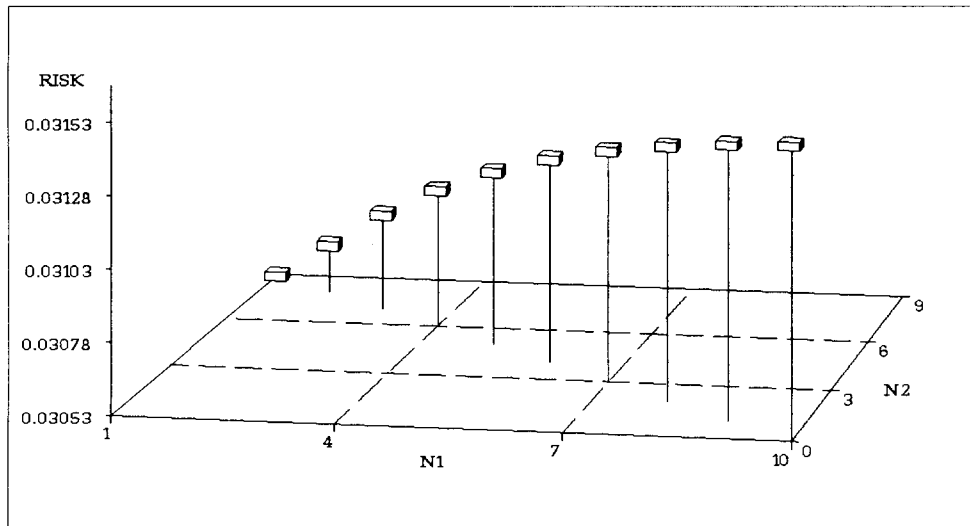
여기서  $c_1 = 0.001642$ 로 잡은 것은 (4.3)식에서  $(r-p)^2$ 의 단위 비용을 1로 보았을 때에 상대 비용이며,  $c_2$ 는 비용이 작어 무시할수 있는 경우로  $c_2 = 0$ 으로 잡은 것이다.

그리고, 모의실험 결과에 따르면  $c_1$ 값이 클수록  $n_1$ 값이 작아진다. 또한  $p$ 값이 변하면  $p$ 가 0이나 1에 가까워질수록 예비표본을 추출하는 것이 더 효율 적이다.

표1.  $n_1 + n_2 = 10$ ,  $c_1 = 0.001642$  (0.001644)이고,  $p = 0.25$  일때의 베イズ 위험

예비표본 n1	추가표본 n2	$c_1 = 0.001642$ 일때의 베イズ 위험		$c_1 = 0.001644$ 일때의 베イズ 위험	
		일양분포	베타분포	일양분포	베타분포
1	9	0.0303089	0.0180236	0.0305289	0.0180436
2	8	0.0304825	0.0181902	0.0307025	0.0182102
3	7	0.0306561	0.0183618	0.0308761	0.0183818
4	6	0.0308049	0.0185378	0.0310249	0.0185578
5	5	0.0309289	0.0187176	0.0311489	0.0187376
6	4	0.0310323	0.0189007	0.0312523	0.0189207
7	3	0.0311191	0.0190868	0.0313391	0.0191068
8	2	0.0311927	0.0192754	0.0314127	0.0192954
9	1	0.0312559	0.0194661	0.0314759	0.0194861
10	0	0.0313105	0.0196587	0.0315305	0.0196787

그림1.  $n_1 + n_2 = 10$ ,  $c_1 = 0.001664$ 이고,  $p = 0.25$  일때의 베イズ 위험



6. 결  
론

이 논문은 모집단을 집락화하여 일단 집락 추출시에 과산포 문제가 일어나므로 집락의 크기를 적정히 정하여 과산포가 일어나지 않도록 한다. 그리고, 우리가 목표로하는 오차범위 내에서 기대되는 포함확률이  $(1 - \alpha)$ 가 되도록 표본의 크기를 정한다. 더 나아가 주어진 표본의 크기 내에서 예비표본과 추가표본의 크기를 적절히 정하여 베イズ 위험을 최소로 되도록 하는 표본설계를 한다.



앞으로 연구할 과제는 과대 산포 문제를 해결하기 위해서는 과대 산포를 나타내는 모수를 포함하는 모형을 만들어 모수들의 최대우도 추정값을 구하여야 한다. 그리고, 하나의 집락은 한 생산 라인으로부터 생산된 제품으로 이루어 졌다고 가정하면 모든 라인이 동일한 조건에서 제품을 생산할수 없다. 따라서, 이러한 다른 조건들을 반영하는 공변량으로 일반화 선형모형을 만들어 자료 분석을 하여야 한다.

## 참 고 문 헌

- [1] 박성현, 박영현(1995). 통계적 품질관리, 민영사.
- [2] 신민웅, 이주영(1997). 집락표본추출에 있어서 이-단계 표본추출, 한국통계학회 논문집, 제4권 2호, 403-409.
- [3] Abramowitz M. and Stegun I. A.(1965). Hand book of Mathematical Functions, New York :Dover Publications.
- [4] Adock, C. J.(1988). A Bayesian approach to calculating sample size, *Statistician*, Vol. 37, 433-439.
- [5] Adock, C. J.(1987). A Bayesian approach to calculating sample size for multinomial sampling. *Statistician*, Vol. 36, 155-159.
- [6] Adock, C. J.(1992). Bayesian approaches to the determination of sample sizes for binomial and multinomial sampling-some comments on the paper by Pham-Gia and Tukkan, *Statistician*, Vol. 41. 399-397.
- [7] Cohen, A and Sakrowitz, H, B.(1984a). Bayes double sample estimation procedures, *The Annals of Statistics*, Vol. 12, 1035--1049.
- [8] Cohen, A and Sakrowitz, H, B.(1984b). Results in double sample estimation for the binomial distribution, *The Annals of Statistics*, Vol. 12, 1109--1116.
- [9] Diaconis, P. and Ylvisaker, D.(1979). Conjugate priors to for exponential families, *The Annals of Statistics*, Vol. 7, 269--281.
- [10] Joseph, L. Wolfson, D. B and du Berger, R (1995). Sample size calculations for binomial proportion via highest posterior density intervals. *Statistician*, Vol. 44, 143--154.
- [11] McCullagh,P. and Nelder,J.A.(1989). Generalized Linear Models. 2nd Edition, Chapman and Hall, London and New York.
- [12] Moor. D.F.(1987). Modelling the extraneous variance in the presence of extra-binomial variation. *Applied statistics* Vol. 38, 8--14.
- [13] Miller , I. and Freund, J. E.(1977). Probability and Statistical for Engineers, Second Ed. Prentice-Hall, Englewood Cliffs, New Jersey.
- [14] Pham-Gia T. and Tukkan, N.(1992). Sample size determination in Bayesian analysis. *Statistician*, Vol. 41, 389-397.