

Discussion Paper No. 76J

空間集計単位の変換による
個数データの誤差から見た
空間集計システムの評価

貞広幸雄*

1998年6月

*東京大学空間情報科学研究センター

〒113 東京都文京区本郷7-3-1

1. はじめに

空間解析では、ある空間集計単位で集計された個数データを別の空間集計単位に変換するという操作がよく行われる。例えば、鉄道駅を中心とする一定距離内の地域の居住者を知りたい場合、駅を中心とする円を描き、町丁目別人口データと重ね合わせて人口を推定する。コンビニエンスストアの商圈人口を推定する場合には、コンビニエンスストアの分布に基づいたポロノイダイアグラムを作成し、ポロノイ領域内の人口を推定する。このような操作はareal interpolationと呼ばれており（Goodchild and Lam, 1980; Lam, 1983; Flowerdew and Green, 1991），その方法についてはこれまで様々なものが提案されている。

例えば、最も簡単な方法としてはpoint-in-polygon法（Burrough and McDonnell, 1998）があり、これは、元の空間集計単位の代表点が新たな空間集計単位に含まれる場合のみ、元の空間集計単位の持つ個数データをそのまま加算するという方法である。point-in-polygon法は操作が非常に簡単であるという利点を持つが、一方では多くのデータ誤差をもたらすという問題点もある。別の方法としては、面積按分法（Fisher and Langford, 1995），pynophylactic法（Tobler, 1979），さらに、補助的な別の空間データを用いる方法（Flowerdew, 1988; Goodchild et al., 1993）など、数多くの方法が提案され、それぞれの特性について議論が行われてきている（Fisher and Langford, 1995）。

これらの方法の中で、方法の簡便さと結果の安定性のバランスが優れていることから広く用いられているのが面積按分法である。この方法では、元の空間集計単位の内部で空間オブジェクトが均一に分布しているものと仮定し、新たな空間集計単位と重なる部分の面積に応じて個数データを比例配分する。しかしこの方法は、均一分布という強い仮定に基づいているため、得られる推定値は通常、真の値とは異なる。この誤差が小さければそれほど問題ではないが、大きい場合にはその後の分析に大きな影響をもたらすおそれがある。従って面積按分法を用いる場合には、推定値の誤差に十分気をつける必要がある。

推定値の誤差を小さくするには、できるだけ小さな空間集計単位のデータを元データとして用いることが望ましいと考えられる。例えば様々な店舗の商圈人口を推定しようとする場合、元データとしては街区単位、町丁目単位、市区町村単位、メッシュ単位など、様々な空間集計単位の人口データが利用可能であり、それらの中から費用と精度に見合ったデータを選択することができる。しかし現在のところ、空間集計単位と推定値精度の関係は未だ明らかになっておらず、適当な精度のデータを選択することは現実には難しい。

このような問題を解決する一つの方法は、空間集計単位と推定値精度の関係を定量的に記述し、空間集計システムの評価を行うことである。小さな空間集計単位のデータを用いるとどのくらい推定値精度が良くなるのか、空間集計単位の形状は推定値精度に影響を与えるのか、といった問いに明確に答えることができれば、データ選択の際に役立つ指針を提供することができる。

そこで本論文では、面積按分法を用いた個数データの推定誤差を、元の空間集計単位、新たな空間集計単位、空間オブジェクト分布の3つに基づいて定量的に記述、評価を行う。推定値の誤差はこれら3つの要素に依存しており、これらが全て既知の場合には、誤差を正確に知ることが可能である。しかし通常は、真の空間オブジェクト分布を知ることは不可能であり、全オブジェクトの個数か、各集計単位地区での集計値くらいしか得ることができない。また、データの利用が曖昧な場合、新たな空間集計単位を定めることも難しい。そこでここでは、真の空間オブジェクト分布や新たな空間集計単位については様々な可能性を考えられるように確率的モデルを用いて記述し、推定誤差との関係を明らかにする。

以下、2節では面積按分法を用いた個数データの推定誤差を確率的モデルに基づいて定式化する。3節では、2節で得られた結果をもとに、いくつかの典型的な場合について推定誤差の期待値と分散を導出する。4節では、3節で得られた分析結果を更に解析的に検討し、推定誤差と空間集計システムの関係进行分析する。5節では、実際によく用いられる空間集計システムについて、推定誤差 ε の分散を数値的に計算し、それら相互間の関係を明らかにする。6節では、結論をまとめ今後の課題について述べる。

2. 推定誤差の定式化と期待値・分散の導出

前節において述べたように、面積按分法を用いた個数データの推定誤差は、元の空間集計単位、新たな空間集計単位、空間オブジェクト分布の3つに大きく依存している。そこで本節では、これらの関係を定量的に記述するために、確率的モデルに基づいた定式化を行う。

いま、面積 A_0 を持つ領域 S_0 の中に N 個の点オブジェクトがそれぞれ独立に確率密度関数 $f(\mathbf{x})$ に従って分布しているものと仮定する。そして領域 S_0 は L 個の単位地区に区切られており、単位地区 U_i の面積を A_i とする。各単位地区 U_i にはその中に含まれる点オブジェクトの個数 n_i が記録されている。

一方、一つの新たな空間集計単位 S （面積を A とする）領域 S_0 に内包されるようにある確率分布に従って分布するものと仮定する。そして、その内部に含まれる点オブジェクトの個数は、各単位地区 i に記録されている個数 n_i から面積按分法によって推定する。

このような前提のもとで、個数の推定誤差を考えてみよう。まず、点 j の位置を \mathbf{z}_j とする。また、関数 $c(\mathbf{x})$ を以下のように定義する。

$$c(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S \\ 0 & \text{else} \end{cases} \quad (2-1)$$

すると、 S に含まれる点オブジェクトの真の個数 M は、

$$M = \sum_j c(\mathbf{z}_j) \quad (2-2)$$

と表される。さらに、関数 $R_i(\mathbf{x})$ を

$$R_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in U_i \\ 0 & \text{else} \end{cases} \quad (2-3)$$

と定義すると、面積按分法による点オブジェクト個数の推定値 \hat{M} は、

$$\hat{M} = \sum_i \left\{ \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}}{A_i} \sum_j R_i(\mathbf{z}_j) \right\} \quad (2-4)$$

となる。従って、点オブジェクト個数の推定誤差 ε は、

$$\begin{aligned} \varepsilon &= M - \hat{M} \\ &= \sum_j c(\mathbf{z}_j) - \sum_i \left\{ \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}}{A_i} \sum_j R_i(\mathbf{z}_j) \right\} \end{aligned} \quad (2-5)$$

となる。

本論文では、上の式で表される推定誤差に基づいて空間集計システムの評価を行うわけであるが、この推定誤差は点オブジェクトの分布や新たな空間集計単位 S の位置によって確率的に変動する値であり、値をそのまま用いた評価はできない。そこで以下では、推定誤差 ε の期待値と分散を算出し、それらの大きさによって空間集計システムを評価する。

まず準備のため、関数 $c(\mathbf{z}_j)$ と $R_i(\mathbf{x})$ の期待値を求めておこう。関数 $c(\mathbf{z}_j)$ の期待値 $E[c(\mathbf{z}_j)]$ は点 j が S に含まれる確率を表しており、これは微小領域 $d\mathbf{x}$ が S に含まれ、同時に点 j が微小領域 $d\mathbf{x}$ に含まれる確率を、領域 S_0 内で積分することにより与えられる。即ち、

$$\begin{aligned} E[c(\mathbf{z}_j)] &= \int_{\mathbf{x} \in S_0} \Pr[(\mathbf{z}_j \in d\mathbf{x}) \cap (d\mathbf{x} \in S)] d\mathbf{x} \\ &= \int_{\mathbf{x} \in S_0} \Pr[\mathbf{z}_j \in d\mathbf{x}] \Pr[d\mathbf{x} \in S] \end{aligned} \quad (2-6)$$

となる。ここで、 $d\mathbf{x}$ は微小領域であるので、

$$\Pr[d\mathbf{x} \in S] \approx \Pr[\mathbf{x} \in S] \quad (2-7)$$

とおくことができる。この式を式2-6に代入すると、

$$\begin{aligned}
E[c(\mathbf{z}_j)] &= \int_{\mathbf{x} \in S_0} \Pr[\mathbf{z}_i \in d\mathbf{x}] \Pr[d\mathbf{x} \in S] \\
&= \int_{\mathbf{x} \in S_0} f(\mathbf{x}) d\mathbf{x} \Pr[\mathbf{x} \in S] \\
&= \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x}
\end{aligned} \tag{2-8}$$

を得る．一方 $R_i(\mathbf{z}_j)$ の期待値とは，点 j が単位地区 U_i に含まれる確率であり，これは

$$\begin{aligned}
E[R_i(\mathbf{z}_j)] &= \Pr[\mathbf{z}_j \in S] \\
&= \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{2-9}$$

と表される．

それでは，これらの計算結果をふまえ，推定誤差 ε の期待値を見てみよう．

$$\begin{aligned}
E[\varepsilon] &= E\left[\sum_j c(\mathbf{z}_j) - \sum_i \left\{ \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}}{A_i} \sum_j R_i(\mathbf{z}_j) \right\} \right] \\
&= \sum_j E[c(\mathbf{z}_j)] - \sum_i \frac{1}{A_i} E\left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \sum_j R_i(\mathbf{z}_j) \right] \\
&= N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x} - \sum_i \frac{1}{A_i} E\left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \sum_j R_i(\mathbf{z}_j) \right]
\end{aligned} \tag{2-10}$$

ここで， $c(\mathbf{x})$ と $R_i(\mathbf{z}_j)$ は独立であるので，第2項を書き換えて

$$\begin{aligned}
E[\varepsilon] &= N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x} - \sum_i \frac{1}{A_i} E\left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \right] E\left[\sum_j R_i(\mathbf{z}_j) \right] \\
&= N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x} - N \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} E[c(\mathbf{x})] d\mathbf{x} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \\
&= N \left\{ \int_{\mathbf{x} \in S_0} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} - \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} \Pr[\mathbf{x} \in S] d\mathbf{x} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \right\} \\
&= N \sum_i \left\{ \int_{\mathbf{x} \in U_i} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} - \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} \Pr[\mathbf{x} \in S] d\mathbf{x} \right\}
\end{aligned} \tag{2-11}$$

となる．

式2-11を見ると， $E[\varepsilon]$ は常に0になるというわけではないことが分かる． $E[\varepsilon]$ が0になるのは，例えば $f(\mathbf{x})$ が各単位領域 U_i 内で均一分布である場合である．これは即ち，面積按分法が合理的であるとみなされる「点オブジェクトが一様に分布している」という条件に相当する．あるいは， $\Pr[\mathbf{x} \in S]$ が \mathbf{x} によらず一定値である場合にも， $E[\varepsilon]$ は0になる．これは即ち， S が領域 S_0 に関して境界条件なしに一様分布する，ということを表しており，Sadahiro (1999)にあるように，1) S_0 が長方形や正六角形などのように，latticeとして平面を完全に被覆することのできる図形である場合，2) S が S_0 に比べて大きさを無視できるほど小さい場合，には $\Pr[\mathbf{x} \in S]$ が \mathbf{x} によらず一定であるという仮定を置くことが可能である．このうち1)については後で詳述する．しかし，一般的には $E[\varepsilon]$ は0ではなく，これが，面積按分法が合理的であるとは限らないという論拠になっている．

それでは，推定誤差 ε の分散はどのようになるのであろうか．

$$\begin{aligned}
V[\varepsilon^2] &= E[\varepsilon^2] - (E[\varepsilon])^2 \\
&= E\left[(M - \hat{M})^2 \right] - (E[\varepsilon])^2 \\
&= E[M^2] - 2E[M\hat{M}] + E[\hat{M}^2] - (E[\varepsilon])^2
\end{aligned} \tag{2-12}$$

各項を個別に見てゆくと，第1項は

$$\begin{aligned}
E[M^2] &= E\left[\left\{\sum_j c(\mathbf{z}_j)\right\}^2\right] \\
&= E\left[\sum_j \sum_{j'} c(\mathbf{z}_j)c(\mathbf{z}_{j'})\right] \\
&= \sum_j \sum_{j'} E[c(\mathbf{z}_j)c(\mathbf{z}_{j'})] \\
&= 2 \sum_{j \neq j'} E[c(\mathbf{z}_j)c(\mathbf{z}_{j'})] + \sum_j E[\{c(\mathbf{z}_j)\}^2]
\end{aligned} \tag{2-13}$$

となる．ここで，各点はそれぞれ独立に分布するので， $c(\mathbf{z}_j)$ と $c(\mathbf{z}_{j'})$ （但し $j \neq j'$ ）は独立である．また，関数 $c(\mathbf{x})$ は2値関数であるので，式2-13は

$$\begin{aligned}
E[M^2] &= 2 \sum_{j \neq j'} E[c(\mathbf{z}_j)c(\mathbf{z}_{j'})] + \sum_j E[\{c(\mathbf{z}_j)\}^2] \\
&= 2 \sum_{j \neq j'} E[c(\mathbf{z}_j)]E[c(\mathbf{z}_{j'})] + \sum_j E[c(\mathbf{z}_j)]
\end{aligned} \tag{2-14}$$

と表される．さらに式2-8を用いると，

$$\begin{aligned}
E[M^2] &= 2 \sum_{j \neq j'} E[c(\mathbf{z}_j)]E[c(\mathbf{z}_{j'})] + \sum_j E[c(\mathbf{z}_j)] \\
&= N(N-1) \left\{ \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x} \right\}^2 + N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] \\
&= N(N-1) \int_{\mathbf{t} \in S_0} \int_{\mathbf{x} \in S_0} f(\mathbf{x}) f(\mathbf{t}) E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{x} d\mathbf{t} + N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x} \\
&= N(N-1) \int_{\mathbf{t} \in S_0} \int_{\mathbf{x} \in S_0} f(\mathbf{x}) f(\mathbf{t}) \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x} d\mathbf{t} + N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x}
\end{aligned} \tag{2-15}$$

となる．

次に，第2項を見てみよう．

$$\begin{aligned}
E[M\hat{M}] &= E\left[\left\{\sum_j c(\mathbf{z}_j)\right\} \left\{\sum_i \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}}{A_i} \sum_j R_i(\mathbf{z}_j)\right\}\right] \\
&= E\left[\sum_i \sum_j \sum_{j'} \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}}{A_i} R_i(\mathbf{z}_j)c(\mathbf{z}_{j'})\right] \\
&= \sum_i \sum_j \sum_{j'} \frac{1}{A_i} E\left[R_i(\mathbf{z}_j)c(\mathbf{z}_{j'}) \int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}\right]
\end{aligned} \tag{2-16}$$

さらに，

$$\begin{aligned}
E[M\hat{M}] &= \sum_i \sum_j \sum_{j'} \frac{1}{A_i} E\left[R_i(\mathbf{z}_j)c(\mathbf{z}_{j'}) \int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}\right] \\
&= 2 \sum_i \sum_{j \neq j'} \frac{1}{A_i} E\left[R_i(\mathbf{z}_j)c(\mathbf{z}_{j'}) \int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}\right] + \sum_i \sum_j \frac{1}{A_i} E\left[R_i(\mathbf{z}_j)c(\mathbf{z}_j) \int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}\right] \\
&= 2 \sum_i \sum_{j \neq j'} \frac{1}{A_i} E[R_i(\mathbf{z}_j)]E[c(\mathbf{z}_{j'})]E\left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}\right] + \sum_i \sum_j \frac{1}{A_i} \int_{\mathbf{x} \in U_i} E[R_i(\mathbf{z}_j)c(\mathbf{z}_j)c(\mathbf{x})] d\mathbf{x}
\end{aligned} \tag{2-17}$$

である．ここで第2項だけに着目すると，

$$\begin{aligned}
E[R_i(\mathbf{z}_j)c(\mathbf{z}_j)c(\mathbf{x})] &= \int_{\mathbf{t} \in U_i} \Pr[(\mathbf{z}_j \in d\mathbf{t}) \cap (d\mathbf{t} \in S) \cap (\mathbf{x} \in S)] \\
&= \int_{\mathbf{t} \in U_i} \Pr[\mathbf{z}_j \in d\mathbf{t}] \Pr[(d\mathbf{t} \in S) \cap (\mathbf{x} \in S)] \\
&\approx \int_{\mathbf{t} \in U_i} f(\mathbf{t}) d\mathbf{t} \Pr[(\mathbf{t} \in S) \cap (\mathbf{x} \in S)] \\
&= \int_{\mathbf{t} \in U_i} f(\mathbf{t}) E[c(\mathbf{t})c(\mathbf{x})] d\mathbf{t}
\end{aligned} \tag{2-18}$$

であるので，これを式2-17に代入すると，

$$\begin{aligned}
E[MM] &= 2 \sum_i \sum_{j \neq i} \frac{1}{A_i} E[R_i(\mathbf{z}_j)] E[c(\mathbf{z}_j)] E\left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}\right] \\
&\quad + \sum_i \sum_j \frac{1}{A_i} \int_{\mathbf{x} \in U_i} E[R_i(\mathbf{z}_j)c(\mathbf{z}_j)c(\mathbf{x})] d\mathbf{x} \\
&= N(N-1) \int_{\mathbf{x} \in S_0} f(\mathbf{x}) E[c(\mathbf{x})] d\mathbf{x} \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} E[c(\mathbf{x})] d\mathbf{x} \\
&\quad + N \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} \int_{\mathbf{t} \in U_i} f(\mathbf{t}) E[c(\mathbf{t})c(\mathbf{x})] d\mathbf{t} d\mathbf{x} \\
&= N(N-1) \int_{\mathbf{x} \in S_0} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} \Pr[\mathbf{x} \in S] d\mathbf{x} \\
&\quad + N \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} \int_{\mathbf{t} \in U_i} f(\mathbf{t}) \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x}
\end{aligned} \tag{2-19}$$

を得る．最後に，

$$\begin{aligned}
E[\hat{M}^2] &= E\left[\left\{\sum_i \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x}}{A_i} \sum_j R_i(\mathbf{z}_j)\right\}^2\right] \\
&= E\left[\sum_i \sum_{i'} \frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} c(\mathbf{x}) d\mathbf{x}}{A_i A_{i'}} \sum_j R_i(\mathbf{z}_j) \sum_{j'} R_{i'}(\mathbf{z}_{j'})\right] \\
&= \sum_i \sum_{i'} \sum_j \sum_{j'} E\left[\frac{\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} c(\mathbf{x}) d\mathbf{x}}{A_i A_{i'}} R_i(\mathbf{z}_j) R_{i'}(\mathbf{z}_{j'})\right] \\
&= \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} E\left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} c(\mathbf{x}) d\mathbf{x}\right] \sum_j \sum_{j'} E[R_i(\mathbf{z}_j)] E[R_{i'}(\mathbf{z}_{j'})]
\end{aligned} \tag{2-20}$$

であり，式2-9を用いて変形すると

$$\begin{aligned}
E[\hat{M}^2] &= N(N-1) \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} E \left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} c(\mathbf{x}) d\mathbf{x} \right] E[R_i(\mathbf{z}_j)] E[R_{i'}(\mathbf{z}_{j'})] \\
&\quad + N \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} E \left[\int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} c(\mathbf{x}) d\mathbf{x} \right] E[R_i(\mathbf{z}_j) R_{i'}(\mathbf{z}_{j'})] \\
&= N(N-1) \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_{i'}} E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i \frac{1}{A_i^2} E \left[\left\{ \int_{\mathbf{x} \in U_i} c(\mathbf{x}) d\mathbf{x} \right\}^2 \right] E \left[\{R_i(\mathbf{z}_j)\}^2 \right] \\
&= N(N-1) \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_{i'}} E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i \frac{1}{A_i^2} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_i} E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \tag{2-21} \\
&= N(N-1) \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_{i'}} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i \frac{1}{A_i^2} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_i} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x}
\end{aligned}$$

を得る．従って，式2-11，2-15，2-19，2-21を式2-12に代入して，

$$\begin{aligned}
V[\varepsilon^2] &= E[M^2] - 2E[MM\hat{M}] + E[\hat{M}^2] - (E[\varepsilon])^2 \\
&= N(N-1) \int_{\mathbf{t} \in S_0} \int_{\mathbf{x} \in S_0} f(\mathbf{x})f(\mathbf{t}) \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x} d\mathbf{t} \\
&\quad + N \int_{\mathbf{x} \in S_0} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} \\
&\quad - 2N(N-1) \int_{\mathbf{x} \in S_0} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} \Pr[\mathbf{x} \in S] d\mathbf{x} \\
&\quad - 2N \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} \int_{\mathbf{t} \in U_i} f(\mathbf{t}) \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\
&\quad + N(N-1) \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_{i'}} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} f(\mathbf{x}) d\mathbf{x} \tag{2-22} \\
&\quad + N \sum_i \frac{1}{A_i^2} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_i} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x} d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \\
&\quad - N^2 \left[\sum_i \left\{ \int_{\mathbf{x} \in U_i} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} - \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} \Pr[\mathbf{x} \in S] d\mathbf{x} \right\} \right]^2
\end{aligned}$$

が誤差の分散である．

式2-22は，空間集計システム，点オブジェクトが従う確率密度関数 $f(\mathbf{x})$ が規定され， $\Pr[\mathbf{x} \in S]$ と $\Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)]$ が任意の \mathbf{x} や \mathbf{t} について計算できれば，数値積分を用いて推定誤差 ε の分散を計算することができる，ということを表している．しかし残念ながら， $\Pr[\mathbf{x} \in S]$ や $\Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)]$ の計算はそれほど容易ではなく，様々な空間集計システムを比較検討することは難しい．そこで次節では，多少限定的な場合について推定誤差 ε の期待値や分散を導出し，空間集計システムの評価を行うためのより扱いやすい枠組みを構成する．

3. いくつかの場合における推定誤差の期待値と分散

本節では，前節で得られた結果を基に，いくつかの限定的な場合について推定誤差 ε の期待値と分散を導出する．

1) periodical continuum

まず第一に，periodical continuumと呼ばれる仮定（Stoyan and Stoyan, 1995）が成り立つ場合を考えてみよう．この仮定は， S_0 が長方形や正六角形，平行四辺形など，合同な図形を重ねることなく並べて平面を被覆できる図形するとき（図1）に利用可能であり，点パターンモデルなどの分析で良く用いられるものである．

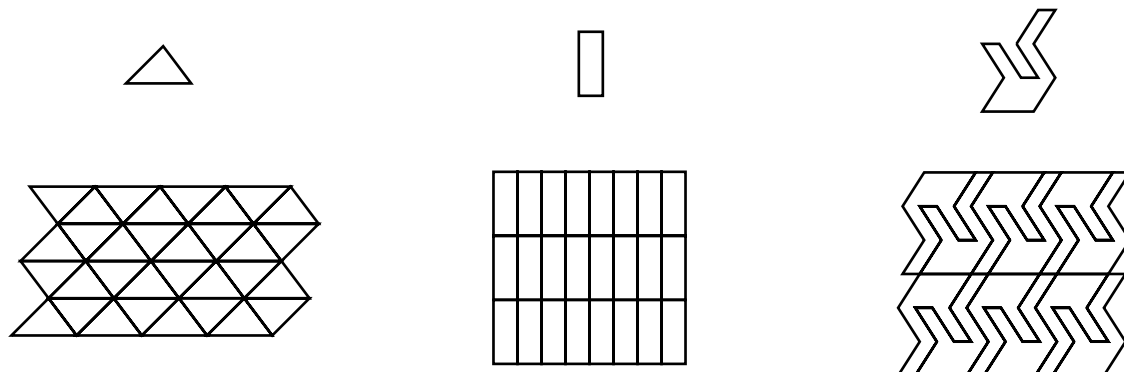


図1 領域 S_0 の例

S_0 がこのような図形するとき，periodical continuumでは以下の2つを仮定する．

- 1) S_0 の存在する平面は， S_0 の繰り返しパターンで完全に被覆されており， S_0 内の $f(\mathbf{x})$ も S_0 と同様に繰り返している．
- 2) S は，平面上に一様に分布する．

仮定1)は， S_0 が図1下段に示されるように並んでおり，各セルの内部では $f(\mathbf{x})$ が同じように分布している，ということの意味している．この仮定は非常に強いと思われるかもしれないが，分析対象領域 S_0 として長方形領域が用いられることは非常に多く，空間集計システムを評価する上ではそれほど問題にはならないと思われる．また仮定2)は， S の分布に偏りが無いということを表しており，これは，新たな空間集計単位の位置に関する情報が存在しない場合には，理由不十分の原則に従う妥当な仮定であろう．なお，periodical continuumに関する詳細な議論についてはRipley (1981)，Stoyan and Stoyan (1995)，Sadahiro (1999)を参照していただきたい．

これらの仮定の下では， $\Pr[\mathbf{x} \in S]$ は一定値となるので，

$$\Pr[\mathbf{x} \in S] = \frac{A}{A_0} \quad (3-1)$$

推定誤差 ε の期待値は以下のように表される．

$$\begin{aligned} E[\varepsilon] &= N \sum_i \left\{ \int_{\mathbf{x} \in U_i} f(\mathbf{x}) \Pr[\mathbf{x} \in S] d\mathbf{x} - \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} \Pr[\mathbf{x} \in S] d\mathbf{x} \right\} \\ &= N \sum_i \left\{ \int_{\mathbf{x} \in U_i} f(\mathbf{x}) \frac{A}{A_0} d\mathbf{x} - \frac{1}{A_i} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_i} \frac{A}{A_0} d\mathbf{x} \right\} \\ &= 0 \end{aligned} \quad (3-2)$$

既に前節でも述べたとおり， S が一様に分布する場合には， $f(\mathbf{x})$ が均一に分布していなくても ε の期待値は0となる．即ち，面積按分法による個数データの推定値が不偏推定量であるということを表しており，これは，面積按分法の推定値としての正当性を支持する重要な点の一つである．

次に推定誤差 ε の分散を求めておこう．

$$\begin{aligned}
V[\varepsilon^2] &= N(N-1) \int_{\mathbf{t} \in S_0} \int_{\mathbf{x} \in S_0} f(\mathbf{x})f(\mathbf{t}) \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x}d\mathbf{t} \\
&\quad + N \left(\frac{A}{A_0} \right) - 2N(N-1) \left(\frac{A}{A_0} \right)^2 \\
&\quad - 2N \sum_i \frac{1}{A_i} \int_{\mathbf{x} \in U_i} \int_{\mathbf{t} \in U_i} f(\mathbf{t}) \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t}d\mathbf{x} \\
&\quad + N(N-1) \sum_i \sum_{i'} \frac{1}{A_i A_{i'}} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_{i'}} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x}d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i \frac{1}{A_i^2} \int_{\mathbf{t} \in U_i} \int_{\mathbf{x} \in U_i} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{x}d\mathbf{t} \int_{\mathbf{x} \in U_i} f(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{3-3}$$

ここで， $\Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)]$ は領域 S の形状と， \mathbf{x} と \mathbf{t} の間の距離にのみ依存し， \mathbf{x} や \mathbf{t} の位置にはよらない．従って，距離 l だけ離れた2点が領域 S に含まれる確率を $g_S(l)$ とおくと，

$$\Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] = g_S(|\mathbf{x} - \mathbf{t}|) \tag{3-4}$$

と書くことができる．この関数は，領域 S がいくつかの典型的な形の場合には，解析的に与えられることが知られている（Santaló, 1976; Sadahiro, 1999）．例えば S が半径 r の円の場合には，

$$g_S(l) = \frac{1}{2\pi A_0} \begin{cases} \pi \left\{ 4r^2 \arccos\left(\frac{l}{2r}\right) - l\sqrt{4r^2 - l^2} \right\} & (l \leq 2r), \\ 0 & (l > 2r). \end{cases} \tag{3-5}$$

である．また， S が短辺 b ，長辺 c の長方形の場合には，

$$g_S(l) = \frac{1}{2\pi A_0} \begin{cases} 2\{ \pi bc - 2(b+c)l + l^2 \} & (l \leq b), \\ 2\{ 2c\sqrt{l^2 - b^2} - 2cl - b^2 + 2bc \arcsin \frac{b}{l} \} & (b < l \leq c), \\ 2\{ 2c\sqrt{l^2 - b^2} + 2b\sqrt{l^2 - c^2} - (b^2 + c^2 + l^2) + 2bc \left(\arcsin \frac{c}{l} - \arccos \frac{b}{l} \right) \} & (c < l \leq \sqrt{b^2 + c^2}), \\ 0 & (\sqrt{b^2 + c^2} < l). \end{cases} \tag{3-6}$$

となる．さらに， S が一辺 d の正方形の場合には，

$$g_S(l) = \frac{1}{2\pi A_0} \begin{cases} 2\{ \pi d^2 - 4dl + l^2 \} & (l \leq d), \\ 2\{ 4d\sqrt{l^2 - d^2} - (2d^2 + l^2) + 2d^2 \left(\arcsin \frac{d}{l} - \arccos \frac{d}{l} \right) \} & (d < l \leq \sqrt{2}d), \\ 0 & (\sqrt{2}d < l). \end{cases} \tag{3-7}$$

となる．従って，領域 S がこれらの典型的な形の場合には，式3-5～3-7と式3-3を用いることで，推定誤差 ε の分散を求めることが可能となる．

それでは，領域 S がこのような形ではない場合にはどのようにすれば良いであろうか．このときには，以下の式を用いることで関数 $g_S(l)$ を数値計算で求めることが可能である（証明はAppendix参照）．

$$g_S(l) = \frac{A}{2\pi A_0} \cdot \frac{1}{l} h_S(l) \tag{3-8}$$

ここで，関数 $h_S(l)$ は，領域 S 内でランダムに2点をとったときの2点間の距離の確率密度関数を表す．関数 $h_S(l)$ の数値計算は比較的容易である（例えばMonte-Carlo simulation）ので，推定誤差 ε の分散を数値的に求めることは可能である．

2) 1)に加えて $f(\mathbf{x})$ が一様分布である場合

次にperiodical continuumに加えて， $f(\mathbf{x})$ は S_0 内の一様分布であるという場合を考えてみよう．この仮定は即ち，

$$f(\mathbf{x}) = \frac{1}{A_0} \tag{3-9}$$

を意味し，推定誤差 ε の分散は以下ようになる．

$$\begin{aligned}
V[\varepsilon^2] &= \frac{2N(N-1)}{A_0^2} \int_{t \in S_0} \int_{x \in S_0} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dx dt \\
&\quad + N \left(\frac{A}{A_0} \right) - 2N(N-1) \left(\frac{A}{A_0} \right)^2 \\
&\quad - \frac{N}{A_0} \sum_i \frac{1}{A_i} \int_{x \in U_i} \int_{t \in U_i} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt dx
\end{aligned} \tag{3-10}$$

ここで,

$$\begin{aligned}
\int_{t \in S_0} \int_{x \in S_0} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dx dt &= \int_{t \in S_0} \int_{x \in S_0} E[c(\mathbf{x})c(\mathbf{t})] dx dt \\
&= E \left[\int_{t \in S_0} \int_{x \in S_0} c(\mathbf{x})c(\mathbf{t}) dx dt \right] \\
&= E \left[\int_{x \in S_0} c(\mathbf{x}) dx \int_{t \in S_0} c(\mathbf{t}) dt \right] \\
&= E \left[\left\{ \int_{x \in S_0} c(\mathbf{x}) dx \right\}^2 \right] \\
&= E[A^2] \\
&= A^2
\end{aligned} \tag{3-11}$$

であることに注意すれば, 式3-10は,

$$V[\varepsilon] = \frac{N}{A_0} \left\{ A - \sum_i \frac{1}{A_i} \int_{x \in U_i} \int_{t \in U_i} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt dx \right\} \tag{3-12}$$

となる.

式3-12は, 推定誤差 ε の分散が点の個数 N に比例し, 領域 S_0 の面積に反比例することを表している. 言い換えれば, 点の密度に比例するということになる. さらにこの式は, $V[\varepsilon]$ が各単位地区の形状に依存し, 単位地区同士の空間的位置関係にはよらない, ということを示している. 即ち, 図2のような空間集計システムは全て推定誤差 ε の分散が等しい.

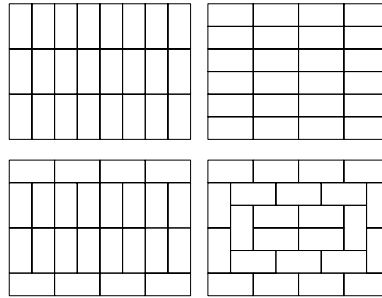


図2 空間的位置関係の異なる空間集計システム

式3-12の実際の計算には, 式3-3と類似した方法を用いることができる. 即ち, 距離 l だけ離れた2点が領域 S に含まれる確率 $g_S(l)$ と, 領域 U_i 内でランダムに2点をとったときの2点間の距離の確率密度関数 $h_i(l)$ を用いると,

$$\int_{x \in U_i} \int_{t \in U_i} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt dx = A_i^2 \int_{l=0}^{\infty} g_S(l) h_i(l) dl \tag{3-13}$$

となるので, 式3-12は

$$V[\varepsilon] = \frac{N}{A_0} \left\{ A - \sum_i A_i \int_{l=0}^{\infty} g_S(l) h_i(l) dl \right\} \tag{3-14}$$

と表すこともできる. あるいは, $g_S(l)$ の計算は難しいので, 式3-8を用いて

$$V[\varepsilon] = \frac{N}{A_0} \left\{ A - \frac{1}{2\pi A_0} \sum_i A_i^3 \int_{l=0}^{\infty} \frac{h_s(l)h_i(l)}{l} dl \right\} \quad (3-15)$$

としても良い。

3) 2)に加えて単位地区が全て合同な場合

メッシュデータのように各単位地区が全て合同なデータは数多い。そのような場合，式3-12は

$$V[\varepsilon] = \frac{N}{A_0} \left\{ A - \frac{A_0}{A_U^2} \int_{\mathbf{x} \in U} \int_{\mathbf{t} \in U} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt d\mathbf{x} \right\} \quad (3-16)$$

となる。但し， A_U は単位地区の面積， U は一つの単位地区領域をそれぞれ表す。あるいは，式3-15を元にして

$$V[\varepsilon] = \frac{NA}{A_0} \left\{ 1 - \frac{A}{2\pi} \int_{l=0}^{\infty} \frac{h_s(l)h_U(l)}{l} dl \right\} \quad (3-17)$$

としても良い。ここで $h_U(l)$ は，領域 U 内でランダムに2点をとったときの2点間の距離の確率密度関数である。この場合には， U と S の図形的性質が $V[\varepsilon]$ に大きな影響をもたらすことになる。

以上のように本節では，periodical continuumに基づいた3つの限定的な場合を想定し，推定誤差 ε の期待値と分散を導出した。これらの値はいずれも数値計算で求めることが可能であり，それによって様々な空間集計システムの評価を行うことができるようになる。

4. 点オブジェクトが一様分布に従う場合における空間集計システムの解析的評価

前節では，periodical continuumという条件の下での推定誤差 ε の期待値と分散を導出した。これらの値はいずれも数値計算で求めることが可能であり，空間集計システムの評価や比較には有用である。しかし，多数のシステムを相互比較する場合には，数値計算ではなく解析的な評価を行う方がより望ましい。そこで本節では，前節の2)の場合について，さらにいくつかの典型的な空間集計システムを取り上げ，推定誤差 ε の解析的評価を行う。

1) 領域 S_0 がいくつかの決まった形の単位地区で構成されている場合

まず，領域 S_0 がいくつかの決まった形の単位地区のみによって構成されている場合を考えてみよう。いま，領域 S_0 が，1, 2, ..., m という m 種類の形の単位地区から成り立っており，各単位地区の S_0 に占める面積的割合を $\alpha_1, \alpha_2, \dots, \alpha_m$ とする（ $\alpha_1 + \alpha_2 + \dots + \alpha_m = 1$ ）。すると，式3-12は

$$\begin{aligned} V[\varepsilon] &= \frac{N}{A_0} \left\{ A - \sum_j \frac{\alpha_j A_0}{A_j} \frac{1}{A_j} \int_{\mathbf{x} \in U_j} \int_{\mathbf{t} \in U_j} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt d\mathbf{x} \right\} \\ &= \frac{N}{A_0} \left\{ A - A_0 \sum_j \frac{\alpha_j}{A_j^2} \int_{\mathbf{x} \in U_j} \int_{\mathbf{t} \in U_j} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt d\mathbf{x} \right\} \end{aligned} \quad (4-1)$$

となる。あるいは式3-15を用いて，

$$\begin{aligned} V[\varepsilon] &= \frac{N}{A_0} \left\{ A - \frac{1}{2\pi A_0} \sum_j \frac{\alpha_j A_0}{A_j} A_j^3 \int_{l=0}^{\infty} \frac{h_s(l)h_j(l)}{l} dl \right\} \\ &= \frac{N}{A_0} \left\{ A - \frac{1}{2\pi} \sum_j \alpha_j A_j^2 \int_{l=0}^{\infty} \frac{h_s(l)h_j(l)}{l} dl \right\} \end{aligned} \quad (4-2)$$

である。

式4-1において，領域 S_0 が図形 j のみで構成されている場合を考えると，そのときの推定誤差 ε の分散 $V_j[\varepsilon]$ は

$$V_j[\varepsilon] = \frac{N}{A_0} \left\{ A - \frac{A_0}{A_j^2} \int_{\mathbf{x} \in U_j} \int_{\mathbf{t} \in U_j} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] dt d\mathbf{x} \right\} \quad (4-3)$$

である．この式を変形すると，

$$\int_{\mathbf{x} \in U_j} \int_{\mathbf{t} \in U_j} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} = \left(A - \frac{A_0 V_j[\varepsilon]}{N} \right) \frac{A_j^2}{A_0} \quad (4-4)$$

である．同様に全ての j について $V_j[\varepsilon]$ を導出し，式4-1を変形すると，

$$\begin{aligned} V[\varepsilon] &= \frac{N}{A_0} \left\{ A - \sum_j \frac{\alpha_j A_0}{A_j} \frac{1}{A_j} \int_{\mathbf{x} \in U_j} \int_{\mathbf{t} \in U_j} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \right\} \\ &= \frac{N}{A_0} \left\{ A - A_0 \sum_j \frac{\alpha_j}{A_j^2} \left(A - \frac{A_0 V_j[\varepsilon]}{N} \right) \frac{A_j^2}{A_0} \right\} \\ &= \frac{N}{A_0} \left\{ A - \sum_j \alpha_j \left(A - \frac{A_0 V_j[\varepsilon]}{N} \right) \right\} \\ &= \sum_j \alpha_j V_j[\varepsilon] \end{aligned} \quad (4-5)$$

を得る．即ち，領域 S_0 がいくつかの種類の単位地区から成り立っているときには， $V[\varepsilon]$ は各単位地区のみで構成されている場合の $V_j[\varepsilon]$ の線形結合で与えられ，それらの場合の中間的な値をとることが分かる．例えば，図3cのような空間集計システムの $V[\varepsilon]$ は，図3aと図3bの $V_j[\varepsilon]$ の面積比を乗じた線形結合となる．

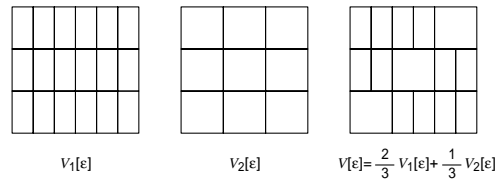


図3 いくつかの種類の単位地区で構成されている空間集計システム

2) 単位地区の結合が $V[\varepsilon]$ に与える影響について

次に，単位地区を結合することによって $V[\varepsilon]$ がどのように変化するかを見てみよう．これは例えば，県，市町村，町丁目というように，空間集計システムが階層構造をとっているとき，階層によって $V[\varepsilon]$ がどのように異なるのかということを考えることになる．

この問題については，2つの単位地区を1つに結合する場合だけを考えれば十分であるので，いま，単位地区 U_1 （面積 A_1 ）と U_2 （面積 A_2 ）を結合して単位地区 U とする場合を取り上げよう． $V[\varepsilon]$ を与える式3-12のうち，この結合によって変化する項だけを考えると，結合前は

$$v[\varepsilon] = -\frac{N}{A_0} \left\{ \frac{1}{A_1} \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} + \frac{1}{A_2} \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \right\} \quad (4-6)$$

であったのが，結合後は

$$v'[\varepsilon] = -\frac{N}{A_0} \cdot \frac{1}{A_1 + A_2} \int_{\mathbf{x} \in U_1 \cup U_2} \int_{\mathbf{t} \in U_1 \cup U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \quad (4-7)$$

となる．これらの差をとると，

$$\begin{aligned} \Delta V[\varepsilon] &= v'[\varepsilon] - v[\varepsilon] \\ &= -\frac{N}{A_0} \cdot \frac{1}{A_1 + A_2} \int_{\mathbf{x} \in U_1 \cup U_2} \int_{\mathbf{t} \in U_1 \cup U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &\quad + \frac{N}{A_0} \left\{ \frac{1}{A_1} \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} + \frac{1}{A_2} \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \right\} \end{aligned} \quad (4-8)$$

である．さらに計算を進めると，

$$\begin{aligned}
\Delta V[\varepsilon] &= \frac{N}{A_0} \left\{ \begin{aligned} &\frac{1}{A_1} \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &+ \frac{1}{A_2} \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &- \frac{1}{A_1 + A_2} \int_{\mathbf{x} \in U_1 \cup U_2} \int_{\mathbf{t} \in U_1 \cup U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \end{aligned} \right\} \\
&= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} \left\{ \begin{aligned} &A_2 (A_1 + A_2) \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &+ A_1 (A_1 + A_2) \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &- A_1 A_2 \int_{\mathbf{x} \in U_1 \cup U_2} \int_{\mathbf{t} \in U_1 \cup U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \end{aligned} \right\} \\
&= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} \left\{ \begin{aligned} &A_2 (A_1 + A_2) \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &+ A_1 (A_1 + A_2) \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &- A_1 A_2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &- A_1 A_2 \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &- 2A_1 A_2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \end{aligned} \right\} \tag{4-9}
\end{aligned}$$

$$\begin{aligned}
\Delta V[\varepsilon] &= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} \left\{ \begin{aligned} &A_2^2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &+ A_1^2 \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \\ &- 2A_1 A_2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_2} \Pr[(\mathbf{x} \in S) \cap (\mathbf{t} \in S)] d\mathbf{t} d\mathbf{x} \end{aligned} \right\} \\
&= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} \left\{ \begin{aligned} &A_2^2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{t} d\mathbf{x} \\ &+ A_1^2 \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{t} d\mathbf{x} \\ &- 2A_1 A_2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_2} E[c(\mathbf{x})c(\mathbf{t})] d\mathbf{t} d\mathbf{x} \end{aligned} \right\} \\
&= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} \left\{ \begin{aligned} &E \left[A_2^2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_1} c(\mathbf{x})c(\mathbf{t}) d\mathbf{t} d\mathbf{x} \right] \\ &+ E \left[A_1^2 \int_{\mathbf{x} \in U_2} \int_{\mathbf{t} \in U_2} c(\mathbf{x})c(\mathbf{t}) d\mathbf{t} d\mathbf{x} \right] \\ &- 2E \left[A_1 A_2 \int_{\mathbf{x} \in U_1} \int_{\mathbf{t} \in U_2} c(\mathbf{x})c(\mathbf{t}) d\mathbf{t} d\mathbf{x} \right] \end{aligned} \right\} \tag{4-10}
\end{aligned}$$

$$\begin{aligned}
\Delta V[\varepsilon] &= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} E \left[\begin{array}{l} A_2^2 \int_{\mathbf{x} \in U_1} \int_{t \in U_1} c(\mathbf{x}) c(t) dt d\mathbf{x} \\ + A_1^2 \int_{\mathbf{x} \in U_2} \int_{t \in U_2} c(\mathbf{x}) c(t) dt d\mathbf{x} \\ - 2A_1 A_2 \int_{\mathbf{x} \in U_1} \int_{t \in U_2} c(\mathbf{x}) c(t) dt d\mathbf{x} \end{array} \right] \\
&= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} E \left[\left\{ A_2 \int_{\mathbf{x} \in U_1} c(\mathbf{x}) d\mathbf{x} \right\}^2 + \left\{ A_1 \int_{\mathbf{x} \in U_2} c(\mathbf{x}) d\mathbf{x} \right\}^2 - 2A_1 A_2 \int_{\mathbf{x} \in U_1} c(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in U_2} c(\mathbf{x}) d\mathbf{x} \right] \quad (4-11) \\
&= \frac{N}{A_0 A_1 A_2 (A_1 + A_2)} E \left[\left\{ A_2 \int_{\mathbf{x} \in U_1} c(\mathbf{x}) d\mathbf{x} - A_1 \int_{\mathbf{x} \in U_2} c(\mathbf{x}) d\mathbf{x} \right\}^2 \right] \\
&\geq 0
\end{aligned}$$

となる。従って、階層構造を持つ空間集計システムについては、単位地区の小さいシステムを用いれば、推定誤差 ε の分散を小さくすることができる。このことは、単位地区や新しい空間集計単位の絶対的な大きさや形状にはよらない普遍的な結論である。但し、ここでは新しい空間集計単位の位置や点オブジェクトの分布について様々な場合を確率的に扱っているため、単位地区の小さいシステムを用いれば常に ε が小さくなるというわけではないということに注意していただきたい。

3) 単位地区が全て合同であるときに単位地区の大きさが $V[\varepsilon]$ に与える影響について

次に、単位地区が全て合同であるときに、単位地区の大きさによって $V[\varepsilon]$ にどのように変化するかを見てみよう。これは例えば、図4のような空間集計システムを比較するということである。

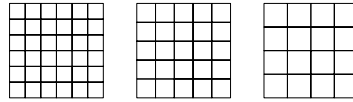


図4 単位地区の大きさが異なる空間集計システム

ここでも前節と同じように、単位地区の大きさだけが異なる2つの空間集計システムを比較しよう。いま、単位地区の大きさがそれぞれ A_1 と A_2 ($A_1 < A_2$) の空間集計システムを考えると、それぞれの $V[\varepsilon]$ は式3-14より、

$$V_1[\varepsilon] = N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S(l) h_1(l) dl \right\} \quad (4-12)$$

$$V_2[\varepsilon] = N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S(l) h_2(l) dl \right\} \quad (4-13)$$

となる。但し、 $h_1(l)$ 及び $h_2(l)$ はそれぞれ単位地区1と単位地区2内に2点をランダムにとったときの点間の距離 l の確率密度関数である。ここで、単位地区1と単位地区2は相似形であるので、

$$h_2(l) = \sqrt{\frac{A_1}{A_2}} h_1\left(\sqrt{\frac{A_1}{A_2}} l\right) \quad (4-14)$$

が成り立つ。式4-14を式4-13に代入すると、

$$\begin{aligned}
V_2[\varepsilon] &= N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S(l) \sqrt{\frac{A_1}{A_2}} h_1\left(\sqrt{\frac{A_1}{A_2}} l\right) dl \right\} \\
&= N \left\{ \frac{A}{A_0} - \sqrt{\frac{A_1}{A_2}} \int_{l=0}^{\infty} g_S(l) h_1\left(\sqrt{\frac{A_1}{A_2}} l\right) dl \right\} \quad (4-15)
\end{aligned}$$

となる。さらに、

$$l' = \sqrt{\frac{A_1}{A_2}} l \quad (4-16)$$

とおけば，

$$\frac{dl}{dl} = \sqrt{\frac{A_1}{A_2}} \quad (4-17)$$

より，式4-15は

$$\begin{aligned} V_2[\varepsilon] &= N \left\{ \frac{A}{A_0} - \sqrt{\frac{A_1}{A_2}} \int_{l=0}^{\infty} g_S \left(\sqrt{\frac{A_2}{A_1}} l \right) h_1(l) \sqrt{\frac{A_2}{A_1}} dl \right\} \\ &= N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S \left(\sqrt{\frac{A_2}{A_1}} l \right) h_1(l) dl \right\} \\ &= N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S \left(\sqrt{\frac{A_2}{A_1}} l \right) h_1(l) dl \right\} \end{aligned} \quad (4-18)$$

となる．従って， $V_1[\varepsilon]$ と $V_2[\varepsilon]$ の差をとると，

$$\begin{aligned} V_2[\varepsilon] - V_1[\varepsilon] &= N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S \left(\sqrt{\frac{A_2}{A_1}} l \right) h_1(l) dl \right\} - N \left\{ \frac{A}{A_0} - \int_{l=0}^{\infty} g_S(l) h_1(l) dl \right\} \\ &= N \int_{l=0}^{\infty} \left\{ g_S \left(\sqrt{\frac{A_2}{A_1}} l \right) - g_S(l) \right\} h_1(l) dl \end{aligned} \quad (4-19)$$

となる．

ここで， $\sqrt{A_2/A_1} > 1$ かつ $h_1(l) \geq 0$ であるので，もし $g_S(l)$ が単調減少関数であれば，単位地区や S の形状によらず，単位地区の大きさが大きくなると $V[\varepsilon]$ は必ず大きくなる．この条件が成立する場合を考えてみると， S がconvexの場合がそれに相当する．なぜならば S がconvexの場合， $g_S(l)$ は長さ l の線分を含む図形 S と合同な図形の集合の測度であり，これは，図形 S に内包される長さ l の線分の集合の測度と等しい．そこで，後者について考えると，長さ $l' (> l)$ の線分のとり得る位置は常に長さ l の線分のとりうる位置を含んでいる．従って，図形 S に内包される長さ l の線分の集合の測度は，図形 S に内包される長さ l' の線分の集合の測度よりも大きく， $l' > l$ であれば $g_S(l) > g_S(l')$ と言える．これは $g_S(l)$ が単調減少関数であることを示している．この結論は， S がconvexの場合，合同な図形で構成される空間集計システムの中では，できるだけ単位地区の大きさが小さいものを選んだ方が良い，ということを表す．

但し残念ながら， S がconvexではない場合について，単位地区の大きさと $V[\varepsilon]$ の関係を解析的に明らかにすることはできない．いくつかの数値計算では，単位地区を大きくしたときに $V[\varepsilon]$ が小さくなるような場合は見られず，「単位地区が大きくなると $V[\varepsilon]$ が大きくなる」ということは広い範囲で成立するものと予想される．この点については，今後更に研究を進める必要がある．

5. 数値計算

本節では前節までの分析結果を利用し，様々な空間集計システムと新しい空間集計単位について，推定誤差 ε の分散を数値的に計算し，それら相互間の関係を明らかにする．

まず，最も代表的な空間集計システムである，正方形格子網について考えてみよう．この場合， $g_U(l)$ 及び $h_U(l)$ は解析的な形で与えられる（式3-5～3-7）．一方，領域 S としては円，正方形，長方形を取り上げる．これらの形についても，いずれも $g_S(l)$ 及び $h_S(l)$ が解析的な形で与えられるので， $V[\varepsilon]$ は式3-17を用いて数値積分により求めることができる．但し，

$$V[\varepsilon] = \frac{NA}{A_0} \left\{ 1 - \frac{A}{2\pi} \int_{l=0}^{\infty} \frac{h_S(l)h_U(l)}{l} dl \right\} \quad (5-1)$$

のうち，括弧の外は議論する意味がないので，ここでは

$$v[\varepsilon] = 1 - \frac{A}{2\pi} \int_{l=0}^{\infty} \frac{h_S(l)h_U(l)}{l} dl \quad (5-2)$$

として $v[\varepsilon]$ の値を求める．分析は，領域 S の面積を一定（1とする）とし，正方形格子の大きさを様々に変えて行う．また，領域 S の長方形としては，縦横比が1:2，1:3，1:4，1:8，1:16，1:32の6通りの場合を取り

上げる。

以下，図5に分析結果を示す。

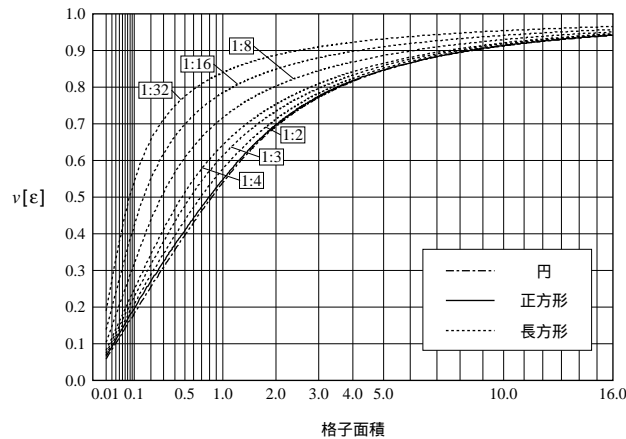


図5 正方形格子網における格子の大きさと $v[\epsilon]$ の関係

ここで取り上げている領域 S は全て凸多角形である。従って前節で述べたとおり，格子の大きさが大きくなると $v[\epsilon]$ は単調に増加することになり，図5でこれを再確認できる。 $v[\epsilon]$ の増加は，格子の大きさが小さい時には急激であるが，格子がある程度以上大きくなると， $v[\epsilon]$ の増加は緩やかになる。即ち，格子が S に対してある程度以上の大きさである場合には，格子の大きさは $v[\epsilon]$ にあまり影響を与えないということである。

一方，領域 S の形状による影響を見てみると，円，正方形，長方形の順に良い結果をもたらしており，長方形の中では縦横比が1:1に近いものほど $v[\epsilon]$ が小さくなっている。このことは，領域 S が円に近い図形であるほど推定誤差が小さくなる，あるいは，格子が大きくても推定誤差を小さく押さえることができる，ということの意味している。例えば， $v[\epsilon]=0.5$ としてみると，領域 S が正方形の場合には格子面積は0.77であるが，縦横比1:4の長方形では0.46，縦横比1:16の長方形では0.16となる。つまり，同じ推定精度を得るためには，縦横比1:16の長方形では正方形の場合の約1/4の大きさの格子で構成される格子網データが必要となり，4倍のデータ量を要するということであり，

次に，格子網が正三角形，正六角形で構成されている場合を見てみよう（図6，図7）。いずれも，正方形格子網の場合と非常に類似した傾向を示していることがわかる。そこで，円，縦横比が1:4の長方形，縦横比が1:16の長方形の3つの場合について，これらを比較してみよう（図8）。すると，正方形格子網と正六角形格子網のグラフはほとんど重なってしまい，正三角形格子網のグラフはこれら2つよりも常に上方にあることが分かる。つまり，正方形格子網と正六角形格子網では推定誤差の分散にほとんど差がなく，正三角形格子網の推定誤差の分散はこれらよりも多少大きくなる。さらに詳しく調べて見ると，正六角形格子網における $v[\epsilon]$ の値は常に正方形格子網の値よりも小さい。従って，推定誤差の分散という観点から見ると，正六角形格子網，正方形格子網，正三角形格子網の順に望ましいと言えよう。但し，これらの差はそれほど大きなものではなく，一般的に良く用いられている正方形格子網は非常によい性質を持っていると言って良いであろう。

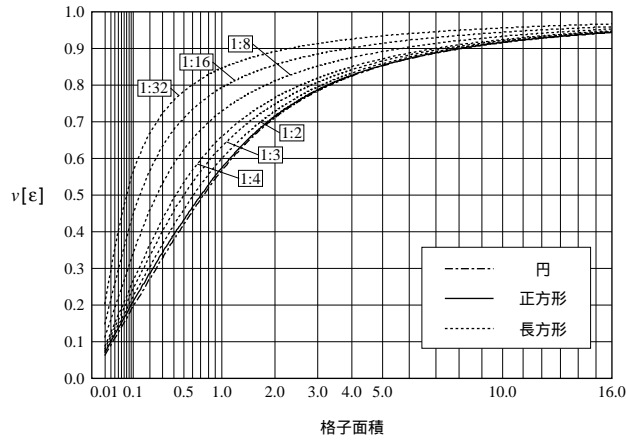


図6 正三角形格子網における格子の大きさと $v[\epsilon]$ の関係

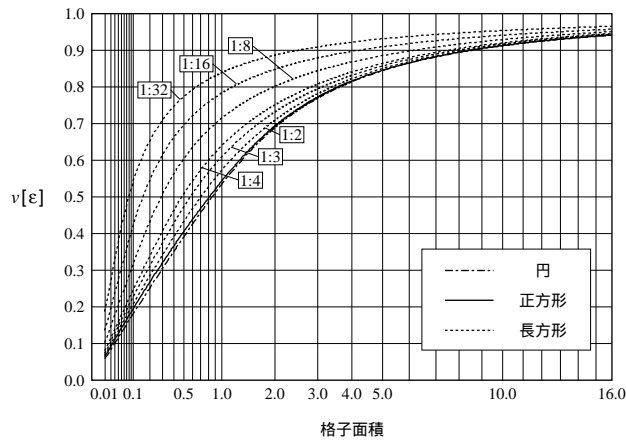


図7 正六角形格子網における格子の大きさと $v[\epsilon]$ の関係

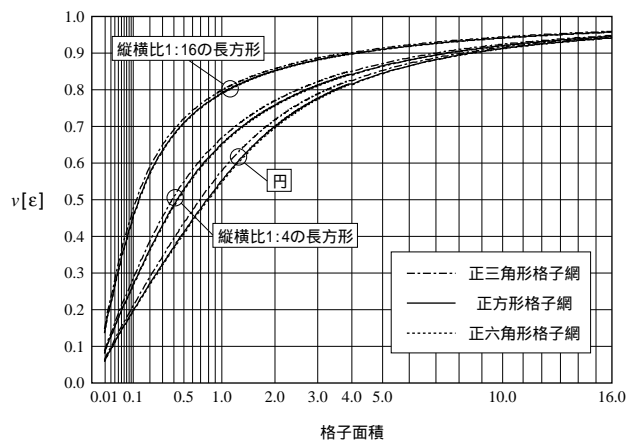


図8 3つの格子網における格子の大きさと $v[\epsilon]$ の関係

最後に、長方形格子網の場合についてみてみよう。実際には、長方形の格子網はあまり使われないが、格子図形の形状と推定誤差の関係を知る上で重要な手がかりを与えると思われる。また、空間集計システムのうち一部が長方形で構成されているものを考えるときには、前節で述べたとおり全て合同図形の場合について $v[\epsilon]$ を計算し、それらの線形和が $V[\epsilon]$ を与える。

以下、紙面の都合上、縦横比が1:2, 1:4, 1:16の長方形格子網について、格子の大きさと $v[\varepsilon]$ の関係を図9~11に示す。いずれも全体的な傾向は正多角形の格子網と類似しており、格子が大きくなると $v[\varepsilon]$ は最初は急激に、やがて緩やかに増加していく。

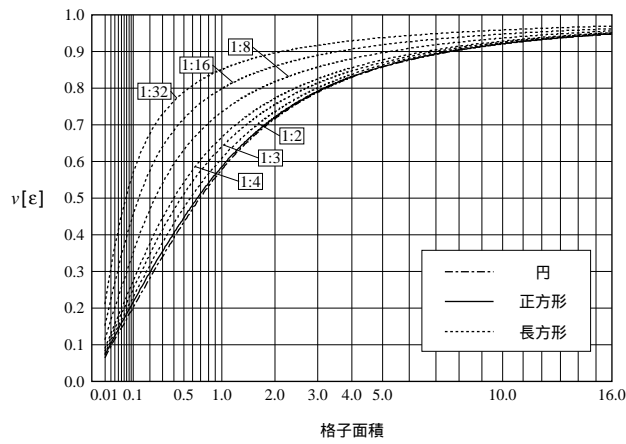


図9 縦横比1:2の長方形格子網における格子の大きさと $v[\varepsilon]$ の関係

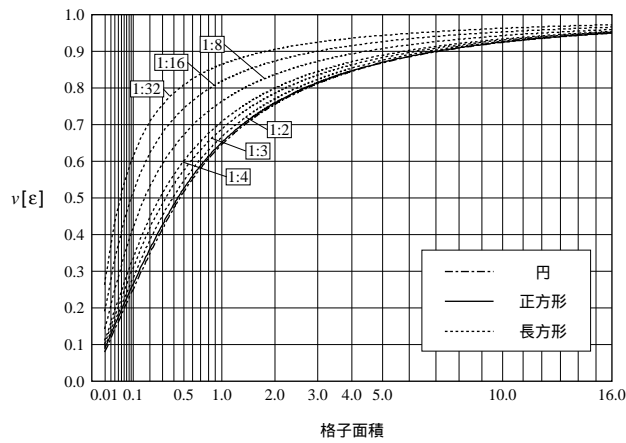


図10 縦横比1:4の長方形格子網における格子の大きさと $v[\varepsilon]$ の関係

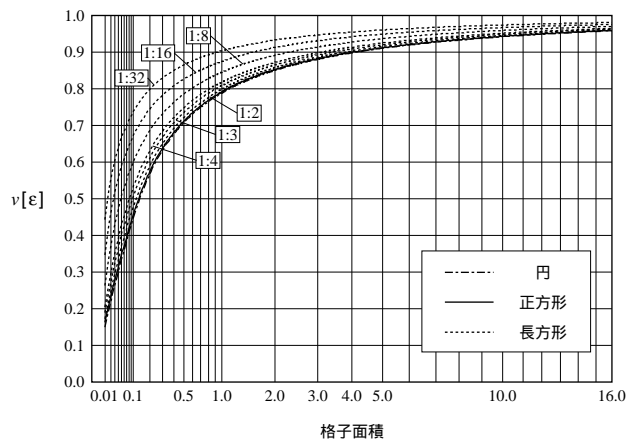


図11 縦横比1:16の長方形格子網における格子の大きさと $v[\varepsilon]$ の関係

次に、格子網を構成する長方形の縦横比が異なる場合について相互比較してみよう。ここでは、領域 S

が円，縦横比1:4の長方形，縦横比1:16の長方形の3つの場合について，図12～14に結果を掲載する．これらの結果を見ると，領域 S が円の場合には，格子図形の形状によって $v[\varepsilon]$ がある程度影響を受けるが，領域 S が縦横比1:16の長方形の場合では，格子図形の形状はあまり大きな影響力を持たないことが分かる．また，格子図形の形状によらず，領域 S が円の場合には格子の大きさによる $v[\varepsilon]$ の変化が緩やかであるのに対し， S が縦横比1:16の長方形の場合には格子の大きさによる $v[\varepsilon]$ の変化が急激である．即ち， S が縦横比1:16の長方形の場合には，格子の大きさの選定には十分注意を払う必要があると言える．特に，高い精度を得ようとする場合には，かなり小さな格子による空間集計システムが必要である．

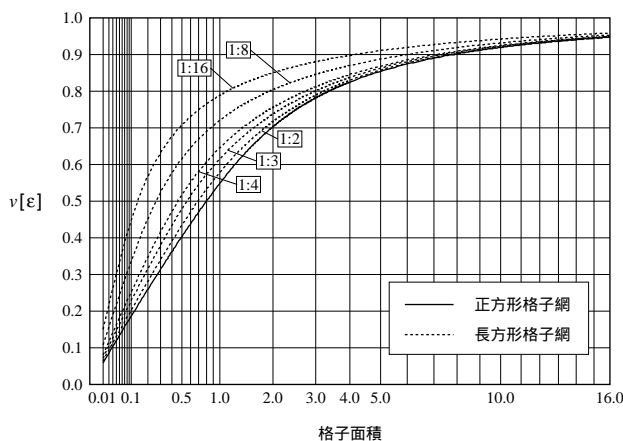


図12 縦横比が異なる長方形格子網における格子の大きさと $v[\varepsilon]$ の関係（領域 S が円の場合）

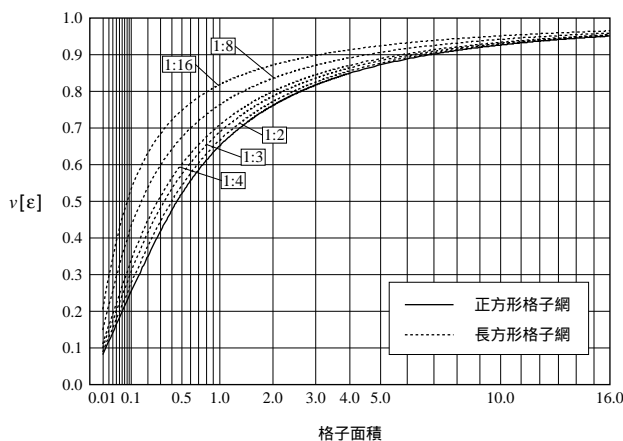


図13 縦横比が異なる長方形格子網における格子の大きさと $v[\varepsilon]$ の関係（領域 S が縦横比1:4の長方形の場合）

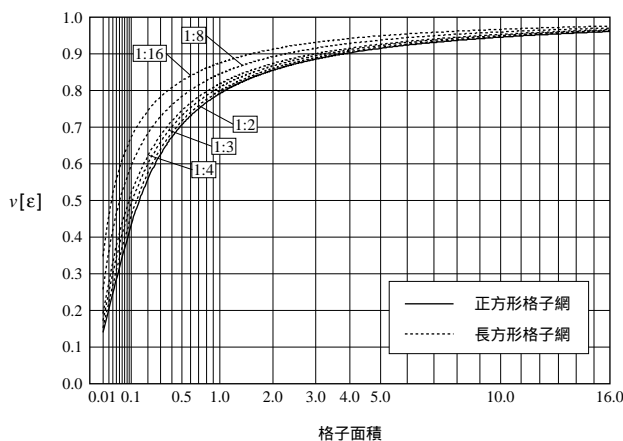


図14 縦横比が異なる長方形格子網における格子の大きさと $v[\varepsilon]$ の関係（領域 S が縦横比1:16の長方形の場合）

ところで、以上の数値計算の結果から、領域 S 、及び、格子図形の形状が円に近いほど $v[\varepsilon]$ は小さくなり、細長い長方形になるほど $v[\varepsilon]$ が大きくなるという傾向が窺える。このことについては、理論的な裏付けは困難であるが、おそらく図形の周長と密接な関係があるものと思われる。いま、 S と単位地区の位置関係について考えてみると、 S が複数の単位地区にまたがっている方が、単一の単位地区に含まれているよりも $v[\varepsilon]$ は大きくなる。例えば、 S が二つの単位地区にまたがっている場合だけに限定すれば、 S がどちらか一方のみに含まれる場合に $v[\varepsilon]$ が最小になる。同様に、三つ以上の単位地区にまたがっているについても、領域 S のまたがる単位地区が少ないほど $v[\varepsilon]$ は小さくなり、反対に、多くの単位地区にまたがるほど、 $v[\varepsilon]$ は大きくなるものと推測される。

一方、周長 L 、面積 A の領域 S が単位地区面積 A_U 、周長 L_U の格子網にランダムに落とされるとき、 S がまたがる単位地区の数 κ の期待値は、

$$E[\kappa] = \frac{2\pi(A_U + A) + L_U L}{2\pi A_U} \quad (5-3)$$

で与えられる。即ち、面積 A 及び A_U 固定して考えると、 L 及び L_U が大きいほど κ の期待値は大きくなる。つまり、 L 及び L_U が大きいほどより多くの単位地区にまたがるということであり、その結果、 $v[\varepsilon]$ が大きくなるものと推測される。

おわりに

以上のように本論文では、ある空間集計システムで記録されている個数データを別の空間集計単位に変換する際に生ずる変換誤差を、空間集計システムや空間集計単位の形状と大きさ、空間オブジェクトの分布との関係から確率的に記述し、その性質を明らかにした。得られた式の中で、式2-11及び式2-22はいずれも非常に広範な適用範囲を持つものであり、様々な場合について変換誤差の評価が可能になる。

得られた解析結果を用いた数値計算では、その結果、空間集計単位の大きさについては小さいほど、またその形状については円に近い図形であるほど、変換誤差が小さくなることが分かった。前者は経験的に我々が知っていたことであるが、後者について大きさに匹敵するほどの影響が見られたことは、非常に有用な知見であろう。図形の形状の影響については、おそらく図形の周長と密接な関係があるものと推測され、今後行うべき研究の一つの方向性を与えている。

参考文献

- Burrough, P. A. and McDonnell, R. A. (1998): *Principles of Geographical Information Systems*. New York: Oxford University Press.
- Fisher, P. F. and Langford, M. (1995): Modelling the Errors in Areal Interpolation between Zonal Systems by Monte Carlo Simulation. *Environment and Planning A*, **27**, 211-224.
- Flowerdew, R. (1988): Statistical Methods for Areal Interpolation: Predicting Count Data from a Binary Variable. *Research Report 15*, Northern Regional Research Laboratory.
- Flowerdew, R. and Green, M. (1991): Data Integration: Statistical Methods for Transferring Data between Zonal Systems. In *Handling Geographical Information: Methodology and Potential Applications*, edited by I. Masser and M. Blakemore (New York: Longman), 38-54.
- Goodchild, M. F. and Lam, N. N-S. (1989): Areal Interpolation: a Variant of the Traditional Spatial Problem. *Geo-processing*, **1**, 297-312.
- Goodchild, M. F., Anselin, L., and Deichmann, U. (1993): A Framework for the Areal Interpolation of Socioeconomic Data. *Environment and Planning A*, **25**, 383-397.
- Ghosh, B. (1951): Random Distances within a Rectangle and between Two Rectangles. *Bulletin of the Calcutta Mathematics Society*, **43**, 17-24.
- Lam, N. N-S. (1983): Spatial Interpolation Methods: a Review. *American Cartographer*, **10**, 129-149.

Ripley, B. D. (1981) *Spatial Statistics*. John Wiley, New York.

Sadahiro, Y. (1999): Statistical Methods for Analyzing the Distribution of Spatial Objects in Relation to a Surface. *Geographical Systems*, to appear.

Santaló, L. A. (1976) *Integral Geometry and Geometric Probability*. Addison-Wesley, London.

Stoyan, D. and Stoyan, H. (1994): *Fractals, Random Shapes and Point Fields*. John Wiley, New York.

Tobler, W. R. (1979): Smooth Pycnophylactic Interpolation for Geographical Regions. *Journal of the American Statistical Association*, **74**, 519-530.

Appendix 関数 $g_S(l)$ と関数 $h_S(l)$ の関係

ここでは、距離 l だけ離れている2点が領域 S に内包される確率 $g_S(l)$ と、領域 S 内にランダムに2点をとった時の2点間の距離の確率密度関数 $h_S(l)$ の関係を考える。

まず、以下のような関数を考える。

$$C(\mathbf{x}, \mathbf{t}, l) = \begin{cases} 1 & \text{if } |\mathbf{x} - \mathbf{t}| \leq l \\ 0 & \text{else} \end{cases} \quad (\text{A-1})$$

そして、この関数を用いて $h_S(l)$ の確率分布関数 $H_S(l)$ を表すと、

$$H_S(l) = \frac{1}{A^2} \int_{\mathbf{t} \in S} \int_{\mathbf{x} \in S} C(\mathbf{x}, \mathbf{t}, l) d\mathbf{x} d\mathbf{t} \quad (\text{A-2})$$

であるので、 $h_S(l)$ は

$$\begin{aligned} h_S(l) &= \lim_{dl \rightarrow 0} \left\{ \frac{H_S(l+dl) - H_S(l)}{dl} \right\} \\ &= \frac{1}{A^2} \int_{\mathbf{t} \in S} \int_{\mathbf{x} \in S} \lim_{dl \rightarrow 0} \frac{C(\mathbf{x}, \mathbf{t}, l+dl) - C(\mathbf{x}, \mathbf{t}, l)}{dl} d\mathbf{x} d\mathbf{t} \end{aligned} \quad (\text{A-3})$$

となる。

次に、関数 $g_S(l)$ を考えてみよう。そのためにまず、距離 l だけ離れている2点を内包する領域 S の測度 $m_S(l)$ を用いて

$$g_S(l) = \frac{m_S(l)}{2\pi A_0} \quad (\text{A-4})$$

とし、測度 $m_S(l)$ の導出を行う。測度 $m_S(l)$ は、領域 S に内包される距離 l だけ離れた2点の測度と等しい。そこでここでは、距離が $l \sim l+dl$ だけ離れた2点对の測度を求める。まず、 $\mathbf{x}_1, \mathbf{x}_2$ のうちまず \mathbf{x}_1 を固定して考えてみよう。すると、 \mathbf{x}_2 の取りうる領域は、半径 $l \sim l+dl$ の間の円環領域のうち S と重なる部分である(図A-1)。この部分の面積 s_c は、関数 $C(\mathbf{x}, \mathbf{t}, l)$ を用いて

$$s_c = \int_{\mathbf{t} \in S} \{C(\mathbf{x}, \mathbf{t}, l+dl) - C(\mathbf{x}, \mathbf{t}, l)\} d\mathbf{t} \quad (\text{A-5})$$

と表される。面積 s_c はまた、部分円環領域の中心角 $\theta(\mathbf{x}_1)$ (図A-1、複数部分に分かれている場合には、各領域の中心角の合計をとる)を用いて次のようにも表される。

$$\begin{aligned} s_c &= \frac{\theta(\mathbf{x}_1)}{2\pi} \{ \pi(l+dl)^2 - \pi l^2 \} \\ &= \frac{\theta(\mathbf{x}_1)}{2} \{ 2ldl - dl^2 \} \\ &\approx \theta(\mathbf{x}_1) l dl \end{aligned} \quad (\text{A-6})$$

式A-5, A-6より、

$$\theta(\mathbf{x}_1) = \frac{\int_{\mathbf{t} \in U_i} \{C(\mathbf{x}, \mathbf{t}, l+dl) - C(\mathbf{x}, \mathbf{t}, l)\} d\mathbf{t}}{ldl} \quad (\text{A-7})$$

を得る。さらに、

$$\lim_{dl \rightarrow 0} \theta(\mathbf{x}_1) = \lim_{dl \rightarrow 0} \frac{\int_{\mathbf{t} \in U_i} \{C(\mathbf{x}, \mathbf{t}, l+dl) - C(\mathbf{x}, \mathbf{t}, l)\} d\mathbf{t}}{ldl} \quad (\text{A-8})$$

を考えると、これは即ち、一端を \mathbf{x}_1 とし、距離が l だけ離れた点 \mathbf{x}_2 の測度を表す。従って測度 $m_S(l)$ は、式

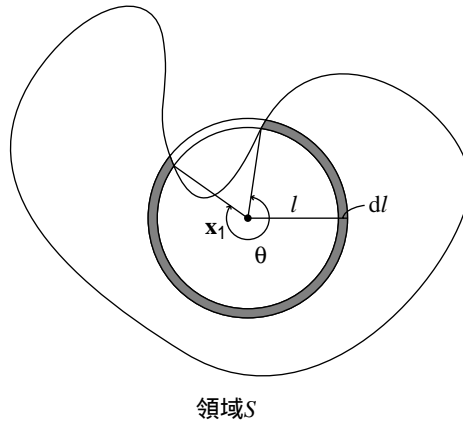
A-8を \mathbf{x}_1 について積分した

$$\begin{aligned}
 m_S(l) &= \int_{\mathbf{x}_1 \in S} \lim_{dl \rightarrow 0} \theta(\mathbf{x}_1) d\mathbf{x}_1 \\
 &= \lim_{dl \rightarrow 0} \int_{\mathbf{x}_1 \in S} \frac{1}{dl} \int_{\mathbf{x}_2 \in S} \{C(\mathbf{x}_1, \mathbf{x}_2, l+dl) - C(\mathbf{x}_1, \mathbf{x}_2, l)\} d\mathbf{x}_2 d\mathbf{x}_1 \\
 &= \frac{1}{l} \int_{\mathbf{x}_1 \in S} \int_{\mathbf{x}_2 \in S} \lim_{dl \rightarrow 0} \frac{C(\mathbf{x}_1, \mathbf{x}_2, l+dl) - C(\mathbf{x}_1, \mathbf{x}_2, l)}{dl} d\mathbf{x}_2 d\mathbf{x}_1 \\
 &= \frac{A^2}{l} h_S(l)
 \end{aligned} \tag{A-9}$$

となり，式A-4を用いて，

$$\begin{aligned}
 g_S(l) &= \frac{m_S(l)}{2\pi A_0} \\
 &= \frac{A^2}{2\pi A_0} \cdot \frac{1}{l} h_S(l)
 \end{aligned} \tag{A-10}$$

を得る．



図A-1 \mathbf{x}_2 の取りうる領域（灰色の部分）