

# 적합탐색 관찰을 이용한 총화 공간표본설계에서의 추정 \*

변종석<sup>1)</sup>

## 요약

관찰 단위들간 특정한 공간 종속관계를 지닌 공간모집단에서 사각형의 칸들로 분할한 후 각 칸마다 하나의 표본점을 임의추출하여 관심 변수의 모수를 추정할 때 탐색 관찰조건을 만족하는 인접한 표본단위만을 추가 관찰하여 모수를 추정하는 적합탐색 추정 방법을 총화 공간표본설계에 적용시켜 보았다. 모의자료로 설정한 가상의 2차원 공간모집단을 총화 공간표본설계에 의해 총화시킨 후 적합 탐색 추정방법을 적용시켜 본 결과, 단순히 공간모집단을 분할하는 전통적인 공간표본설계보다 적은 수의 표본이 관찰되었으며, 효율성이 크게 감소하지 않는 결과를 얻음으로써 총화효과와 적합탐색 관찰효과가 동시에 존재하는 적절한 추정 결과를 얻을 수 있었다.

주요용어: 공간추출법, 총화 공간표본설계, 적합탐색 추정.

## 1. 서론

일반적으로 동-식물 모집단이나 대기(aerial) 모집단과 같이 자연상태를 모집단으로 하는 공간모집단에서는 공간좌표를 표본점으로 추출하는 공간표본추출법으로 관심변수의 분포 면적이나 비율, 밀도 등과 같은 모수를 추정한다.

평면에서의 표본점 추출방법에 대한 연구는 Quenouille(1949)가 자연상태로 존재하거나 평면상에 존재하는 2차원 공간에서 표본점을 추출하는 방법에 대하여 처음으로 제안하였다. Quenouille는 각 축마다 평면좌표를 계통적으로 추출하는 계통추출이 조사 전 지역의 특성을 반영하고 있기 때문에 단순임의추출보다 효율적임을 보였다. 그외에도 Bellhouse(1977, 1981), Matern(1980), Koop(1990) 등에 의해 2차원 평면에서의 공간표본설계와 공간표본추출법이 연구되었다. 이들의 연구에서 공간모집단을 분할하는 과정을 총화로, 분할된 사각형을 층으로 표현하였다. Matern은 연구지역 전체를 작은 사각형의 칸들로 분할하여 추정하는 총화추출법이 가장 효율적인 총화 방법임을 보였다. Koop은 2차원 공간좌표를 표본점으로 추출하는 경우 계통추출법, 총화추출법 등의 전통적인 방법들을 서로 조합하여 모든 가능한 공간추출법과 그에 따른 추정량을 정리하였다.

그리고 공간모집단에서 인접한 표본단위들의 강한 종속 관계와 서로 밀집되어 형성된 집락의 특성을 이용한 적합표본설계에 대한 연구는 Thompson(1990, 1992), Thompson, Ramsey와 Seber(1992)에 의해 시도되었다. 이들의 연구에서는 표본단위로부터 분할된 사각형

\* 이 논문은 1999년도 한신대학교 학술연구비 지원에 의해 연구되었음.

1) (447-791) 경기도 오산시 양산동 411번지, 한신대학교 정보통계학과, 조교수

E-mail: jsbyun@hucc.hanshin.ac.kr

의 칸들중 적합 조건을 만족하는 초기 표본과 이웃한 칸들을 표본으로 선정하여 각 초기 표본에 대한 네트워크로부터 모수를 추정하는 방법을 제안하고 있다. 또한 변과 남궁(1996)은 분할된 칸마다 공간추출법으로 추출된 표본점의 관찰값을 확률변수로 가정하여 적합탐색 조건을 만족하는 경우에만 인접한 칸들을 표본으로 추가하여 관찰하는 적합탐색 추정에 대한 연구를 시도하였다. 이들의 연구는 공간표본추출법에 관심을 두고서 단순히 공간모집단을 분할하는 전통적인 공간표본설계에 근거하여 적합탐색 관찰조건을 만족하는 칸들로 구성된 네트워크를 이용하여 모수를 추정하고 있다.

이와 같이 공간모집단의 표본설계에서는 표본추출 방법과 표본점의 관찰 방법에 관심을 두고 있다. 공간표본설계의 관점에서는 연구지역의 형태, 밀집도 등의 외적 특성을 이용하거나 표본단위사이의 공간적 상관을 이용하여 설계할 수 있을 것이다. 그러나 본 연구에서는 표본단위의 관찰 방법에 더욱 큰 관심을 두었기에 표본단위사이에 존재하는 공간 상관을 이용하여 양고 공간모집단의 외적 특성만을 반영하는 공간 표본설계 방법만을 고려하여 적합탐색 추정의 효율성을 살펴보고자 한다.

이에 본 연구는 공간모집단에서 특정한 변수의 분포 면적 추정에 관심을 가진다고 가정할 때 변과 남궁(1996)에 의해 제안된 적합탐색 추정방법을 연구지역의 형태나 규모, 밀집도 등과 같은 외적 특성을 층화변수로 이용하여 설계하는 층화 공간표본설계에 확장시켜 보았다.

## 2. 층화 공간표본설계에서의 적합탐색 추정

### 2.1. 층화 공간표본설계

길이가  $L$ 이고 폭이  $M$ 인 2차원 공간모집단(그림 2.1)에서 관심영역  $\alpha$ 의 면적을 추정하고자 할 때 사전에 미리 정한 표본크기  $n$ 개의 표본점 좌표를 공간추출법에 의해 추출한다. 표본점 좌표를 추출하는 공간추출법은  $Y$ 축에서  $b$ 개의 좌표를,  $X$ 축에서  $a$ 개의 좌표를 추출하여  $ab$ 개의 좌표 쌍으로  $n(=ab)$ 개 표본점 좌표를 결정하게 된다. 공간표본설계에서는 각 축의 좌표를 추출할 때 적용하는 추출법에 따라 많은 공간추출법이 존재하게 된다. 본 연구에서는 표본점이 전체 조사 지역에서 추출될 수 있도록 조사 지역 전체를 작은 사각형의 크기로 분할하고, 분할된 각 칸마다 독립적으로 표본점을 추출하는 층화-층화추출법( $st_0st_0$ )만을 가정하여 검토하기로 한다.

만약 그림 2.1과 같이 관심영역의 형태가 불규칙적이고 굴곡이 심한 경우에는 관심영역의 형태나 밀집도 등을 이용하는 층화 공간표본설계를 고려하는 것이 바람직할 것이다. 먼저, 공간모집단에서 적절한 층화변수를 이용하여  $H$ 개의 층으로 층화하는 층화 공간표본설계에서의 적합탐색 관찰방법에 의한 면적추정량을 살펴보기에 앞서 필요한 기호들을 다음과 같이 정의한다.

- $A^\alpha$  : 공간모집단에서 관심영역의 면적(모수)
- $A_h$  :  $h$ 층 부모집단의 조사지역 면적
- $A_{m_h}$  :  $h$ 층에서 적합탐색으로 관찰되는 조사지역의 면적
- $A_{n_h - m_h}$  :  $h$ 층에서 적합탐색으로 관찰되지 않은 조사지역의 면적

- $A_h^\alpha$  :  $h$ 층에서 관심변수의 특성을 나타내는 영역의 면적
- $A_{m_h}^\alpha$  :  $h$ 층에서 적합탐색으로 관찰된 영역의 면적
- $A_{n_h - m_h}^\alpha$  :  $h$ 층에서 적합탐색으로 관찰되지 않은 영역의 면적
- $ph_{ij}$  :  $h$ 층  $(i, j)$ 사각형에서의 비율
- $n_h$  :  $h$ 층에서의 표본점의 수(표본크기)

임의의  $h$ 층에서 층화-층화추출법을 적용하기 위해서는 먼저 X축과 평행한  $b_h$ 개 사각형, Y축과 평행한  $a_h$ 개 사각형으로 분할한다. 이 때  $h$ 층의 공간모집단은 분할된  $a_h$ 개,  $b_h$ 개의 사각형 조합으로  $n_h = a_h b_h$ 개의 분할된 칸으로 구성된다. 공간모집단을 분할할 때, 칸들의 크기를 가능한 동일하게 분할하여  $h$ 층의 공간모집단이 크기가 동일한  $n_h$ 개의 작은 사각형 칸들로 구성되게 한다. 각 칸마다 하나의 표본점을 임의추출하면  $n_h$ 개의 표본점이 얻어지게 된다. 추출된 표본점들간 서로 독립이므로  $h$ 층에서  $n_h$ 개의 표본을 임의추출하는 것과 동일한 의미를 지니게 된다.

$h$ 층의 공간모집단에서 분할된  $(i, j)$ 번째 사각형 칸의 표본점이 관심영역  $\alpha_h$ 내부에 존재하는가 또는 외부에 존재하는가를 표현해 주는 관찰값을  $Z_{hij}$ 로 나타낸다고 하면, 관찰값은 다음과 같은 베르누이 확률변수로 정의된다.

$$Z_{hij} = \begin{cases} 1, & (i, j)\text{번째 사각형에서의 표본점이 관심영역에 속하는 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases}$$

$h$ 층의 모든  $(i, j)$ 번째 사각형 칸에 대한  $Z_{hij}$ 의 기대값과 분산은 표본점 추출방법에 따라 다르게 표현되지만, 각 칸마다 임의로 표본점을 추출하는 층화-층화추출법( $st_0st_0$ )을 적용한다고 하면 확률변수  $Z_{hij}$ 에 대한 기대값과 분산은 다음과 같다.

$$E(Z_{hij}) = ph_{ij} \quad , \quad \text{Var}(Z_{hij}) = ph_{ij} (1 - ph_{ij})$$

여기서  $ph_{ij} = \text{Pr}(Z_{hij} = 1)$ 을 의미한다.

그리하여 공간모집단에서 면적을 추정할 때 공간추출법으로 결정한 표본점  $n_h$ 개에 대한 좌표들이 관심영역  $\alpha_h$  내부에 존재하는 표본점의 수를 셈(count)하여 면적을 추정하면 된다.(Bellhouse, 1981) 즉, 층화 공간표본설계로부터  $h$ 층의 면적추정량과 층화 면적추정량은 다음과 같다. 여기서  $A_h$ 는  $h$ 층의 조사지역 면적이므로 알려져 있다고 가정한다.

$$\begin{aligned} h\text{층의 면적추정량} & ; \quad \widehat{A}_h^\alpha = \frac{A_h}{n_h} \sum_{(ij)} z_{hij} \\ \text{층화 면적추정량} & ; \quad \widehat{A}^\alpha = \sum_h \widehat{A}_h^\alpha \end{aligned}$$

그리고 각 층마다 독립적으로 표본점이 추출된다고 가정하면 층화 공간표본설계에 의한 면적추정량의 분산은 다음과 같다.

$$\text{Var}(\widehat{A}^\alpha) = \sum_h \text{Var}(\widehat{A}_h^\alpha)$$

이제 층화 공간표본설계를 위한 표본크기의 배분, 표본크기의 결정에 대하여 간단히 정리해 보기로 한다.

### 2.1.1 표본 크기의 배분

공간모집단이 적절한 층화기준에 의해 H개층으로 층화되어 있고 표본크기  $n$ 이 주어졌다고 가정할 때 각 층으로 표본크기를 배분하는 방법에 대하여 살펴보기로 한다.

#### 1) 비례배분

공간모집단에 대한 층화표본설계에서 각 층의 면적  $A_h$ 는 사전에 알고 있으나 층내변동은 알 수 없다고 가정할 때, 표본크기는  $n_h = nA_h/A$ 으로 각 층의 면적에 비례하여 표본을 배분하면 된다. 비례배분에 의한 층화 공간표본설계에서의 면적추정량과 분산을 나타내면 다음과 같다.

$$\widehat{A^\alpha} = \frac{A}{n} \sum_h \sum_{(ij)} z_{hij}, \quad \text{Var}(\widehat{A^\alpha}) = \frac{A^2}{n^2} \sum_h \sum_{(ij)} \text{Var}(Z_{hij})$$

#### 2) 최적배분

최적배분법으로 표본을 배분하기 위하여 층화 공간표본설계에서 고려되는 비용함수를  $C = c_0 + \sum_h n_h c_h$ 로 가정한다. 여기서  $c_0$ 는 고정비용이고,  $c_h$ 는  $h$ 층에서의 단위당 평균조사비용이다. 비용함수에서  $c_0$ 는 고정비용으로써 일정한 값을 가지게 되므로 식 (2.1)과 같이 표현할 수 있다.

$$C = C - c_0 = \sum_h n_h c_h \quad (2.1)$$

그리고 임의의  $h$ 층에서 표본점들을 독립적으로 추출한다고 가정하면 층화 공간표본설계에 의한 추정량의 분산은 식 (2.2)와 같이 나타낼 수 있다.

$$\begin{aligned} \text{Var}(\widehat{A^\alpha}) &= \sum_h \frac{A_h^2}{n_h^2} \sum_{(ij)} \text{Var}(Z_{hij}) \\ &= \sum_h \frac{A_h^2}{n_h} p_h (1 - p_h) \end{aligned} \quad (2.2)$$

따라서 고정된 비용함수(식 (2.1))를 조건으로 하여 추정량의 분산(식 (2.2))을 최소화하는 식 (2.3)과 같은 Lagrange 승수  $\lambda$ 를 이용하면 최적배분에 의한 표본 크기  $n_h$ 를 유도할 수 있다.

$$\Phi = \sum_h \frac{A_h^2}{n_h} p_h (1 - p_h) + \lambda \left( \sum_h c_h n_h - C \right) \quad (2.3)$$

식 (2.3)으로부터 유도된 최적배분에  $h$ 의해 배분된 층의 표본크기와 추정량의 분산은 다음과 같다.

$$n_h = n \frac{A_h \sqrt{p_h(1-p_h)/c_h}}{\sum_h A_h \sqrt{p_h(1-p_h)/c_h}}$$

$$\text{Var}(\widehat{A}^\alpha) = \frac{1}{n_h} \left[ \sum_h A_h \sqrt{p_h(1-p_h)c_h} \right] \left[ \sum_h A_h \sqrt{p_h(1-p_h)/c_h} \right]$$

3) 네이만배분

네이만(Neyman)배분은 최적배분의 특별한 경우로써 각 층의 단위당 평균조사비용  $c_h$ 가 모든 층에서 동일하다는 가정하에서  $n_h$ 의 최적값을 계산하는 방법이다. 네이만배분에 의한 각 층으로 배분되는  $h$ 층의 표본크기와 추정량의 분산은 다음과 같다.

$$n_h = n \frac{A_h \sqrt{p_h(1-p_h)}}{\sum_h A_h \sqrt{p_h(1-p_h)}}$$

$$\text{Var}(\widehat{A}^\alpha) = \frac{1}{n} \left[ \sum_h A_h \sqrt{p_h(1-p_h)} \right]^2$$

2.1.2 표본 크기의 결정

층화 공간표본설계에서의 표본크기는 전통적인 방법에서와 마찬가지로 표본크기의 배분방법에 의해 결정된다. 첫째, 총조사비용을 고정시켰을 때 둘째, 추정량의 분산을 고정시켰을 때 표본크기  $n$ 을 결정할 수 있다.

먼저 고정시킨 총조사비용을  $C'$ 이라고 가정하면 식 (2.1)과 같은 비용함수는  $C' - c_0 = \sum n_h c_h$ 으로 나타낼 수 있다.

또한 각 층마다 독립적으로 표본점을 추출한다고 가정하여 고정시킨 추정량의 분산은 다음과 같이  $\text{Var}'$ 으로 표현하자.

$$\text{Var}'(\widehat{A}^\alpha) = \sum_h \frac{A_h^2}{n_h^2} \sum_i \sum_j \text{Var}(Z_{hij}) = \sum_h \frac{A_h^2}{n_h} p_h(1-p_h)$$

층화 공간표본설계에서도 필요한 표본크기는 적용되는 표본배분법에 따라 유도하면 적절한 표본크기를 결정할 수 있다.

1) 비례배분

총조사비용이 고정된 경우 : 여기서  $W_h = A_h/A$ 일 때

$$n = (C' - c_0) \frac{A}{\sum_h A_h c_h} = \frac{C' - c_0}{\sum_h W_h c_h}$$

추정량의 분산이 고정된 경우 :

$$n = \frac{A}{\text{Var}'(\widehat{A^\alpha})} \sum_h A_h p_h (1 - p_h)$$

## 2) 최적배분

총조사비용이 고정된 경우 :

$$n = (C' - c_0) \frac{\sum A_h \sqrt{p_h(1-p_h)}/c_h}{\sum A_h \sqrt{p_h(1-p_h)}c_h}$$

추정량의 분산이 고정된 경우 :

$$n = \frac{1}{\text{Var}'(\widehat{A^\alpha})} \left[ \sum_h A_h \sqrt{p_h(1-p_h)}c_h \right] \left[ \sum_h A_h \sqrt{p_h(1-p_h)}/c_h \right]$$

## 3) 네이만배분

총조사비용이 고정된 경우 :

$$n = (C' - c_0) \frac{\sum A_h \sqrt{p_h(1-p_h)}}{\sum A_h \cdot c_h \sqrt{p_h(1-p_h)}}$$

추정량의 분산이 고정된 경우 :

$$n = \frac{1}{\text{Var}'(\widehat{A^\alpha})} \left[ \sum_h A_h \sqrt{p_h(1-p_h)} \right]^2$$

## 2.2. 층화 공간표본설계에서의 적합탐색 관찰

변과 남궁(1996)에 의해 제안된 적합탐색 추정법은 단순히 공간모집단을 분할하는 전통적인 공간표본설계에 근거하여 적합탐색 조건을 만족하는 칸들과 인접한 칸만을 추가 관찰하여 관심영역의 면적을 추정하는 방법이다. 그러나 그림 2.1과 같이 하나의 큰 집락으로 분포되어 있고 추정하려는 관심영역의 형태가 불규칙적이고 굴곡이 심한 공간모집단에서 면적을 추정할 때, 연구지역의 형태, 밀집도 등의 외적 특성을 이용하거나 표본단위사이의 공간적 상관관계를 이용하는 층화 공간표본설계를 고려하는 것이 일반적이다. 따라서 본 연구는 표본단위의 관찰 방법에 더욱 큰 관심을 두고 있으므로 표본단위사이에 존재하는 공간상관관계를 이용하여 얇고 공간모집단의 외적 특성만을 반영하는 공간표본설계 방법만을 고려하여 변과 남궁에 의해 제안된 적합탐색 추정방법을 확장시켜 보고자 한다.



층화 공간표본설계로 층화한 임의의  $h$ 층 공간모집단에서 선택된 초기 칸들에 대해 공간추출법으로 표본점의 좌표를 추출하여 표본점의 관찰값  $Z_{hij}$ 들을 먼저 관찰한다. 이 때 초기 칸의 관찰값  $Z_{hij}$ 는 관심영역내의 존재여부를 판단하는 베르누이 확률변수를 따르므로  $h$ 층에서의 적합탐색 조건은 초기 칸들의 관찰값들을 이용하여 탐색조건을 결정하면 된다. 이 때 초기 칸들의 관찰값이 0 또는 1로 관찰되므로 층에서의 적합탐색 조건은 임의로  $z_{hij} = 0$  또는 1로 설정한다. 그리고 탐색 조건을 만족하는 칸들의 상/하, 좌/우의 인접한 칸들을 추가 관찰해 나감으로써 초기 출발 칸에 대하여 관찰되는 칸들로 탐색관찰 네트워크를 구성하게 된다. 이후에도 탐색 조건을 만족하여 탐색된 칸들에 대해서는 인접한 칸들을 계속적으로 탐색하여 하나의 네트워크를 구성하게 된다. 이 때, 어떤 네트워크에서 한번 관찰된 칸은 다시 관찰하지 않는다. 그리고 탐색조건을 만족하지 않아서 더 이상의 인접 칸들을 지속적으로 탐색할 수 없는 경우를 가장자리(edge) 칸이라고 하며, 자신만 탐색 네트워크에 포함되게 된다. 즉, 하나의 네트워크에는 탐색조건을 만족하여 관찰된 칸들과 더 이상 추가 관찰이 진행 되지 않는 가장자리 칸들로 구성하게 된다.

이와 같이 적합탐색 관찰에서는 적합탐색 조건에 의해 구성된 네트워크의 정보를 조건부로 하면 경계내의 관찰되지 않은 칸의 결과를 알 수 있게 된다. 따라서 임의의  $h$ 층에 대한 적합탐색 추정에서는 적합탐색 관찰로 구성된 네트워크로부터 추정된 결과와 관찰되지 않고서 계산되는 결과를 합하여  $h$ 층의 면적을 추정하게 된다.

### 2.3. 층화 공간표본설계에서 적합탐색 추정

층화 공간표본설계에 적합탐색 관찰을 적용하면, 어떤  $h$ 층의 관심영역 면적( $A_h^\alpha$ )은 적합탐색에 의해 관찰된  $m_h$ 개의 칸들로 구성된 네트워크의 면적( $A_{m_h}^\alpha$ )과 이 네트워크를 조건부로 하여 관찰되지 않은  $n_h - m_h$ 개의 칸들에 대한 면적( $A_{n_h - m_h}^\alpha$ )의 합으로 계산될 수 있기 때문에 식 (2.4)와 같이 표현할 수 있다.

$$A_h^\alpha = A_{m_h}^\alpha + A_{n_h - m_h}^\alpha \tag{2.4}$$

이와 같이 층화 공간표본설계에 의해 층화된 층화 공간모집단에서 적합탐색 관찰에 의한  $h$ 층의 면적추정량은 적합탐색의 관찰 조건으로 관찰된 칸들의 관찰값  $z_{hij}$ 들로 구성된 적합탐색 네트워크의 면적과 관찰되지 않은 칸들의 면적을 결합함으로써 식 (2.5)와 같이 추정할 수 있다.

$$\begin{aligned} \widehat{A}_h^\alpha &= A_{m_h} \cdot \frac{1}{m_h} \sum_{(ij) \in \psi_{h(m)}} z_{hij} + A_{n_h - m_h} \cdot \frac{1}{n_h - m_h} \sum_{(ij) \in \psi_{h(n-m)}} z_{hij} \\ &= \widehat{A}_{m_h}^\alpha + \widehat{A}_{n_h - m_h}^\alpha \end{aligned} \tag{2.5}$$

여기서  $\psi_{h(m)}$ 은  $h$ 층에서 적합탐색 관찰로 관찰된 칸들의 네트워크를 의미한다.

정리 2.1  $h$ 층에서의 적합탐색에 의한 면적 추정량  $\widehat{A}_h^\alpha$ 은  $A_h^\alpha$ 의 불편추정량이다.

식 (2.5)의 추정량  $\widehat{A}_{n_h - m_h}^\alpha$ 는 관찰되지 않은  $n_h - m_h$ 개 칸들로부터 추정하게 된다. 그러나 적합탐색으로 관찰되지 않는  $n_h - m_h$ 개 칸들은 특정 칸들이 탐색조건을 만족하지 않아 더 이상의 인접 칸들을 탐색하지 않게 되므로 탐색되지 않는 칸들의 특성은 탐색 관찰 조건과 반대 속성을 지니게 된다. 따라서 2.2절의 적합탐색 관찰 조건과 가정에 의해 적합탐색으로 관찰되지 않는 칸들은  $\sum_{(ij) \in \psi_{h(m)}} z_{hij} = n_h - m_h$ 이 되므로  $\widehat{A}_{n_h - m_h}^\alpha = A_{n_h - m_h}$ 의 관계가 성립하게 된다. 여기서  $h$ 층의 모집단 크기( $A_h$ )와 적합탐색 관찰에 의한 영역의 크기( $A_{m_h}$ )는 알 수 있으므로  $A_h = A_{m_h} + A_{n_h - m_h}$ 의 관계로부터  $A_{n_h - m_h}$ 의 계산이 가능하다. 그러므로  $\widehat{A}_{n_h - m_h}^\alpha$ 는 실제로 관찰하지 않고서 적합탐색 조건과  $A_{n_h - m_h}$ 에 의해 자연스럽게 알 수 있게 된다.

또한  $h$ 층에서 크기가 동일한  $n_h$ 개의 (정)사각형 칸으로 분할된 각 칸의 면적이  $A_{hc}(= A_h/n_h)$ 로 동일하다면,  $h$ 층에서의 적합탐색 관찰에 의한 면적 추정량은 식 (2.6)과 같이 표현할 수 있다.

$$\begin{aligned} \widehat{A}_h^\alpha &= A_{hc} \sum_{(ij) \in \psi_{h(m)}} z_{hij} + A_{hc} \sum_{(ij) \in \psi_{h(n-m)}} z_{hij} \\ &= \frac{A_h}{n_h} \left\{ \sum_{(ij) \in \psi_{h(m)}} z_{hij} + \sum_{(ij) \in \psi_{h(n-m)}} z_{hij} \right\} \end{aligned} \tag{2.6}$$

$h$ 층에서 적합탐색 관찰로 추정된 면적 추정량의 분산은 적합탐색 관찰에 의한 부분과 관찰되지 않고서 알 수 있는 부분으로 분리하여 유도하면 된다. 그러나 앞서 표현한 바와 같이 추정량  $\widehat{A}_{n_h - m_h}^\alpha$ 은 적합탐색 조건에 의해 자연스럽게 결정되므로 상수이다. 따라서  $h$ 층에서  $(i, j)_h$ 번째 칸의 관찰값인  $Z_{hij}$ 는 2.1절에서와 같이 베르누이 확률변수이므로  $Z_{hij}$ 의 분산을 이용하면  $h$ 층에서의 적합탐색 관찰로 추정된 면적 추정량의 분산은 다음과 같이 정리할 수 있다.

**정리 2.2**  $h$ 층에서 적합탐색 관찰에 의한 면적 추정량  $\widehat{A}_h^\alpha$ 의 분산은 다음과 같다.

$$\begin{aligned} \text{Var}(\widehat{A}_h^\alpha) &= A_{m_h}^2 \cdot \frac{1}{m_h^2} \left[ \sum_{(ij) \in \psi_{h(m)}} \text{Var}(Z_{hij}) \right] \\ &= A_{m_h}^2 \cdot \frac{1}{m_h^2} \left[ \sum_{(ij) \in \psi_{h(m)}} p_{hij}(1 - p_{hij}) \right] \end{aligned} \tag{2.7}$$

그러나 분산추정량은 적합탐색 관찰로 관찰되는 표본크기를 알 수 없고, 각 칸마다 하나의 표본점만을 추출하기 때문에 직접 구할 수 없다. 적합탐색 추정에 대한 분산추정량에 대한 결과는 변과 남궁(1996)의 결과를 참고하면 된다.

이제 층화 공간표본설계에서 적합탐색 관찰로 추정된 면적 추정량을 살펴보기로 한다. 식 (2.4)과 (2.5)를 이용하여 각 층에서의 면적추정량을 합하면 층화 공간표본설계에서 적합탐색 관찰로 추정된 면적 추정량을 정리 2.3과 같이 얻을 수 있다.

**정리 2.3** 층화 공간표본설계에서 적합탐색 관찰에 의한 면적 추정량  $A^\alpha$ 은 식(2.8)과 같이 추정된다.

$$\begin{aligned}\widehat{A}^\alpha &= \sum_h \widehat{A}_h^\alpha \\ &= \sum_h \widehat{A}^\alpha_{m_h} + \sum_h \widehat{A}^\alpha_{n_h - m_h}\end{aligned}\quad (2.8)$$

식 (2.8)의 추정량에 대한 평균과 분산은 정리 2.1과 정리 2.2로부터 다음과 같이 정리할 수 있으며, 이 결과로부터 층화 공간표본설계에서의 적합탐색 관찰에 의한 면적 추정량이  $A^\alpha$ 의 불편추정량이 됨을 알 수 있다.

$$E(\widehat{A}^\alpha) = A^\alpha \quad , \quad \text{Var}(\widehat{A}^\alpha) = \sum_h \text{Var}(\widehat{A}^\alpha_{m_h})$$

#### 2.4. 층화 공간표본설계에서 적합탐색 관찰에 의한 표본크기의 감소 효과

층화 공간표본설계에서 적합탐색 관찰로 관찰된 표본 수의 감소효과를 살펴보면, 2.1절의 표본크기 배분 방법에 따라 배정된  $h$ 층의  $n_h$ 개 표본 중 적합탐색 조건에 의해  $m_h$ 개만이 관찰된다. 따라서 각 층마다  $n_h - m_h$ 개의 표본 수만큼 관찰 표본수가 감소하게 되므로 층화 공간표본설계에서 적합탐색 관찰로 관찰되는 표본 수는  $m = \sum_h m_h (\leq n (= \sum_h n_h))$ 만큼 관찰하게 될 것이다. 따라서 층화 공간표본설계에서 적합탐색 관찰에 의해 관찰되는 표본의 수는 사전에 계획된  $n$ 개 중  $m$ 개만이 관찰되므로  $n - m$ 개는 관찰하지 않게 된다. 적합탐색 관찰로 관찰되는 표본 수와 사전에 계획된 표본 수와의 관계는 항상  $m \leq n$ 의 관계를 나타내게 되므로 적합탐색 관찰을 이용하면 항상 표본 수가 감소된다는 사실을 예상할 수 있다.

특히 관심영역의 분포 형태나 면적을 고려하여 층화하는 층화 공간표본설계에서 사전에 계획된 표본 크기  $n$ 을 각 층으로 배분할 때, 분포 형태가 불규칙적이고 굴곡이 심한 층에 많은 수의 표본을 배정하도록 한다. 이는 분할된 칸의 크기를 충분히 작게 하여 적합탐색 관찰로 표본점을 관찰할 수 있기 때문에 실제 관찰되는 표본 수에 대한 감소 효과는 더욱 크게 나타날 것이다.

### 3. 모의자료를 이용한 실증분석

#### 3.1. 공간 모집단의 설정

적합탐색 관찰에 의한 층화 공간표본설계에서의 면적 추정량을 살펴보기 위하여 크기가  $40 \times 40$ 인 가상의 2차원 공간모집단(부록)을 모의자료로 설정하였다. 그리고 추정하고

표 3.1: 가상의 공간모집단에 대해 비례배분된 층화 공간표본설계의 내용

표본크기		모집단 크기	관심영역의 면적	배분된 표본크기	표본점의 조합	분할된 칸의 크기
64	층 1	25 × 40 (1000)	516	40	5 × 8	5 × 5
	층 2	15 × 40 ( 600)	44	24	3 × 8	
100	층 1	24 × 40 ( 960)	500	60	6 × 10	4 × 4
	층 2	16 × 40 ( 600)	60	40	4 × 10	

자 하는 관심 영역의 분포 형태가 불규칙적이고 치우친 정도와 굴곡이 심한 공간모집단으로 표현하였다. 부록의 공간모집단에서 “1”은 관심영역의 분포내에 존재하는 것을 의미하며 “0”은 분포밖에 존재하는 것을 나타낸다. 가상의 공간모집단은 관심영역의 분포 면적( $A^a$ )이 560이고, 분포 비율( $p$ )은 0.35로 구성되어 있다.

### 3.2. 적합탐색 관찰을 위한 층화 공간표본설계

가상의 공간모집단에서 관심영역의 분포 면적을 추정하기 위한 표본크기로는  $n = 64, 100$ 의 두 가지 경우만을 고려하였다. 공간모집단에서 표본점을 결정하는 공간 표본추출법으로는 임의-임의추출법( $r_0r_0$ )과 층화-층화추출법( $st_0st_0$ )만을 적용하여 비교해 보았다. 임의-임의추출법은 2차원 평면의 각 축에서 표본크기만큼 임의로 좌표를 선택하여 표본점을 결정하는 방법이며, 층화-층화추출법은 표본크기만큼 공간모집단을 사각형의 칸으로 분할하여 각 칸마다 임의-임의추출법으로 하나의 표본점 좌표를 결정하는 방법이다. 이 때 공간모집단은 분할이 간단하고 효율적인 결과를 제공해 주는 정사각형으로 분할하였다. 따라서 가상의 공간모집단에서 분할된 칸의 형태 및 표본 수는 정사각형 격자로 분할된  $8 \times 8, 10 \times 10$  으로 정하였다.

또한 층화 공간표본설계를 적용하기 위하여 관심영역의 분포 형태를 고려하여 2개의 층으로만 층화하였으며, 관찰 표본크기는 층의 모집단 크기에 비례하는 비례배분에 의해 표본 크기를 배분하였다. 그러나 층화된 공간모집단에서도 정사각형의 칸으로 분할할 수 있도록 하기 위하여 계획된 표본 크기에 따라 층화 공간모집단내에서 분할 칸의 면적을 다르게 조정하여 이용하였다. 가상의 공간모집단에서 층화 공간표본설계에 의한 층화된 모집단 현황 및 공간 표본설계 내용을 표 3.1에 나타내었다.

### 3.3. 적합탐색 관찰 조건

기본적인 층화 공간표본설계에서는 표본설계단계에서 각 층으로 배분된 표본 크기만큼 공간모집단을 정사각형으로 분할하고, 분할된 모든 칸의 관찰값을 이용하여 면적을 추정하게 된다. 층화 공간표본설계에서 적합탐색 추정을 위한 표본설계는 층화 공간표본설계와 동일하지만 적합탐색 조건에 의해 관찰되는 칸들의 결과를 조건부로 하여 면적을 추정하게 된다. 이 때 적합탐색 관찰조건이 추정 결과에 영향을 준다는 사실을 직관적으로 알 수 있다.

적합탐색 관찰조건은 처음에 임의추출하여 선택되는 초기 칸들의 관찰값에 의해 결정된다. 이와 같이 적합탐색 관찰에 의한 추정에서는 초기 칸의 선택이 적합탐색 관찰 조건을 결정하는 중요한 의미를 지니고 있다. 그러나 관심 영역이 하나의 큰 집락으로 분포되어 있거나 고립된 지역이 존재하지 않는다고 가정할 때 관심영역 내부 혹은 외부에 탐색하는 가하는 탐색 조건의 결정에 초기 칸의 선택이 큰 의미를 지니지 않게 된다.

따라서 본 연구에서는 각 층마다 관심 영역 외부에 존재하는 임의의 칸을 선택하여 적합탐색 관찰 조건을  $z_{hij} = 0$ 으로 설정하고, 관심영역의 외부에 존재하는 칸들을 적합탐색 조건에 의해 관찰하기로 하였다. 그리하여 초기 칸을 임의추출하기보다는 각 층마다 관심영역의 외부에 존재하는 4개의 모서리칸을 초기 칸으로 정하였다. 이와 같이 층화 공간표본설계에서의 적합탐색 관찰은 각 층내에서 초기 칸마다 독립적으로 임의-임의추출법을 적용하여 추출된 표본점을 관찰하게 되며, 적합탐색 조건을 만족하여 관찰되는 칸들을 각 초기 칸들에 대한 하나의 네트워크를 구성해 가면서 탐색 조건을 만족하지 않는 가장자리 칸이 나타날 때까지 반복하여 관찰하게 된다.

가상의 공간모집단과 같이 조사하려는 관심영역이 하나의 큰 집락으로 형성되어 있는 경우  $z_{hij} = 1$ 로 설정하여 탐색한다면 하나의 탐색 네트워크로 모두 관찰이 가능해지게 된다. 그러므로 보다 많은 네트워크를 구성하여 표본설계와 추정량의 효율성을 높일 수 있도록 가상의 공간 모집단에 대한 모의실험에서는  $z_{hij} = 0$ 으로 설정하게 된 것이다.

### 3.4. 면적 추정 결과

가상의 공간모집단에서 관심영역의 면적을 추정할 때 적합탐색 관찰을 이용한 층화 공간표본설계에서의 추정 방법을 적용시켜 보았다.

가상의 공간모집단에서 층화 표본설계효과와 층화 설계에서 적합탐색 관찰효과를 살펴 보기 위하여 공간추출법중 임의-임의추출법과 층화-층화추출법을 이용하였다. 면적 추정을 위한 표본설계로는 1) 기본적인 공간표본설계(비층화), 2) 층화 공간표본설계, 3) 적합탐색을 이용한 층화 공간표본설계를 적용하였다. 그리고 층화효과와 적합탐색 관찰에 의한 추정 효과를 비교하기 위하여 적용되는 공간추출법과 표본크기를 고려하여 1000번의 반복 실험을 실시하였으며, 면적추정량은 1000번 반복 결과에 대한 평균으로 계산하였다. 적합탐색 관찰에 의한 추정에서는 실제로 관찰된 표본 수를 제시하여 감소된 표본 크기를 비교할 수 있도록 하였다. 그러나 기존 공간표본설계에서의 표본 수와 적합탐색 관찰에서의 표본 수가 서로 다르기 때문에 최소제곱오차(MSE)의 직접적인 비교가 곤란하지만, 각 방법

표 3.2: 기본적인 공간표본설계와 층화 공간표본설계와의 비교

공간추출법	비층화 공간표본설계			층화 공간표본설계				상대효율
	표본 크기	면적 추정량	MSE(1)	표본크기		면적 추정량	MSE(2)	
				층 1	층 2			
임의-임의추출법 ( $r_0r_0$ )	64	587.38	9474.38	40	24	583.68	7925.88	0.8366
	100	590.46	6213.12	60	40	582.03	5017.34	0.8075
층화-층화추출법 ( $st_0t_0$ )	8 × 8	565.48	2579.88	5 × 8	3 × 8	566.98	2458.63	0.9530
	10 × 10	552.22	1273.34	6 × 10	4 × 10	550.98	1271.30	0.9984

여기서 상대효율 = MSE(2)/MSE(1)

표 3.3: 적합탐색 관찰을 이용한 층화 공간표본설계

공간추출법	표본크기		적합탐색 관찰 표본크기			면적 추정량	MSE(3)	상대 효율 (1)	상대 효율 (2)
	층 1	층 2	최소	최대	평균				
층화-층화추출법 ( $st_0st_0$ )	5 × 8	3 × 8	48	63	54.30	577.88	2655.63	1.0294	1.0801
	6 × 10	4 × 10	75	92	82.18	564.56	1382.66	1.0858	1.0876

여기서 상대효율(1)=MSE(3)/MSE(1), 상대효율(2)=MSE(3)/MSE(2)

에 따른 추정 결과의 효율성을 살펴볼 수 있도록 참고로 MSE를 제시하였다. 가상의 모의 자료에 대한 면적 추정 결과는 적용시킨 표본설계에 따라 표 3.2와 표 3.3에 나타내었다.

비층화 공간표본설계와 층화 공간표본설계를 적용하여 추정된 표 3.2의 결과를 요약하면 i) 공간추출법중 층화-층화추출법이 임의-임의추출법보다 더 효율적이고, ii) 표본점의 수가 증가하거나 표본 칸의 면적이 작을수록 효율성이 증가한다는 사실을 알 수 있으며, iii) 비층화 공간표본설계와 층화 공간표본설계와의 효율성을 비교하여 보면 임의-임의추출법의 경우에는 효율성이 매우 높게 나타났으며, 층화-층화추출법에서는 효율성이 크게 증가하지는 않았지만 대체로 층화효과가 존재하고 있음을 알 수 있다. 이러한 결과는 본 연구에서 편의상 조사지역을 두 개의 층으로 단순하게 분리하여 검토하였기 때문에 나타난 결과라고 판단된다. 실제로 층화 표본설계에서는 층화 변수나 방법에 따라 추정에 많은 영향을 주게 되므로 공간모집단에서도 분포의 밀도나 형태와 같은 적절한 층화 변수를 이용하여 층화한다면 층화의 효율성이 더욱 높아지게 될 것이다.

적합탐색 관찰을 이용한 층화 공간표본설계로 추정된 표 3.3의 결과를 요약하면 i) 표 3.2에서와 마찬가지로 표본점의 수가 증가하거나 표본 칸의 면적이 작을수록 효율성이 증가한다는 사실과 ii) 층화 공간표본 설계에서 적합탐색 관찰을 이용하여 적은 수의 표본으로 면적을 추정하는 경우 상대효율(1)과 상대효율(2)에서 볼 수 있듯이 표 3.2의 결과에 비해 상대적인 효율성이 다소 떨어지지만 그 차이가 크지 않음을 알 수 있다. 이러한 결과는

두 관찰 방법의 표본 크기가 서로 다르기 때문에 나타나는 현상으로써 적은 수의 표본을 관찰하게 되는 적합탐색 관찰에 의한 추정량과의 MSE를 비교하게 되면 상대 효율이 떨어지는 것은 당연한 결과일 것이다. 따라서 두 방법에 대한 상대효율을 비교한다는 것은 곤란하다고 판단된다.

그러나 이 결과를 관찰 표본 크기의 관점에서 살펴보면, 표 3.2의 공간표본설계에서는 주어진 모든 칸들을 관찰하지만 적합탐색 관찰을 이용한 층화 공간표본설계에서는 표본크기가 64인 경우 평균 54.30개의 표본이 관찰되어 실제 관찰 표본 수를 약 15.2%(9.70개)정도 감소시켰으며, 표본크기가 100인 경우 평균 82.18개의 표본을 관찰함으로써 약 17.8%(17.72개) 정도의 실제 관찰 표본 수를 감소시키는 것으로 나타났다. 또한 조사지역에 대한 관심영역의 분포 비율이 크거나 분포 형태의 굴곡이 심하지 않게 불룩한 형태의 공간모집단에서 분할된 칸의 면적이 작거나 칸의 수가 커질수록 적합탐색에 의해 관찰되는 표본 수의 감소효과는 더욱 높아지는 것을 알 수 있다.

#### 4. 결론

본 연구에서는 2차원의 공간모집단에서 표본 수만큼 칸을 분할하고 각 칸마다 독립적으로 공간 좌표를 표본점으로 임의추출하여 면적을 추정하는 경우, 조사과정에서 관찰조건을 만족하는 칸과 인접한 칸들을 추가 관찰하여 탐색해 가는 적합탐색 추정을 층화 공간표본설계에 적용시켜 보았다.

이는 공간모집단에서의 추정에서 모집단 분포의 형태나 밀집도 등을 고려하는 층화 공간표본설계가 일반적으로 널리 적용될 수 있는 표본설계방법이기 때문이다.

구체적으로 공간 모집단에서 면적을 추정할 때 적용할 수 있는 층화 공간표본설계 과정에서 표본크기를 계산하고 배분하는 방법에 대하여 정리한 후, 층화 공간표본설계에서 적합탐색 관찰에 의한 추정을 살펴보았다. 가상의 모의자료에 대한 분석에서 보았듯이 1) 전통적인 공간 표본설계보다는 층화 공간표본설계에서 보다 효율적인 추정 결과를 보였으며, 2) 본 연구에서 제안된 방법을 적용하여 적은 수의 표본을 관찰하여 추정하더라도 효율성이 크게 떨어지지 않는 결과가 얻어짐을 알 수 있었다.

따라서 본 연구에서 제안한 바와 같이 공간모집단에서도 관심 영역의 면적이나 형태에 따라 적절한 방법으로 층화하여 층마다 분할 칸의 형태를 상이하게 조정하여 설계할 수 있는 층화 공간표본설계에서 적합탐색 관찰을 이용하여 분포의 면적이나 비율, 밀집도 등을 추정한다면 층화효과와 적합탐색 관찰에 의한 관찰 표본 수의 감소효과는 더욱 높아질 것으로 예상된다.

그리고 본 연구에서 고려한 층화 공간표본설계에서는 표본점들간 존재하는 공간 상관관계를 고려하지 못하였다. 그러므로 향후 공간모집단에서 표본점들간 존재하는 공간 상관관계에 대한 규명과 더불어 상관관계를 반영하는 표본설계 방법에 대한 연구가 추가적으로 요구된다. 또한 환경분야나 생물분야 등과 같이 자연상태를 모집단으로 설정하는 분야에서 적용할 수 있는 다양한 연구들이 지속적으로 시도되기를 기대해 본다.

## 감사의 글

본 논문에 대하여 많은 조언을 해 주신 심사위원님들께 감사를 드립니다.

## 참고문헌

- [1] 변중석, 남궁평 (1996). 적합탐색 관찰방법을 이용한 추정, <응용통계연구>, 제9권 2호, 145-159.
- [2] Bellhouse, D.R. (1977). Some optimal designs for sampling in two dimensions, *Biometrika*, 64, 605-612.
- [3] Bellhouse, D.R. (1981). Area estimation by point-counting techniques, *Biometrics*, 37, 303-312.
- [4] Koop, J.C. (1990). Systematic sampling of two-dimensional surfaces and related problems, *Commun. Statist. - Theory and Methods*, 9(5), 1701-1750.
- [5] Matern, B. (1980). *Spatial Variation* (2 eds), Lecture Note in Statistics 36, Springer-Verlag.
- [6] Overton, W.S. and Stehman, S.V. (1993). Properties of Design for sampling continuous spatial resources from a triangular grid, *Commun. Statist. - Theory and Methods*, 22(9), 2641-2660.
- [7] Quenouille, M.H. (1949). Problems in plane sampling, *Annals of Mathematical Statistics*, vol 20, 355-375.
- [8] Thompson, S.K. (1990). Adaptive cluster sampling, *J. Amer. Statist. Assoc.*, 85, 1050-1059.
- [9] Thompson, S.K. (1992). *Sampling*, New York : Wiley.
- [10] Thompson, S.K., Ramsey, F.L. and Seber, G.A.F. (1992). An adaptive procedure for sampling animal populations, *Biometrics*, 48, 1195-1199.

[ 1999년 8월 접수, 2000년 3월 채택 ]



## Adaptive Searching Estimation in Stratified Spatial Sample design \*

Jongseok Byun<sup>1)</sup>

### ABSTRACT

We systematized an stratified spatial sample design(SSSD) that uses the adequate stratification criteria such as the shapeness or the dispersion of an interesting region in a spatial population. And we proposed an adaptive searching estimation method in the SSSD to estimate the area of region of interest in two-dimensional surfaces. When we adopt the proposed adaptive searching estimation method in SSSD, the observing sample size is more decreased than a classical sample design that all the designed sample size is observed. Nevertheless it has been shown that we can produce the moderate result but the efficiency is a slight reduced.

*Keywords:* Spatial Sampling; Stratified Spatial Sample Design; Adaptive Searching Estimation.

---

\* This paper was supported by Hanshin University Research Fund, 1999.

1) Assistant Professor, Department of Statistics and Information, Hanshin University.  
E-mail: jsbyun@hucc.hanshin.ac.kr