응용통계연구 제 9권 2호
1996년 9월 pp. 145-159

적합탐색 관찰방법을 이용한 추정

변 종석[1)], 남궁 평[2)]

요 약

2차원의 공간모집단에서 모수를 추정하기 위하여 모집단내에서 존재하는 위치들간의 공간적 종속관계를 이용하여 표본단위를 관찰하는 적합탐색 관찰방법을 제안하고, 공간모집단에서 관심영역의 면적이나 비율을 추정하고자 할 때 적합탐색 관찰방법을 적용하여 얻은 추정량에 대하여 살펴본다. 각 표본점들 사이에 존재하는 공간종속관계를 이용한 적합탐색 관찰방법을 적용한다면 조사이전에 미리 정해지는 표본크기 n에 대하여 추출된 표본단위들을 모두 관찰하는 전통적인 표본설계에 비하여 보다 적은 수 $m(m \leq n)$개의 관찰로써 추정할 수 있을 것이다. 이로써 표본조사시 발생하는 비용과 시간 등의 절감효과가 기대되며, 추정하려는 관심영역의 형상에 대한 사전 정보가 존재하지 않더라도 적합탐색 관찰방법으로 그 형상을 파악할 수 있게 된다.

1. 서 론

2차원 공간에 존재하는 몇몇 공간모집단은 임의의 한 장소와 일정한 거리 내에 존재하는 장소들 간에 특정한 관계를 가지고 있다. 그 특성은 첫째, 관찰된 한 장소에서 그 거리가 가까울수록 공간적 종속관개가 강해지며, 반대로 거리가 멀어질수록 공간적 종속관계가 약해지는 특성을 나타낸다. 둘째, 공간모집단에서는 공간적 종속관계의 특성으로 인하여 서로 밀집된 집락을 형성하며 분포되어 존재한다.

공간모집단에서의 표본설계에 대한 연구 방향은 가정된 공간종속관계의 모형하에서 진행되고 있다. 그 주된 연구방향은 공간모집단을 어떤 적절한 형상(shape)으로 분할할 것인가 하는 분할형상에 대한 연구(Matern 1980, Yfantis, Flatman과 Behar 1987, Olea 1984, McBratney, Webster와 Burgess 1981a,b, Ripley 1981, Overton과 Stehman 1993)와 공간모집단에서의 관찰위치에 해당되는 표본점들의 추출방법에 대한 연구(Quenouille 1949, Matern 1980, Bellhouse 1977, 1981, 1988, Koop 1990), 공간 변동의 특성을 반영한 적합탐색 관찰방법에 의한 추정방법의 연구(Thompson 1990, 1991a,b, 1992, Thompson, Ramsey와 Seber 1992) 등으로 진행되고 있다.

공간적 특성을 지닌 공간모집단에서 평균, 총계, 비율 등과 같은 모수를 추정하기 위하여 전통적인 표본추출방법으로 표본설계를 하는 것보다는 공간모집단의 공간변동을 반영하여 표본설계를 하는 것이 더 바람직할 것이다. 사전에 결정된 표본에 대하여 실제조사를 진행해 나가면

1) (110-745) 서울 종로구 명륜동 3가 53, 성균관대학교 통계학과 강사.
2) (110-745) 서울 종로구 명륜동 3가 53, 성균관대학교 통계학과 교수.

서 모집단의 변동을 반영하여 관찰가능한 표본크기를 조정하는 적합탐색 관찰방법으로 추정해 보고자 한다. 이는 변화되는 공간모집단의 특성을 실제 조사에 반영하여 추정하는 방법이다.

2. 적합탐색 관찰과정

공간모집단에서 관심영역에 대한 면적이나 비율을 추정하기 위하여 충분히 많은 n개의 칸들로 분할한 후 각 칸(cell)에서 하나의 표본점을 추출하여 표본점계산(point-counting)기법으로 추정하는 방법에 관심을 갖는다. 이 때, 각 칸에서 표본점들은 다음과 같은 베르누이 확률변수가 된다. 즉, 모든 (i, j)번째 칸에 대하여

$$z_{ij} = \begin{cases} 1, & \text{표본점이 정의된 변수의 특성을 나타내는 경우} \\ 0, & \text{그렇지 않은 경우} \end{cases} \qquad (2\text{-}1)$$

이다. 여기서 z_{ij}는 베르누이분포를 따른다.

2-1. 적합탐색 관찰을 위한 가정

공간모집단은 어느 특정한 장소와 일정한 범위 내에 존재하는 이웃 장소들 사이에 특별한 연관성을 가지고 분포하게 된다. 공간모집단을 충분히 많은 칸으로 분할한다고 하면, 여러 개의 집락을 형성하며 분포되는 공간 특성으로 인하여 다음과 같은 상황이 발생할 수 있는 확률은 0이라고 가정한다.

$$\begin{array}{ccc} & 0 & \\ 0 & 1 & 0 \\ & 0 & \end{array} \qquad\qquad \begin{array}{ccc} & 1 & \\ 1 & 0 & 1 \\ & 1 & \end{array}$$

<그림 1> 고립된 지역의 특성

여러 개의 집락으로 형성되어 분포되는 특정한 몇몇 공간모집단의 경우에 적용 가능한 방법이다. 사전에 결정된 표본크기에 대하여 실제 표본조사를 진행하는 관찰과정에서 공간모집단의 특성을 이용한 적합탐색방법으로 추정하려는 목적하에 다음과 같은 사항을 가정한다.

- **첫째,** 조사연구지역이 충분히 큰 지역으로 조사연구지역 내에서의 추정하려는 영역의 형상이 하나의 폐곡선으로 형성된다.
- **둘째,** 조사연구지역 내에서 추정하려고 하는 영역의 형상은 고립된(isolated) 지역으로 존재하지 않으며 또한, 하나의 큰 형상으로 이루어진다고 하더라도 도우넛형태로는 존재하지 않는다고 가정한다.
- **셋째,** 전체 면적의 추정에 관심을 가지고 있으므로 조사지역 전체를 크기가 동일하고 가능한 정사각형이 되도록 충분히 많은 칸으로 분할이 가능하다고 가정한다.

2-2. 적합탐색 관찰과정

2-2-1. 초기 칸의 선택방법

크기가 동일하고 정사각형에 근사하도록 n개의 칸으로 분할한 공간모집단에서 단순임의추출법이나 계통추출법에 의해 초기 칸을 선택한다. 추출된 초기 칸들로 부터 탐색조건을 결정하고 이 초기 칸들을 네트워크의 출발 칸으로 정하여 탐색조건에 따라 네트워크를 구성한다.

2-2-2. 적합탐색 관찰조건의 결정

분할된 공간모집단에서 추출된 초기 칸들에 대하여 공간추출법으로 표본점의 좌표를 결정하고 표본점의 관찰값들을 먼저 탐색한다. 관찰된 표본점들의 값들은 식 (2-1)과 같은 결과를 가지게 된다.

적합탐색 관찰조건은 초기 칸들의 관찰값들을 이용하여 탐색조건을 결정하게 되는 데 관찰된 초기 칸들의 관찰값들에 따라 다음과 같이 결정한다.

$$Z = \begin{cases} 0 \ , & \text{모든 초기 칸들의 관찰값이 0인 경우} \\ 1 \ , & \text{그렇지 않은 경우} \end{cases} \tag{2-2}$$

여기서 모든 초기 칸들의 관찰 값이 모두 1이거나, 또는 0과 1이 혼합된 경우에는 칸의 관찰값이 1로 나타나는 경우에만 계속 탐색한다고 결정한다.

이와 같이 결정한 탐색조건을 이용하여 칸들의 표본점에 대한 관찰값들이 각 네트워크의 초기값과 일치되는 경우에 한하여 이웃한 칸들을 계속 탐색하게 된다.

2-2-3. 적합탐색 관찰 원칙

분할된 공간모집단에서 초기 칸들을 선택하고 초기 칸들의 관찰값들을 이용하여 식 (2-2)와 같이 탐색조건을 정한 후, 각각의 출발 칸들에 대하여 탐색조건을 만족하는 칸들의 이웃한 칸들을 탐색하여 나감으로써 초기 출발 칸에 대한 네트워크를 구성한다. 그리고 탐색조건을 만족하는 칸에 대해서는 계속적으로 이웃한 칸들을 탐색하여 하나의 네트워크를 구성하게 된다.

이러한 탐색조건을 만족하지 않아서 더 이상의 이웃 칸들에 대하여 지속적으로 탐색할 수 없는 경우에는 자신만 네트워크에 포함되게 되는 데, 이러한 칸을 가장자리(edge) 칸이라고 부르기로 한다. 따라서 하나의 네트워크에는 탐색조건을 만족하여 관찰된 칸들과 자신만 포함되는 가장자리 칸들로 구성되게 될 것이다. 이러한 적합탐색 관찰과정을 실행하기 위한 적합탐색 관찰 원칙을 다음과 같이 정하기로 한다.

첫째, 각각의 초기 출발 칸들에 대하여 네트워크를 구성하게 되는데 식 (2-2)에서 결정된 탐색조건 $\{z_{ij} \mid z_{ij} = z,\ z \in Z\}$을 만족하는 칸들에 대해서는 인접한 칸들에 대하여 오른쪽방향으로 추가 탐색하여 관찰한다. 인접 칸을 관찰할 경우 출발 칸들을 관찰하기 위

해 표본점을 추출한 방법과 동일한 공간추출법을 적용한다.

둘째, 탐색된 칸들은 모두 초기 칸의 네트워크에 포함시키게 되는데 결정된 탐색조건을 만족하는 칸들에 대해서는 이웃한 인접 칸들에 대하여 같은 방향으로 계속 탐색해 나간다.

셋째, 출발 칸의 관찰 값과 다른 관찰 값을 갖는, 즉 탐색조건을 만족하지 않는 가장자리 칸은 해당 칸만 네트워크에 포함시키고 동일한 방향으로의 탐색과정을 멈춘다.

넷째, 다시 최초의 출발 칸에 대하여 왼쪽방향으로 이동하여 가면서 적합탐색 관찰과정의 둘째, 셋째단계를 반복한다. 계속하여 아래쪽, 윗쪽방향으로 인접한 칸들에 대하여 적합탐색과정을 동일하게 적용한다.

다섯째, 적합탐색 관찰과정에서 한번 탐색된 칸에 대해서는 중복적으로 관찰하지 않으며 또한, 여러 네트워크에 중복적으로 포함시키지 않고 가장 가까운 출발 칸의 네트워크에 포함시킨다.

n개의 칸으로 분할된 공간모집단에 대하여 각 칸마다 추출된 표본점을 관찰하여 적합탐색원칙으로 실제 조사를 진행하게 된다. 이 결과로 부터 관찰되지 않는 칸들은 적합탐색조건에 의한 가장자리칸들로 경계지어 구분되게 된다. 즉, 적합탐색조건에 의해 구성된 네트워크의 정보를 조건부로 이용하여 경계내의 관찰되지 않는 칸의 정보를 알 수 있게 될 것이다.

3. 적합탐색 관찰을 이용한 추정

하나의 큰 집락으로 분포되어 있는 2차원 평면상에 존재하는 몇몇 공간모집단에서 적합탐색방법을 이용하여 관심영역의 면적을 추정하는 적합탐색에 의한 추정량은 적합탐색에 의한 관찰칸들로 구성된 네트워크에서의 추정 결과와 적합탐색과정의 네트워크를 조건부로 하여 알 수 있는 탐색되지 않는 칸들의 결과를 결합하여 추정하게 된다. 즉, 본 연구에서 제안하는 공간모집단에서의 전체 면적에 대한 추정량은 가정된 공간모집단의 제약조건하에서 적합탐색으로 관찰된 칸들의 면적과 관찰되지 않는 칸들에 대한 면적으로 분리하여 추정하는 방법을 제안하려는 것이다.

적합탐색 관찰방법에 의한 면적추정량을 검토하기 위하여 필요한 기호들을 다음과 같이 정의한다.

A : 연구조사지역의 면적

A_m : 적합탐색 과정으로 관찰된 전체 네트워크에서의 면적

A_{n-m} : 적합탐색 과정에 의해 관찰되지 않은 면적

A^{α} : 정의된 관심의 특성을 나타내는 관심영역의 면적

A^{α}_{m} : 관심영역중 적합탐색 과정으로 관찰된 전체 네트워크에 포함된 면적

A^{α}_{n-m} : 관심영역중 적합탐색 과정으로 관찰되지 않은 면적

n : 표본크기로서 공간모집단을 분할한 전체 칸의 수

공간모집단에서 추정하고자 하는 관심영역의 전체면적을 추정하는 다른 한 방법으로써 분할된 n개의 칸을 전부 관찰하는 기존의 관찰방법으로 추정한다면, 면적추정량($\widehat{A_0^\alpha}$)은

$$\widehat{A_0^\alpha}=A\cdot\frac{1}{n}\sum_{(ij)}z_{ij} \tag{3-1}$$

이 된다.

그러나 적합탐색 관찰방법으로 관심영역의 면적을 추정하기 위하여 다음과 같이 정의한다.

<정의> 공간모집단의 조사연구지역 내에서 관심영역의 면적(A^α)은 적합탐색에 의해 관찰된 네트워크에서의 면적(A_m^α)과 이 네트워크를 조건부로 하여 관찰되지 않은 칸들의 면적(A_{n-m}^α)에 대한 합으로 계산된다. 즉,

$$A^\alpha = A_m^\alpha + A_{n-m}^\alpha \tag{3-2}$$

이다.

정의에 의하여 2차원의 공간모집단에서 적합탐색 관찰방법으로 추정하는 관심영역의 면적추정량은 정의에 의해 적합탐색에 의한 관찰 칸들로 구성된 네트워크에서의 면적과 적합탐색의 관찰조건으로 부터 알 수 있는 관찰되지 않은 칸들에 대한 면적을 결합하여 추정하게 된다.

$$\begin{aligned}\widehat{A^\alpha} &= A_m\cdot\frac{1}{m}\sum_{(ij)\in\phi_m}z_{ij}+A_{n-m}\cdot\frac{1}{n-m}\sum_{(ij)\in\phi_{n-m}}z_{ij}\\ &= \widehat{A_m^\alpha}+\widehat{A^\alpha}_{n-m}\end{aligned} \tag{3-3}$$

그리고 본 절에서는 2-1절에서 가정한 것처럼 크기가 동일하게 분할된 공간모집단에서 적합탐색 관찰방법으로 관심영역의 면적추정량을 제안해 보려고 한다. 적합탐색에 의한 면적추정량은 정의를 이용하여 다음과 같은 정리들로써 보다 간단히 계산할 수 있다.

<정리 1> 크기가 동일한 n개의 사각형칸으로 분할된 공간모집단의 가정하에서 칸의 면적을 $A_c(=A/n)$라고 한다면, 적합탐색 관찰에 의한 면적추정량은

$$\widehat{A^\alpha} = A_c\sum_{(ij)\in\phi_m}z_{ij}+A_c\sum_{(ij)\in\phi_{n-m}}z_{ij} \tag{3-4}$$

이다.

증명) 크기가 동일한 n개의 칸들로 분할된 공간모집단에서 칸의 면적은 $A_c(=A/n)$이므로 $A_m=m\cdot A_c$이고, $A_{n-m}=(n-m)\cdot A_c$가 된다. 그러므로 $\widehat{A_m^\alpha}$와 $\widehat{A^\alpha}_{n-m}$은 다음과 같다.

$$\begin{aligned}
\widehat{A^a}_m &= A_m \cdot \frac{1}{m} \sum_{(ij)\in\phi_m} z_{ij} \\
&= m \cdot A_c \cdot \frac{1}{m} \sum_{(ij)\in\phi_m} z_{ij} \qquad (3\text{-}5)\\
&= A_c \sum_{(ij)\in\phi_m} z_{ij}
\end{aligned}$$

$$\begin{aligned}
\widehat{A^a}_{n-m} &= A_{n-m} \cdot \frac{1}{n-m} \sum_{(ij)\in\phi_{n-m}} z_{ij} \\
&= (n-m) \cdot A_c \cdot \frac{1}{n-m} \sum_{(ij)\in\phi_{n-m}} z_{ij} \qquad (3\text{-}6)\\
&= A_c \sum_{(ij)\in\phi_{n-m}} z_{ij}
\end{aligned}$$

따라서 크기가 동일한 n개의 칸으로 분할이 가능한 공간모집단에서의 적합탐색 관찰에 의한 추정량은 식 (3-5)와 식 (3-6)을 식 (3-2)에 대입하여 정리하면 다음과 같이 표현되므로 식 (3-4)는 성립한다.

$$\begin{aligned}
\widehat{A^a} &= \widehat{A^a_m} + \widehat{A^a}_{n-m} \\
&= A_c \sum_{(ij)\in\phi_m} z_{ij} + A_c \sum_{(ij)\in\phi_{n-m}} z_{ij}
\end{aligned}$$

<정리 2> 크기가 동일하고 똑같은 면적을 가지도록 공간모집단을 분할한다는 가정 하에서 관심영역에 대한 적합탐색 추정량 $\widehat{A^a}$는 A^a의 불편추정량이다.

증명) 모든 (i, j)번째 칸에서의 표본점들은 적합탐색 관찰 원칙에 의하여

$$\sum_{(ij)} z_{ij} = \sum_{(ij)\in\phi_m} z_{ij} + \sum_{(ij)\in\phi_{n-m}} z_{ij}$$

으로 나뉘어지게 된다. (i, j)번째 칸에서의 표본점에 대한 관찰값인 z_{ij}는 베르누이 확률변수이므로 z_{ij}의 기대값은 $E(z_{ij}) = p_{ij}$이다. 그러므로

$$\begin{aligned}
\sum_{(ij)} E(z_{ij}) &= \sum_{(ij)\in\phi_m} E(z_{ij}) + \sum_{(ij)\in\phi_{n-m}} E(z_{ij}) \\
&= \sum_{(ij)\in\phi_m} p_{ij} + \sum_{(ij)\in\phi_{n-m}} p_{ij} \qquad (3\text{-}7)\\
&= \sum_{(ij)} p_{ij}
\end{aligned}$$

이 된다.

공간모집단에서 크기가 동일하게 분할된다는 가정하에서

$$
\begin{aligned}
E(\widehat{A^{\alpha}}) &= E(\widehat{A^{\alpha}_{m}}) + E(\widehat{A^{\alpha}}_{n-m}) \\
&= A_m \cdot \frac{1}{m} \sum_{(ij) \in \phi_m} E(z_{ij}) + A_{n-m} \cdot \frac{1}{n-m} \sum_{(ij) \in \phi_{n-m}} E(z_{ij}) \\
&= A_c \sum_{(ij) \in \phi_m} E(z_{ij}) + A_c \sum_{(ij) \in \phi_{n-m}} E(z_{ij})
\end{aligned}
$$

이 된다. 따라서 면적추정량에 대한 기대값을 구하면

$$
\begin{aligned}
E(\widehat{A^{\alpha}}) &= A_c \sum_{(ij) \in \phi_m} p_{ij} + A_c \sum_{(ij) \in \phi_{n-m}} p_{ij} \\
&= A_c \cdot \sum_{(ij)} p_{ij} \\
&= A \cdot \frac{1}{n} \sum_{(ij)} p_{ij} \\
&= A \cdot p \\
&= A^{\alpha}
\end{aligned}
$$

이 된다. 여기서 $p = \frac{1}{n} \sum_{(ij)} p_{ij}$이다.

그리하여 공간모집단에서의 관심영역에 대한 면적은 적합탐색으로 관찰된 면적과 관찰되지 않은 칸들의 면적으로 추정하게 된다. 이로써 적합탐색 관찰방법에 의해 관찰되는 $m(\leq n)$개의 표본크기로써 추정하는 적합탐색에 의한 면적추정량이 기존 관찰방법에 의한 추정량과 유사한 추정결과를 제공하게 될 것이다.

4. 추정량의 분산

4.1 분산과 분산 추정량

공간모집단에서 관심영역의 면적에 관한 적합탐색에 의한 면적추정량은 적합탐색원칙에 의해 관찰된 결과와 적합탐색조건을 조건부로 관찰되지 않고 얻어진 결과를 결합하여 추정하고 있다. 이제 적합탐색추정량의 분산과 분산 추정량에 대하여 살펴보고자 한다.

적합탐색 관찰에 의한 추정량의 분산은 적합탐색 관찰에 의한 부분과 적합탐색 관찰조건에 의해 관찰되지 않고 알 수 있는 부분으로 분리하여 추정량의 분산을 검토할 수 있다. 그러나 관찰되지 않고 추정되는 추정량 $\widehat{A^{\alpha}}_{n-m}$은 적합탐색조건에 의해 결정되므로 상수가 된다.

(i, j)번째 칸에서의 표본점에 대한 관찰값인 z_{ij}는 베르누이 확률변수이므로 z_{ij}의 분산은 $Var(z_{ij}) = p_{ij}\ (1-p_{ij})$이다.

이를 이용하여 적합탐색 관찰에 의한 추정량의 분산은

$$\begin{aligned} Var(\widehat{A^{\alpha}}) &= A_m^{\,2} \cdot \frac{1}{m^2}\Big[\sum_{(ij)\in\phi_m} Var(z_{ij}) \Big] \\ &= A_m^{\,2} \cdot \frac{1}{m^2}\Big[\sum_{(ij)\in\phi_m} p_{ij}\ (1-p_{ij}) \Big] \end{aligned} \tag{3-8}$$

으로 표현된다.

분할된 공간모집단에서 각 칸들마다 하나의 표본점을 추출하는 방법을 고려하기 때문에 (i, j)번째 칸에서의 면적비율 p_{ij}를 추정할 수 없고, 적합탐색 관찰방법으로 관찰되는 표본크기 m을 알 수 없으므로 분산의 추정량은 직접 구할 수 없다. n개의 칸으로 구성된 공간모집단에서 적합탐색과정에 의해 구성된 K개의 독립표본으로 부터 추정된 적합탐색추정량 $\widehat{A^{\alpha}}_1$, $\widehat{A^{\alpha}}_2, \cdots, \widehat{A^{\alpha}}_K$ 을 이용하여 분산추정량을 구할 수 있다.

이 경우 K개의 적합탐색추정량의 평균을 $\overline{A^{\alpha}}$라고 하면, 적합탐색추정량의 분산추정량은

$$\widehat{Var}(\widehat{A^{\alpha}}) = \frac{1}{K\cdot(K-1)}\Big[\sum_{k=1}^{K} (\widehat{A^{\alpha}}_k - \overline{A^{\alpha}})^2 \Big] \tag{3-9}$$

으로 간단히 표현할 수 있다. 여기서 $\overline{A^{\alpha}} = \frac{1}{K}\sum \widehat{A^{\alpha}}_k$ 이다.

4.2 효율성비교

공간모집단에서 분할된 n개 칸을 전부 관찰하는 기존의 관찰방법에 의한 추정량과 본 연구에서 제안하는 적합탐색 관찰에 의한 추정량의 효율성을 비교해본다.

분할된 모든 칸을 전부 관찰하여 추정하는 기존 관찰방법에 대한 면적추정량($\widehat{A_0^{\alpha}}$)의 분산은

$$\begin{aligned} Var(\widehat{A_0^{\alpha}}) &= A^2 \cdot \frac{1}{n^2}\Big[\sum_{(ij)} Var(z_{ij}) \Big] \\ &= A^2 \cdot \frac{1}{n^2}\Big[\sum_{(ij)} p_{ij}\ (1-p_{ij}) \Big] \end{aligned} \tag{3-10}$$

으로 나타낼 수 있고, 적합탐색 관찰에 의한 면적추정량의 분산은 식 (3-8)과 같다.

기존의 관찰방법에 의한 추정량과 적합탐색 관찰방법에 의한 추정량의 효율성을 살펴보기 위하여 식 (3-8)과 식 (3-10)를 이용하여 상대효율을 비교하면

$$\frac{Var(\widehat{A^{\alpha}})}{Var(\widehat{A_0^{\alpha}})} = (\frac{A_m}{A})^2 \cdot (\frac{n}{m})^2 \cdot \frac{\sum_{(ij)\in\phi_m} p_{ij}(1-p_{ij})}{\sum_{(ij)} p_{ij}(1-p_{ij})} \tag{3-11}$$

이 된다. 식 (3-11)으로 부터 두 추정량의 효율성은 적합탐색 관찰방법으로 관찰되는 표본크기 m을 알 수 없으므로 효율성은 직접 비교할 수 없으며, 또한

$$\frac{\sum_{(ij)\in\phi_m} p_{ij}(1-p_{ij})}{\sum_{(ij)} p_{ij}(1-p_{ij})}$$

에 영향을 받게 된다.

5. 모의자료를 이용한 실증분석

공간특성이 존재하는 2차원 공간모집단에서 관심영역의 분포면적이나 비율을 추정하는 데 공간특성을 이용한 적합탐색 관찰방법에 의한 추정방법을 모의자료를 이용한 실증분석으로 검토해 보고자 한다.

크기가 40X40인 가상의 공간모집단[부록]을 가정하였고, 하나의 폐곡선으로 나타나는 관심영역의 형상에 대한 면적을 기존의 관찰방법에 의한 추정결과와 적합탐색 관찰방법에 의한 추정결과를 비교한다. 적합탐색 관찰에 의한 추정방법을 검토하기 위하여 공간모집단에서의 표본크기는 n=64, 100으로 정하였다. 공간모집단에서의 표본크기는 모집단을 분할하는 칸의 수와 일치하게 하였다. 따라서 2차원의 각 축으로 할당하는 표본점 위치의 수는 공간모집단을 정사각형 격자로 분할할 수 있도록 8X8, 10X10으로 정하였다. 공간모집단에서 관심영역의 면적을 추정하기 위해 층화-층화추출법중 st_0st_0방법을 적용한다.

가상의 공간모집단에서 적합탐색 관찰과정을 실행하기 위하여 최초로 선택되는 초기 칸들은 확률추출법에 의해 선택해야 하지만 분할된 공간모집단을 모두 탐색한다는 기본원칙에서 최초로 선택된 초기 칸으로 4개의 모서리칸을 선택하였다. 이들 초기 칸에서 임의-임의추출법으로 표본점의 좌표를 정하는 층화-층화추출법으로 표본점을 선택하여 이들 초기 칸에서 표본점을 관찰하여 적합탐색 관찰 조건을 z=0으로 설정하고 적합탐색 원칙으로 표본점을 관찰한다.

적합탐색 관찰에 의한 관심영역 면적의 추정량 및 최소제곱오차(MSE)를 살펴보기 위하여 모의자료를 이용하여 구성된 가상의 공간모집단에서 1000번의 반복실험을 실시하였고, 면적추정량및 최소제곱오차는 반복실험의 평균으로 나타내었다.

추정하고자 하는 관심영역의 형상이 불규칙하고 치우친 정도와 굴곡이 심하게 나타나도록 가정된 공간모집단으로 관심영역의 면적은 A^{α}=560이고, 비율은 p=0.35이다. 모의자료로 구성된 공간모집단에서 주어진 표본크기에 대하여 각각 1000번의 시행을 반복한 실증분석 결과를 요약하면 <표>와 같다.

<표> 굴곡이 심한 불규칙한 형상을 나타내는 공간모집단에 대한 결과

표본 크기	기존 관찰 방법			적합탐색 관찰방법				
	실제 관찰 크기	추정량	MSE	실제관찰크기			추정량	MSE
				평균	관찰범위			
					최소	최대		
64 (8X8)	64	569.28	2371.38	55.67 (1.82)	50	61	573.83	2397.88
100 (10X10)	100	552.10	1295.36	85.00 (2.50)	76	92	555.31	1442.05

굴곡이 심한 불규칙한 형상을 나타내는 공간모집단에 대하여 적합탐색 관찰방법을 이용하여 1000번의 반복실험으로 추정해 본 결과, 주어진 표본크기가 $n=64$일 때 실제 관찰된 표본크기는 평균 $m=55.67$로 나타나 평균 8.33개의 표본은 관찰하지 않는 것으로 나타났으며 표본크기의 범위는 50에서 61로 나타나고 있다. 표본크기가 $n=100$인 경우에도 실제 관찰된 표본크기의 범위는 76에서 92로 나타났으며 평균 관찰 표본크기는 $m=85.00$으로 나타나 15.00개의 표본은 관찰하지 않는 것으로 나타났다.

주어진 표본크기에 대하여 모든 칸들을 관찰하는 기존 관찰방법과 적합탐색 관찰방법에서의 결과들을 비교해 보면 추정량은 차이가 크지는 않지만, 추정량의 효율성비교에서는 $n=64$일 때 1.01(=2397.88/2371.38)로 차이가 크지 않으나, $n=100$일 때 1.11(=1442.05/1295.36)으로 기존 방법에 비하여 본 연구에서 제안된 방법이 다소 효율성이 감소하는 것으로 나타나고 있다. 그러나 주어진 공간모집단에서 적합탐색 관찰방법으로 면적을 추정한다면 기존의 관찰방법에 비하여 약 10%이상의 표본크기를 감소시켜 주게 되고, 추정량의 효율성은 다소 떨어지지만 면적추정량이 큰 차이를 보여 주지 않고 있으므로 면적추정을 위한 효율적인 표본설계방법으로 그 이용 가치가 있다고 생각된다.

그리고 10X10으로 분할된 공간모집단의 칸내에서 적합탐색 관찰방법으로 하나의 표본점을 추출해 가면서 탐색된 관심영역의 형상에 대한 그림은 <그림 2>과 같이 나타났다.

<그림 2>에서 각 칸을 십단위의 숫자로 표현하고 있는 데 십단위는 네트워크를 표현한 것이고 단단위는 표본점의 관찰값을 나타낸 것이다. [부록]의 공간모집단은 하나의 네트워크로 구성이 되고 있음을 볼 수 있으므로 나머지 3개의 네트워크는 고려하지 않았다. 그리고 '9'는 적합탐색조건에 의해 가장자리칸들의 경계내에 포함되어 있기 때문에 적합탐색원칙에 의해 해당 칸내 표본 관찰값의 특성을 알 수 있다는 가정하에서 관찰하지 않는 칸을 나타낸 것이다.

그러므로 분할된 공간모집단의 칸들내에서 표본점을 관찰할 때 '9'의 갯수만큼 실제로 관찰하지 않게 된다. 따라서 <그림 2>에서처럼 칸내에서 표본점의 관찰값이 '1'로 나타나는 칸들을 연결하여 공간모집단에서의 관심영역에 대한 형상을 파악할 수 있게 된다.

10	10	10	10	10	10	10	10	10	10
10	10	10	10	10	10	11	10	10	10
10	10	10	10	11	11	11	10	11	11
10	11	11	11	9	9	9	11	9	11
10	11	9	9	9	9	9	9	11	10
10	10	11	11	11	9	9	11	10	10
10	10	10	10	10	11	11	11	10	10
10	10	10	10	10	10	10	11	10	10
10	10	10	10	10	10	10	10	10	10
10	10	10	10	10	10	10	10	10	10

<그림 2> 가상의 공간모집단에 대하여 적합탐색 관찰방법으로
파악된 관심영역의 형상에 대한 형태

6. 결 론

본 연구에서는 2차원의 공간모집단에서 주어진 표본크기에 대한 표본을 관찰할 때 모집단의 변화하는 특성이나 공간 특성을 이용하여 적합탐색 관찰방법으로 표본을 관찰하여 추정하는 추정량을 제안하는 데 목적을 두고 살펴보았다. 제한적인 몇몇 공간모집단에 대해서 n개의 동일한 크기를 갖는 사각형의 칸들로 분할하여 각 칸에서 추출한 표본점을 적합탐색으로 관찰하는 적합탐색 관찰로 추정하는 방법이다. 즉, 분할된 공간모집단의 칸들에 대한 정보를 이용하여 유사한 특성을 가지는 칸들을 동일한 네트워크로 구성함으로써 이웃 칸들에 대한 추가적인 관찰 여부를 탐색적으로 판단하는 적합탐색 관찰방법에 의한 추정방법이다.

표본조사에서 실제조사 이전에 정해진 표본크기 n에 대한 기존의 관찰방법은 추출된 모든 표본을 관찰하여 표본을 구성하는 데 반하여 본 연구에서 제안된 적합탐색 관찰방법에서는 적합탐색조건을 만족하는 이웃 칸들에 대하여 표본점을 추출하여 관찰하게 된다.

적합탐색 관찰에 의한 추정량은 n개로 분할된 모든 칸에서 표본점을 추출하여 관찰하는 것이 아니라 공간적 특성을 나타내고 있는 칸들의 정보를 이용하고 있다. 적합탐색 관찰로 네트워크를 구성하는 경우, 분할된 n개의 칸중에서 $m(\le n)$개의 칸만을 관찰하여 관심영역의 비율이나 면적을 추정할 수 있다는 점이다.

몇몇 공간모집단의 경우 본 연구에서 제안한 것처럼 특정한 공간 특성을 이용하여 적합탐색 관찰방법으로 추정하게 되면 다음과 같은 효과를 얻게 될 것으로 기대된다.

첫째, 적합탐색 관찰로 관찰하므로 관찰하게 되는 표본크기를 감소시키는 효과가 기대되므로 표본조사에서 발생하는 경비와 시간 등의 절감효과가 나타날 것이며

둘째, 추정하려는 관심영역의 형상에 대한 사전 가정이 존재하지 않더라도 적합탐색 관찰방법에 의해서 관심영역에 대한 정확한 형상을 파악할 수는 없지만 구체적인 형상을 파

악하기 위한 실제 조사에 앞서 관심영역의 형상에 관한 대략적인 윤곽을 쉽게 파악할 수 있다는 점이다.

그리하여 공간모집단에서 정확한 관심영역의 형상은 표현할 수 없지만 관심영역의 대략적 형상을 파악할 수 있으므로 공간모집단의 특성을 파악하기 위한 예비조사나 정확한 형상을 파악하기 위한 정밀조사 등의 사전조사로도 이용할 수 있을 것이다.

특히 동,식물에 대한 분포 면적이나 광물, 석유, 석탄 등의 매장면적, 환경이나 수질오염지역에서의 오염면적과 같은 2차원 평면상에 존재하는 몇몇 공간모집단에서는 가정에서 언급된 것처럼 하나의 집락을 형성하여 분포될 수 있으므로 적합탐색에 의해 구성된 네트워크의 결과와 네트워크를 조건부로 한 탐색되지 않은 칸들의 결과를 결합하여 분포면적이나 매장면적, 오염지역의 면적 등을 추정하는 것이 가능해지게 된다. 이처럼 주어진 공간모집단에서 면적을 추정한다고 할 때 적합탐색 관찰방법으로 면적을 추정한다면 기존의 관찰방법에 비하여 실제 관찰하는 표본크기를 크게 감소시켜 주기 때문에 면적추정을 위한 효율적인 표본설계방법으로 적용할 수 있을 것으로 사료된다.

향후 전통적인 여러 표본설계에 대한 적합탐색 관찰방법의 검토와 2차원의 공간을 확장한 3차원의 공간모집단에서의 공간표본설계방법에 대한 적합탐색 관찰방법의 검토를 고려할 수 있을 것으로 생각된다.

참고문헌

[1] Bellhouse, D.R. (1977). Some optimal designs for sampling in two dimensions, *Biometrika*, Vol. 64, 605-612.

[2] Bellhouse, D.R. (1981). Area estimation by point-counting techniques, *Biometrics*, Vol. 37, 303-312.

[3] Bellhouse, D.R. (1988). Spatial Sampling, *Encyclopedia of statistical. series*, John Wiley and Sons, 8, 581-584.

[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] McBratney, R.W., Webster,R., and Burgess,T.M. (1981a). The design of optimal sampling schemes for local estimation and mapping of regionaliged variables-I, *Computers and Geosciences*, 7, 331-334.

[7] McBratney, R.W., Webster,R., and Burgess,T.M. (1981b). The design of optimal sampling schemes for local estimation and mapping of regionaliged variables-II, *Computers and Geosciences*, 7, 335-336.

[8] Olea, R.A. (1984). Sampling design optimization for spatial functions, *Mathematical Geology*, 16, 369-392.

[9] 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.

[10] Quenouille, M.H., (1949). Problems in plane sampling, *Annals of Mathematical Statistics*, Vol. 20, 355-375.

[11] Ripley, B. D. (1981). *Spatial Statistics*, New York : Wiley.

[12] Thompson, S.K. (1990). Adaptive cluster sampling, *J. Amer. Statist. Assoc.*, Vol. 85, 1050-1059.

[13] Thompson, S.K. (1991a). Adaptive cluster sampling : Designs with primary and secondary units, *Biometrics*, Vol. 47, 1103-1115.

[14] Thompson, S.K. (1991b). Stratified adaptive cluster sampling, *Biometrika*, Vol. 78, 389-397.

[15] Thompson, S.K. (1992). *Sampling*, New York : Wiley.

[16] Thompson, S.K., Ramsey, F.L. and Seber, G.A.F. (1992). An adaptive procedure for sampling animal populations, *Biometrics*, Vol. 48, 1195-1199.

[17] Yfantis, E.A., Flatman, G.T., and Behar, J.V., (1987). Efficency of Kriging estimation for square, trianglar, and hexagonal grids, *Mathematical Geology*, 19, 183-205.

부록. 모의자료에 의한 가상적인 공간모집단

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	0	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1	1	0	0	1	1	0	0	1	1	0	0	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	0	0	1	1	0	0	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	1	1	0	0	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
0	0	0	0	0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
0	0	0	0	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Estimation Using the Method of Adaptive Searching Observation

Jong-seok Byun[3], Pyong Namkung[4]

Abstract

We propose an adaptive searching method using some spatial relations among sample points to estimate the interesting area in the spatial population. The fundamental idea is to observe the neighboring sample points when a sample point is satisfied with some condition of an adaptive searching observation.

For observing the sample points with this method to estimate the area the sample size is decreased. From this result, we may expect to reduce the cost and time consuming in observation the sample points and to draw the shape of the interesting area without prior information of an spatial population.

Some analytical simulation results are also presented.

3) Lecturer, Department of Statistics, Sung Kyun Kwan University, 3-53 MyungRyun-Dong Chongro-Gu, Seoul, 110-745, Korea.

4) Professor, Department of Statistics, Sung Kyun Kwan University, 3-53 MyungRyun-Dong Chongro-Gu, Seoul, 110-745, Korea.