

Discussion Paper No. 77

代表点内包法を用いて推定される個数データの誤差

貞広幸雄*

1998年7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

GISでは、空間データがある空間単位で集計されていることが多く、空間解析の段階でそれらのデータを別の空間単位に変換する必要がしばしば生ずる。このような変換は面補間 (areal interpolation) と呼ばれており (Goodchild and Lam, 1980; Lam, 1983; Flowerdew and Green, 1991)、面積按分法 (Fisher and Langford, 1995)、代表点内包法 (Burrough and McDonnell, 1998)、pynophylactic法 (Tobler, 1979) などといった比較的単純な方法から、RSデータに代表されるような補助的データを用いる方法 (Flowerdew, 1988; Goodchild et al., 1993) まで、様々な方法が提案されてきている。

しかしながら、面補間によって得られる値はあくまで推定値であり、通常は必ず何らかの誤差を含んでいる。そこで本論文では、最も代表的な面補間の方法である代表点内包法において発生する個数データの誤差について、点オブジェクトや代表点の分布、集計単位の大きさなどとの関係を分析する。この種の研究は従来、実証分析の積み重ねや、モンテカルロシミュレーションを用いた数値シミュレーションによって行われることが多かったが (Fisher and Langford, 1995)、ここでは点オブジェクトと集計単位地区について確率分布モデルを導入することにより、そのような計算費用の高い方法を回避する。

以下、2節では確率過程に基づいた面補間モデルと、個数データの推定誤差に関する評価指標を提案する。3節では、このモデルを用いて集計単位の大きさや点オブジェクトの分布形、代表点の位置などが推定誤差に与える影響を数値的に分析する。4節では結論と今後の課題を述べる。

2. 面補間モデル

代表点内包法を用いた個数データの推定誤差は、変換元の空間集計システム、新たな空間集計地区、空間オブジェクト分布、代表点の位置の4つに依存する。本節では、これらの関係を定量的に記述するために、確率過程に基づいた面補間モデルによる定式化を行う。

いま、面積 A_0 を持つ領域 Z_0 の内部がいくつかの単位地区に区切られているものとし、単位地区 Z_i の面積を A_i 、代表点の位置ベクトルを z_i とおく。この空間集計システムを S と表記する。領域 Z_0 内には、 N 個の点オブジェクトがそれぞれ独立に確率密度関数 $f(x)$ に従って分布し、単位地区 Z_i に含まれる点オブジェクトの個数 n_i は代表点 z_i に全て割り当てられる。そして、 S とは一致しない、ある空間集計地区 T (面積を B とする) が、領域 S_0 に重な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 (図1)。

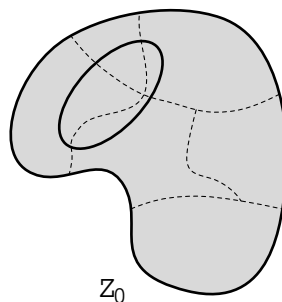


図1 空間集計システムと空間集計地区

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 y_j とする。また、関数 $C(x)$ を以下のように定義する。

$$C(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

すると、 T に含まれる点オブジェクトの真の個数 M は、

$$M = \sum_j C(\mathbf{y}_j) \quad (2)$$

と表される．さらに，関数 $U_i(\mathbf{x})$ を

$$U_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in Z_i \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

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，

$$\hat{M} = \sum_i \left\{ C(\mathbf{z}_i) \sum_j U_i(\mathbf{y}_j) \right\} \quad (4)$$

となる．従って，点オブジェクト個数の推定誤差 ε は，

$$\begin{aligned} \varepsilon &= M - \hat{M} \\ &= \sum_j C(\mathbf{y}_j) - \sum_i \left\{ C(\mathbf{z}_i) \sum_j U_i(\mathbf{y}_j) \right\} \end{aligned} \quad (5)$$

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地区 T 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．そこでここでは，推定誤差 ε の平均自乗誤差Mean Square Error (MSE)を誤差の評価値として採用する．

$MSE[S]$ の導出方法は以下の通りである．まず準備のため，関数 $C(\mathbf{y}_j)$ と $U_i(\mathbf{y}_j)$ の期待値を求めておこう．関数 $C(\mathbf{y}_j)$ の期待値 $E[C(\mathbf{y}_j)]$ は点 j が T に含まれる確率を表しており，これは微小領域 $d\mathbf{x}$ が T 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 Z_0 内で積分することにより与えられる．即ち，

$$\begin{aligned} E[C(\mathbf{y}_j)] &= \int_{\mathbf{x} \in Z_0} \Pr[(\mathbf{y}_j \in d\mathbf{x}) \cap (d\mathbf{x} \in T)] d\mathbf{x} \\ &= \int_{\mathbf{x} \in Z_0} \Pr[\mathbf{y}_j \in d\mathbf{x}] \Pr[d\mathbf{x} \in T] \end{aligned} \quad (6)$$

となる．ここで， $d\mathbf{x}$ は微小領域であるので，

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

とおくことができる．この式を式6に代入すると，

$$\begin{aligned} E[C(\mathbf{y}_j)] &= \int_{\mathbf{x} \in Z_0} \Pr[\mathbf{y}_j \in d\mathbf{x}] \Pr[d\mathbf{x} \in T] \\ &= \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) d\mathbf{x} \Pr[\mathbf{x} \in T] \\ &= \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) E[C(\mathbf{x})] d\mathbf{x} \end{aligned} \quad (8)$$

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

$$\begin{aligned} E[U_i(\mathbf{y}_j)] &= \Pr[\mathbf{y}_j \in Z_i] \\ &= \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \end{aligned} \quad (9)$$

と表される．

それでは，これらの計算結果をふまえ， $MSE[S]$ を求めよう．

$$\begin{aligned} MSE[S] &= E[\varepsilon^2] \\ &= E[(M - \hat{M})^2] \\ &= E[M^2] - 2E[M\hat{M}] + E[\hat{M}^2] \end{aligned} \quad (10)$$

各項を個別に見てゆくと，第1項は

$$\begin{aligned}
E[M^2] &= E\left[\left\{\sum_j C(\mathbf{y}_j)\right\}^2\right] \\
&= E\left[\sum_j \sum_{j'} C(\mathbf{y}_j)C(\mathbf{y}_{j'})\right] \\
&= \sum_j \sum_{j'} E[C(\mathbf{y}_j)C(\mathbf{y}_{j'})] \\
&= 2 \sum_{j \neq j'} E[C(\mathbf{y}_j)C(\mathbf{y}_{j'})] + \sum_j E[\{C(\mathbf{y}_j)\}^2]
\end{aligned} \tag{11}$$

となる．ここで，各点はそれぞれ独立に分布するので， $C(\mathbf{y}_j)$ と $C(\mathbf{y}_{j'})$ （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11は

$$\begin{aligned}
E[M^2] &= 2 \sum_{j \neq j'} E[C(\mathbf{y}_j)C(\mathbf{y}_{j'})] + \sum_j E[\{C(\mathbf{y}_j)\}^2] \\
&= 2 \sum_{j \neq j'} E[C(\mathbf{y}_j)]E[C(\mathbf{y}_{j'})] + \sum_j E[C(\mathbf{y}_j)]
\end{aligned} \tag{12}$$

と表される．さらに式8を用いると，

$$\begin{aligned}
E[M^2] &= 2 \sum_{j \neq j'} E[C(\mathbf{y}_j)]E[C(\mathbf{y}_{j'})] + \sum_j E[C(\mathbf{y}_j)] \\
&= N(N-1) \left\{ \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) E[C(\mathbf{x})] d\mathbf{x} \right\}^2 + N \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) E[C(\mathbf{x})] \\
&= N(N-1) \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) f(\mathbf{t}) E[C(\mathbf{x})C(\mathbf{t})] d\mathbf{x} d\mathbf{t} + N \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) E[C(\mathbf{x})] d\mathbf{x} \\
&= N(N-1) \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) f(\mathbf{t}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} + N \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) \Pr[\mathbf{x} \in T] d\mathbf{x}
\end{aligned} \tag{13}$$

となる．

次に，式10の第2項を見てみよう．

$$\begin{aligned}
E[MM\hat{M}] &= E\left[\left\{\sum_j C(\mathbf{y}_j)\right\} \left\{\sum_i C(\mathbf{z}_i) \sum_j U_i(\mathbf{y}_j)\right\}\right] \\
&= E\left[\sum_i \sum_j \sum_{j'} C(\mathbf{z}_i) U_i(\mathbf{y}_j) C(\mathbf{y}_{j'})\right] \\
&= \sum_i \sum_j \sum_{j'} E[U_i(\mathbf{y}_j) C(\mathbf{z}_i) C(\mathbf{y}_{j'})]
\end{aligned} \tag{14}$$

さらに書き換えると，

$$\begin{aligned}
E[MM\hat{M}] &= \sum_i \sum_j \sum_{j'} E[U_i(\mathbf{y}_j) C(\mathbf{z}_i) C(\mathbf{y}_{j'})] \\
&= 2 \sum_i \sum_{j \neq j'} E[U_i(\mathbf{y}_j) C(\mathbf{z}_i) C(\mathbf{y}_{j'})] + \sum_i \sum_j E[U_i(\mathbf{y}_j) C(\mathbf{z}_i) C(\mathbf{y}_j)] \\
&= 2 \sum_i \sum_{j \neq j'} E[U_i(\mathbf{y}_j)] E[C(\mathbf{z}_i) C(\mathbf{y}_{j'})] + \sum_i \sum_j E[U_i(\mathbf{y}_j) C(\mathbf{z}_i) C(\mathbf{y}_j)]
\end{aligned} \tag{15}$$

である．ここで式15の第1項は

$$\begin{aligned}
E[U_i(\mathbf{y}_j)]E[C(\mathbf{z}_i)C(\mathbf{y}_j)] &= \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{t} \in Z_0} \Pr[(\mathbf{z}_i \in S) \cap (\mathbf{y}_j \in d\mathbf{t}) \cap (d\mathbf{t} \in T)] \\
&= \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{t} \in Z_0} f(\mathbf{t})\Pr[\mathbf{z}_i \cup d\mathbf{t} \in T] \\
&= \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{t} \in Z_0} f(\mathbf{t})E[C(\mathbf{z}_i)C(\mathbf{t})]d\mathbf{t}
\end{aligned} \tag{16}$$

である．一方，式15の第2項は

$$\begin{aligned}
E[U_i(\mathbf{y}_j)C(\mathbf{z}_i)C(\mathbf{y}_j)] &= \int_{\mathbf{t} \in Z_i} \Pr[(\mathbf{y}_j \in d\mathbf{t}) \cap (\mathbf{z}_i \in T) \cap (d\mathbf{t} \in T)] \\
&= \int_{\mathbf{t} \in Z_i} \Pr[\mathbf{y}_j \in d\mathbf{t}]\Pr[\mathbf{z}_i \cup d\mathbf{t} \in T] \\
&\approx \int_{\mathbf{t} \in Z_i} f(\mathbf{t})d\mathbf{t} \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] \\
&= \int_{\mathbf{t} \in Z_i} f(\mathbf{t})E[C(\mathbf{z}_i)C(\mathbf{t})]d\mathbf{t}
\end{aligned} \tag{17}$$

である．従って式15は，

$$\begin{aligned}
E[MM\hat{M}] &= 2 \sum_i \sum_{j \neq i} E[U_i(\mathbf{y}_j)]E[C(\mathbf{z}_i)C(\mathbf{y}_j)] + \sum_i \sum_j E[U_i(\mathbf{y}_j)C(\mathbf{z}_i)C(\mathbf{y}_j)] \\
&= N(N-1) \sum_i \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{t} \in Z_0} f(\mathbf{t})E[C(\mathbf{z}_i)C(\mathbf{t})]d\mathbf{t} \\
&\quad + N \sum_i \int_{\mathbf{t} \in Z_i} f(\mathbf{t})E[C(\mathbf{z}_i)C(\mathbf{t})]d\mathbf{t} \\
&= N(N-1) \sum_i \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{t} \in Z_0} f(\mathbf{t})\Pr[\mathbf{z}_i \cup \mathbf{t} \in T]d\mathbf{t} \\
&\quad + N \sum_i \int_{\mathbf{t} \in Z_i} f(\mathbf{t})\Pr[\mathbf{z}_i \cup \mathbf{t} \in T]d\mathbf{t}
\end{aligned} \tag{18}$$

となる．最後に式10の第3項は，

$$\begin{aligned}
E[\hat{M}^2] &= E\left[\left\{\sum_i C(\mathbf{z}_i)\sum_j U_i(\mathbf{y}_j)\right\}^2\right] \\
&= E\left[\sum_i \sum_{i'} C(\mathbf{z}_i)C(\mathbf{z}_{i'})\sum_j U_i(\mathbf{y}_j)\sum_{j'} U_{i'}(\mathbf{y}_{j'})\right] \\
&= \sum_i \sum_{i'} \sum_j \sum_{j'} E[C(\mathbf{z}_i)C(\mathbf{z}_{i'})U_i(\mathbf{y}_j)U_{i'}(\mathbf{y}_{j'})] \\
&= \sum_i \sum_{i'} E[C(\mathbf{z}_i)C(\mathbf{z}_{i'})]\sum_j \sum_{j'} E[U_i(\mathbf{y}_j)]E[U_{i'}(\mathbf{y}_{j'})]
\end{aligned} \tag{19}$$

である．式9を用いて変形すると，

$$\begin{aligned}
E[\hat{M}^2] &= N(N-1) \sum_i \sum_{i'} E[C(\mathbf{y}_i)C(\mathbf{y}_{i'})] E[U_i(\mathbf{z}_j)] E[U_{i'}(\mathbf{z}_j)] \\
&\quad + N \sum_i \sum_{i'} E[C(\mathbf{y}_i)C(\mathbf{y}_{i'})] E[U_i(\mathbf{z}_j)U_{i'}(\mathbf{z}_j)] \\
&= N(N-1) \sum_i \sum_{i'} E[C(\mathbf{y}_i)C(\mathbf{y}_{i'})] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in Z_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i E[\{C(\mathbf{y}_i)\}^2] E[\{U_i(\mathbf{z}_j)\}^2] \\
&= N(N-1) \sum_i \sum_{i'} E[C(\mathbf{y}_i)C(\mathbf{y}_{i'})] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in Z_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i E[C(\mathbf{y}_i)] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \\
&= N(N-1) \sum_i \sum_{i'} \Pr[\mathbf{y}_i \cup \mathbf{y}_{i'} \in T] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in Z_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i \Pr[\mathbf{y}_i \in T] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{20}$$

となる．従って，式13，18，20を式10に代入して，

$$\begin{aligned}
MSE[S] &= N(N-1) \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x})f(\mathbf{t}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x}d\mathbf{t} \\
&\quad + N \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) \Pr[\mathbf{x} \in T] d\mathbf{x} \\
&\quad - 2N(N-1) \sum_i \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{t} \in Z_0} f(\mathbf{t}) \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] d\mathbf{t} \\
&\quad - 2N \sum_i \int_{\mathbf{t} \in Z_i} f(\mathbf{t}) \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] d\mathbf{t} \\
&\quad + N(N-1) \sum_i \sum_{i'} \Pr[\mathbf{z}_i \cup \mathbf{z}_{i'} \in T] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x} \int_{\mathbf{x} \in Z_{i'}} f(\mathbf{x}) d\mathbf{x} \\
&\quad + N \sum_i \Pr[\mathbf{z}_i \in T] \int_{\mathbf{x} \in Z_i} f(\mathbf{x}) d\mathbf{x}
\end{aligned} \tag{21}$$

が誤差 ε の $MSE[S]$ である．

式21において，空間集計システム S ，点オブジェクトが従う確率密度関数 $f(\mathbf{x})$ が規定され， $\Pr[\mathbf{x} \in T]$ と $\Pr[\mathbf{x} \cup \mathbf{t} \in T]$ が任意の \mathbf{x} や \mathbf{t} について計算可能であれば，数値積分を用いて $MSE[S]$ を計算することができる．しかし残念ながら， $\Pr[\mathbf{x} \in T]$ や $\Pr[\mathbf{x} \cup \mathbf{t} \in T]$ 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．そこで本論文では，periodical continuum (Stoyan and Stoyan, 1995) と呼ばれる仮定に基づいたモデルを提案する．このモデルは，Sadahiro (1998)でも用いられているものであり，以下の様な条件を課すモデルである．

- 1) 領域 Z_0 はそれと合同な図形のlatticeによって平面をおおうことができる図形である (図2) ．
- 2) 領域 Z_0 はそれと合同な図形によって囲まれており，周囲の図形は Z_0 と同じ空間集計システム，点オブジェクト分布を持つ (図3) ．
- 3) 空間集計地区 T は，領域 Z_0 からはみ出す場合，はみ出した部分をそれに対応する領域 Z_0 内の部分に置き換える (図4) ．そして，空間集計地区 T は，修正後の空間集計地区 T のあらゆるパターンがランダムに発生するように分布するものと仮定する ．

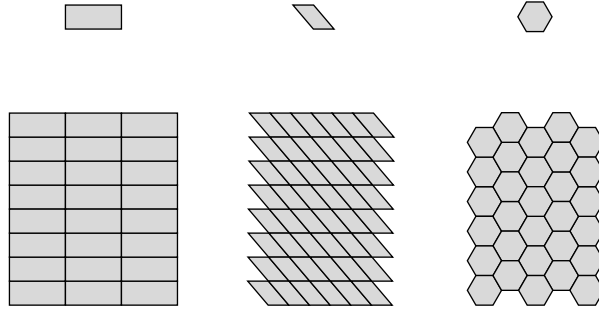


図2 領域 Z_0 の形状

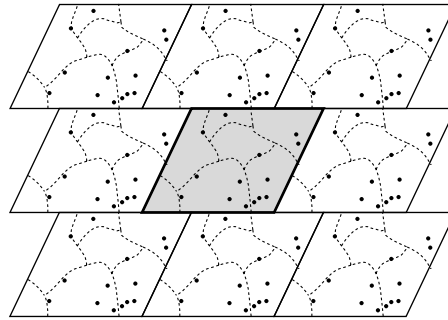


図3 領域 Z_0 とその周辺領域

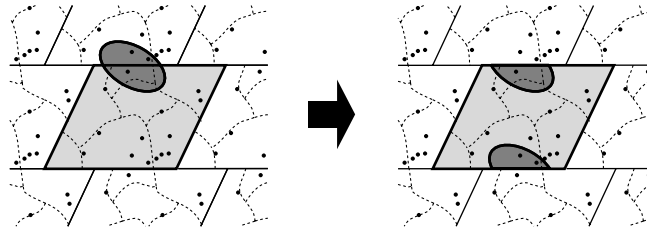


図4 空間集計地区 T の置換

これらの条件のもとでは，

$$\Pr[\mathbf{x} \in T] = \frac{B}{A_0} \quad (22)$$

が成り立ち，式21は

$$\begin{aligned} MSE[S] &= N(N-1) \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x})f(\mathbf{t}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x}d\mathbf{t} \\ &\quad + 2N \left(\frac{A}{A_0} \right) \\ &\quad - 2N(N-1) \sum_i \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{t} \in Z_0} f(\mathbf{t}) \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] d\mathbf{t} \\ &\quad - 2N \sum_i \int_{\mathbf{t} \in Z_i} f(\mathbf{t}) \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] d\mathbf{t} \\ &\quad + N(N-1) \sum_i \sum_{i'} \Pr[\mathbf{z}_i \cup \mathbf{z}_{i'} \in T] \int_{\mathbf{x} \in Z_i} f(\mathbf{x})d\mathbf{x} \int_{\mathbf{x} \in Z_{i'}} f(\mathbf{x})d\mathbf{x} \end{aligned} \quad (23)$$

と書き換えられる．

上式の中で， $\Pr[\mathbf{x} \cup \mathbf{t} \in T]$ の計算は以下のように行うことができる．いま，2点 \mathbf{x} と \mathbf{t} を内包する T と合同な

図形の測度を $m(T; |\mathbf{x}-\mathbf{t}|)$ とおく（この測度は \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ことに注意していただきたい）．すると， T の周囲領域における点 \mathbf{t} の対応点の集合を $\Omega(\mathbf{t})$ として，

$$\Pr[\mathbf{x} \cup \mathbf{t} \in T] = \sum_{\mathbf{u} \in \Omega(\mathbf{t})} \frac{m(T; |\mathbf{x}-\mathbf{u}|)}{2\pi A_0} \quad (24)$$

が成り立つ．測度 $m(T; |\mathbf{x}-\mathbf{t}|)$ は， T が単純な図形の場合には簡単に計算可能である．例えば， T が半径 r の円の場合には，

$$m(T; l) = \begin{cases} 4\pi r^2 \arccos\left(\frac{l}{2r}\right) - \pi l \sqrt{4r^2 - l^2} & (l \leq 2r), \\ 0 & (l > 2r). \end{cases} \quad (25)$$

である．その他の場合については，例えばSadahiro (1998)を参照していただきたい．また， T が複雑な図形の場合には，領域 T 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 $g_T(l)$ として，

$$g_T(l) = \frac{B^2}{l} m(T; l) \quad (26)$$

という式を用いればよい（この式の証明についてはSadahiro (1998)を参照のこと）．

なお参考のため， $f(\mathbf{x})$ が一様分布である場合について $MSE[S]$ を計算しておこう．

$$f(\mathbf{x}) = \frac{1}{A_0} \quad (27)$$

を用いれば，式21は

$$\begin{aligned} MSE[S] &= N(N-1) \left(\frac{1}{A_0}\right)^2 \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} \\ &\quad + 2N \left(\frac{A}{A_0}\right) \\ &\quad - 2N(N-1) \left(\frac{1}{A_0}\right)^2 \sum_i A_i \int_{\mathbf{t} \in Z_0} \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] d\mathbf{t} \\ &\quad - 2N \frac{1}{A_0} \sum_i \int_{\mathbf{t} \in Z_i} \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] d\mathbf{t} \\ &\quad + N(N-1) \left(\frac{1}{A_0}\right)^2 \sum_i \sum_{i'} A_i A_{i'} \Pr[\mathbf{z}_i \cup \mathbf{z}_{i'} \in T] \end{aligned} \quad (28)$$

となる．ここで，

$$\begin{aligned} \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} &= \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} E[C(\mathbf{x})C(\mathbf{t})] d\mathbf{x} d\mathbf{t} \\ &= E \left[\int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} C(\mathbf{x})C(\mathbf{t}) d\mathbf{x} d\mathbf{t} \right] \\ &= E \left[\int_{\mathbf{x} \in Z_0} C(\mathbf{x}) d\mathbf{x} \int_{\mathbf{t} \in Z_0} C(\mathbf{t}) d\mathbf{t} \right] \\ &= E \left[\left\{ \int_{\mathbf{x} \in Z_0} C(\mathbf{x}) d\mathbf{x} \right\}^2 \right] \\ &= E[B^2] \\ &= B^2 \end{aligned} \quad (29)$$

であることに注意すれば，

$$\begin{aligned}
MSE[S] &= N(N-1) \left(\frac{B}{A_0} \right)^2 + 2N \left(\frac{B}{A_0} \right) \\
&\quad - 2N(N-1) \left(\frac{1}{A_0} \right)^2 \sum_i A_i \int_{t \in Z_0} \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] dt \\
&\quad - 2N \frac{1}{A_0} \sum_i \int_{t \in Z_i} \Pr[\mathbf{z}_i \cup \mathbf{t} \in T] dt \\
&\quad + N(N-1) \left(\frac{1}{A_0} \right)^2 \sum_i \sum_{i'} A_i A_{i'} \Pr[\mathbf{z}_i \cup \mathbf{z}_{i'} \in T]
\end{aligned} \tag{30}$$

を得る．これに対し，同じ条件下での面積按分法による $MSE[S]$ は

$$MSE[S] = \frac{N}{A_0} \left\{ B - \sum_i \frac{1}{A_i} \int_{x \in Z_i} \int_{t \in Z_i} \Pr[\mathbf{x} \cup \mathbf{t} \in S] dt dx \right\} \tag{31}$$

であり，代表点内包法と比べて点の個数 N の影響が小さいことが分かる．

3. 正方形格子網における推定精度の評価

前節では，periodical continuumという条件の下での推定誤差 ε の平均自乗誤差 $MSE[S]$ を導出した．そこで本節では，最も代表的な空間集計システムである正方形格子網（square lattice）を取り上げ，様々な状況下での空間集計地区 T や確率密度関数 $f(\mathbf{x})$ と $MSE[S]$ の関係を数値的に分析する．

まず始めに，正方形格子網における $MSE[S]$ を導出しておこう．なおここでは，各セルの代表点は全てセル内の対応する同一箇所に位置し，また，各セルにおける $f(\mathbf{x})$ の形状は全て同一であると仮定する．すると式23は以下のように書き換えることができる．

$$\begin{aligned}
MSE[S] &= N(N-1) \int_{t \in Z_0} \int_{x \in Z_0} f(\mathbf{x}) f(\mathbf{t}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] dx dt \\
&\quad + 2N \left(\frac{B}{A_0} \right) \\
&\quad - 2N(N-1) \int_{t \in Z_0} f(\mathbf{t}) \Pr[\mathbf{z} \cup \mathbf{t} \in T] dt \\
&\quad - 2N \left(\frac{A_0}{A} \right) \int_{t \in U} f(\mathbf{t}) \Pr[\mathbf{z} \cup \mathbf{t} \in T] dt \\
&\quad + N(N-1) \left(\frac{A}{A_0} \right)^2 \sum_i \sum_{i'} \Pr[\mathbf{z}_i \cup \mathbf{z}_{i'} \in T]
\end{aligned} \tag{32}$$

但し， U はセル領域， A はその面積をそれぞれ表す．

有限領域内での数値分析では，領域の境界部分でのboundary effectが生ずることがある．ここでは， T ， $f(\mathbf{x})$ ， $MSE[S]$ の關係に注目しているので，boundary effectを避けるために以下のような条件を付加する．

1) Z_0 は分析で用いられるどの T と比べても十分広い．

2) 点の個数 N は Z_0 の面積 A_0 に比例し，点の密度 $\mu = N/A_0$ は A_0 によらず一定である．

そして，領域 Z_0 が大きくなるときの極限值を計算する．

上の条件の下では，式32は以下のように変形される．

$$\begin{aligned}
MSE[S] &= A_0\mu(A_0\mu - 1) \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x})f(\mathbf{t}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x}d\mathbf{t} \\
&\quad + 2\mu B \\
&\quad - 2A_0\mu(A_0\mu - 1) \int_{\mathbf{t} \in Z_0} f(\mathbf{t}) \Pr[\mathbf{z} \cup \mathbf{t} \in T] d\mathbf{t} \\
&\quad - \frac{2A_0^2\mu}{A} \int_{\mathbf{t} \in U} f(\mathbf{t}) \Pr[\mathbf{z} \cup \mathbf{t} \in T] d\mathbf{t} \\
&\quad + (A_0\mu - 1) \frac{\mu A^2}{A_0} \sum_i \sum_{i'} \Pr[\mathbf{z}_i \cup \mathbf{z}_{i'} \in T]
\end{aligned} \tag{33}$$

さらに書き換えると，

$$\begin{aligned}
MSE[S] &= \frac{\mu^2 A_0}{2\pi} \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x})f(\mathbf{t}) m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x}d\mathbf{t} \\
&\quad + 2\mu B \\
&\quad - \frac{\mu^2 A_0}{\pi} \int_{\mathbf{t} \in Z_0} f(\mathbf{t}) m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \\
&\quad - \frac{\mu A_0}{\pi A} \int_{\mathbf{t} \in U} f(\mathbf{t}) m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \\
&\quad + \frac{\mu^2 A^2}{2\pi A_0} \sum_i \sum_{i'} m(T; |\mathbf{z}_i - \mathbf{z}_{i'}|) \\
&= \frac{\mu^2}{2\pi A_0} \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} \{A_0 f(\mathbf{x})\} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x}d\mathbf{t} \\
&\quad + 2\mu B \\
&\quad - \frac{\mu^2}{\pi} \int_{\mathbf{t} \in Z_0} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \\
&\quad - \frac{\mu}{\pi A} \int_{\mathbf{t} \in U} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \\
&\quad + \frac{\mu^2 A^2}{2\pi A_0} \sum_i \sum_{i'} m(T; |\mathbf{z}_i - \mathbf{z}_{i'}|)
\end{aligned} \tag{34}$$

となる．式34で領域 $Z_0 \rightarrow \infty$ とすると，

$$\begin{aligned}
MSE_\infty[S] &= \lim_{Z_0 \rightarrow \infty} MSE[S] \\
&= \left[\begin{aligned} &\frac{1}{2\pi A} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in Z_0} \{A_0 f(\mathbf{x})\} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x}d\mathbf{t} \\ &- \frac{1}{\pi} \int_{\mathbf{t} \in Z_0} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \\ &+ \frac{A^2}{2\pi A} \sum_i m(T; |\mathbf{z} - \mathbf{z}_i|) \end{aligned} \right] \mu^2 \\
&\quad + \left[2B - \frac{1}{\pi A} \int_{\mathbf{t} \in U} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \right] \mu
\end{aligned} \tag{35}$$

となる．

ここで，式35の各項の持つ意味について検討してみよう．但しここでは，セルの大きさや形状は常に一定であり，また，各セル内での $f(\mathbf{x})$ の形状や代表点の位置はセルによらず全て同一であるものとする．

$$1) \frac{1}{2\pi A} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in Z_0} \{A_0 f(\mathbf{x})\} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x} d\mathbf{t} \quad (\text{QP1})$$

この項は， T の大きさや形状，及び， $f(\mathbf{x})$ の形状に依存し，代表点の位置とは無関係な項である．

$$2) \frac{1}{\pi} \int_{\mathbf{t} \in Z_0} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \quad (\text{QP3})$$

この項は， T の大きさや形状， $f(\mathbf{x})$ の形状，さらに代表点の位置に依存する項である．

$$3) \frac{A^2}{2\pi A} \sum_i m(T; |\mathbf{z} - \mathbf{z}_i|) \quad (\text{QP2})$$

この項は， T の大きさや形状にのみ依存し， $f(\mathbf{x})$ の形状やセル内での代表点の位置とは無関係である．

$$4) \frac{1}{\pi A} \int_{\mathbf{t} \in U} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{z} - \mathbf{t}|) d\mathbf{t} \quad (\text{RP2})$$

この項は， T の大きさや形状， $f(\mathbf{x})$ の形状，さらに代表点の位置に依存する項である．

なお数値分析では，異なる大きさの T 間で $MSE[S]$ が比較可能なように，

$$\frac{MSE_{\infty}[S]}{B^2} = Q\mu^2 + R\mu \quad (36)$$

とおき， Q 及び R の計算を行う．但し簡単のため，セルの面積 A は1とする．

またここでは比較のために，面積按分法を用いた場合の平均自乗誤差 $MSE[S]$ も併せて計算する．

$$\begin{aligned} MSE[S] &= N(N-1) \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) f(\mathbf{t}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} \\ &\quad + N \left(\frac{B}{A_0} \right) + N(N-1) \left(\frac{B}{A_0} \right)^2 \\ &\quad - 2N(N-1) \frac{1}{A} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in Z_0} f(\mathbf{x}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} \\ &\quad - 2N \frac{A_0}{A^2} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in U} f(\mathbf{x}) \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} \\ &\quad + N \frac{1}{A^2} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in U} \Pr[\mathbf{x} \cup \mathbf{t} \in T] d\mathbf{x} d\mathbf{t} \end{aligned} \quad (37)$$

この式を変形すると，

$$\begin{aligned} MSE[S] &= \frac{\mu^2}{2\pi A_0} \int_{\mathbf{t} \in Z_0} \int_{\mathbf{x} \in Z_0} \{A_0 f(\mathbf{x})\} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x} d\mathbf{t} \\ &\quad + \mu B + \mu^2 B^2 \\ &\quad - \frac{\mu^2}{\pi A} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in Z_0} \{A_0 f(\mathbf{x})\} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x} d\mathbf{t} \\ &\quad - \frac{\mu}{\pi A^2} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in U} \{A_0 f(\mathbf{x})\} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x} d\mathbf{t} \\ &\quad + \frac{\mu}{2\pi A^2} \int_{\mathbf{t} \in U} \int_{\mathbf{x} \in U} m(T; |\mathbf{x} - \mathbf{t}|) d\mathbf{x} d\mathbf{t} \end{aligned} \quad (38)$$

となり，やはり先程と同じように

$$MSE_{\infty}[S] = \left[\begin{aligned} & B^2 + \frac{1}{2\pi A} \int_{t \in U} \int_{x \in Z_0} \{A_0 f(\mathbf{x})\} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \\ & - \frac{1}{\pi A} \int_{t \in U} \int_{x \in Z_0} \{A_0 f(\mathbf{x})\} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \end{aligned} \right] \mu^2$$

$$+ \left[\begin{aligned} & B + \frac{1}{2\pi A^2} \int_{t \in U} \int_{x \in U} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \\ & - \frac{1}{\pi A^2} \int_{t \in U} \int_{x \in U} \{A_0 f(\mathbf{x})\} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \end{aligned} \right] \mu \quad (39)$$

とにおいて

$$\frac{MSE_{\infty}[S]}{B^2} = Q\mu^2 + R\mu \quad (40)$$

を分析する（ここでもセルの面積 A は1とする）。なおここでも，先程と同様に，式40の各項を検討しておく。

$$1) \frac{1}{2\pi A} \int_{t \in U} \int_{x \in Z_0} \{A_0 f(\mathbf{x})\} \{A_0 f(\mathbf{t})\} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \quad (QA2)$$

前出の1)と同一の項であり， T の大きさや形状， $f(\mathbf{x})$ の形状に依存する項である。

$$2) \frac{1}{\pi A} \int_{t \in U} \int_{x \in Z_0} \{A_0 f(\mathbf{x})\} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \quad (QA3)$$

この項は， T の大きさや形状， $f(\mathbf{x})$ の形状に依存している。

$$3) \frac{1}{2\pi A^2} \int_{t \in U} \int_{x \in U} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \quad (RA3)$$

T の大きさや形状にのみ依存し， $f(\mathbf{x})$ の形状とは無関係である。

$$4) \frac{1}{\pi A^2} \int_{t \in U} \int_{x \in U} \{A_0 f(\mathbf{x})\} m(T; |\mathbf{x} - \mathbf{t}|) dx dt \quad (RA2)$$

この項は， T の大きさや形状， $f(\mathbf{x})$ の形状に依存する項である。

3.1 空間集計地区 T の大きさと平均自乗誤差 $MSE[S]$ の関係

まずはじめに，空間集計地区 T の大きさと平均自乗誤差 $MSE[S]$ の関係を分析する。 T は円形とし，半径0.1から5.0の範囲で計算を行った。各セルの代表点は全てセルの重心にある。確率密度関数 $f(\mathbf{x})$ としては，1) 一様分布（図5a），2) セル重心直上を頂点とする四角錐型の分布（図5b），3) セル重心を中心とする一辺0.5の正方形を底面とする直方体型の分布（図5c）の3通りの分布形を取り上げた。

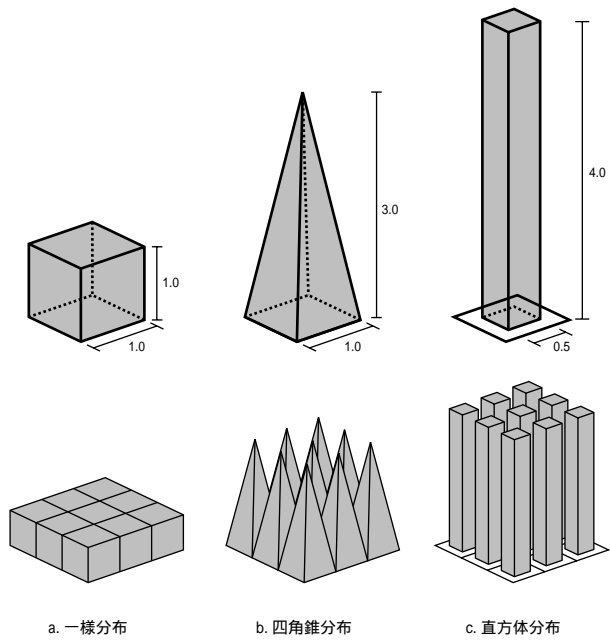


図5 セル内の確率密度分布

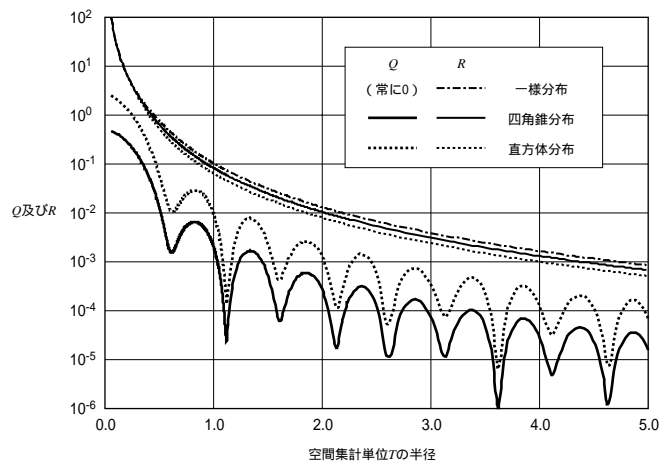


図6 面積按分法を用いた場合の空間集計地区Tと平均自乗誤差MSE[S]の関係

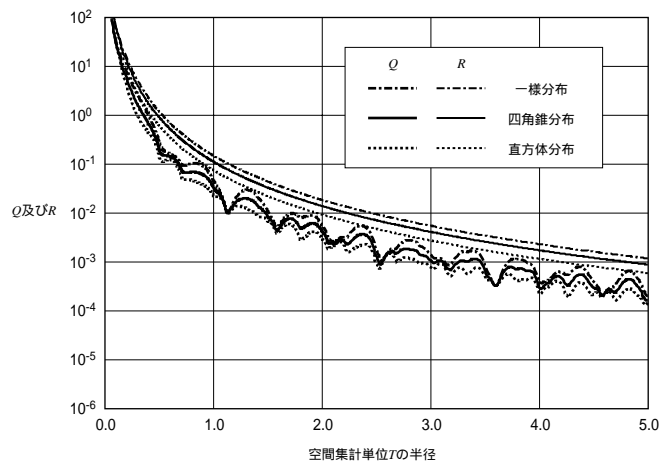


図7 代表点内包法を用いた場合の空間集計地区Tと平均自乗誤差MSE[S]の関係

それでは、図6及び7を検討しよう。全体的な傾向として、空間集計地区 T が大きくなるほど、平均自乗誤差 $MSE[S]$ は相対的に小さくなる。但し面積按分法、代表点内包法のいずれの場合でも、 Q 値は周期的に変化しながら小さくなっている。これは、latticeと $f(x)$ の持つ空間的周期性によるものであり、代表点内包法の場合にはさらに代表点の配置の持つ周期性が加わって複雑な変化を見せている。このことは、点オブジェクトや空間単位地区が等間隔で並んでいるような場合、データの推定誤差に十分注意する必要がある、ということを表している。

面積按分法の場合、 Q 値に関しては一様分布の場合が最も小さく、次いで四角錐分布、直方体分布である。しかし R 値に関しては、逆に一様分布の場合が最も大きく、次いで四角錐分布、直方体分布となっている。代表点内包法の場合には、 Q 値、 R 値のいずれについても、直方体分布の場合が最小であり、次いで四角錐分布、一様分布である。つまり、面積按分法は一様分布に最も良く適合し、直方体分布や四角錐分布のようにある程度集中した分布にはあまりなじまない、反対に代表点内包法は直方体分布に最も良く適合し、一様分布には適当ではない、とそれぞれ考えることができるであろう。

Q 値と R 値を比較すると、いずれの場合でも後者の方が前者よりも大きくなっている。しかし、式36及び40から明らかとなり、点の密度が高い場合には Q 値の方が R 値よりも $MSE[S]$ に対して与える影響が大きい。即ち、点の密度 μ が低い場合には、 Q 値と R 値はどちらも $MSE[S]$ に影響を与えるが、 μ が大きくなると、 Q 値の方が R 値よりも大きな影響力を持つ。そこで、これら二つの値の影響が等しくなるような μ 値を T の大きさとの関連から求めた(図8)。この図から、代表点内包法の場合には、 $\mu=10$ 程度で Q 値と R 値の影響が等しくなることがわかる。即ち、 μ が10よりも十分大きければ、 $MSE[S]$ はほぼ Q 値のみで決定され、 R 値の影響は無視できると言える。しかし面積按分法の場合には、これら二つの影響が等しくなるのは $\mu=100 \sim 1000$ 程度であり、 $MSE[S]$ を考える上で Q 値と R 値のどちらも考慮する必要が高い。

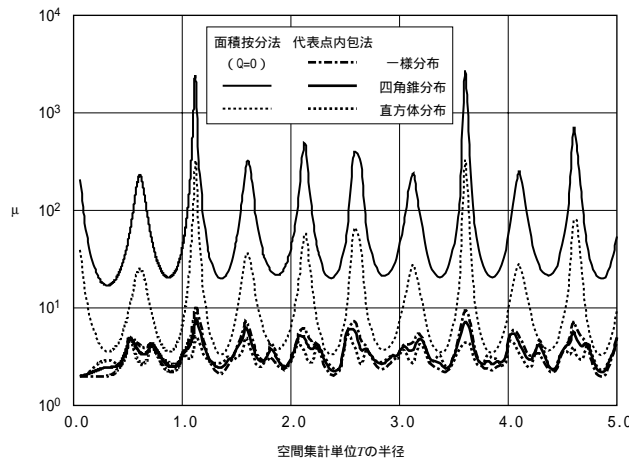


図8 Q 値と R 値の影響が等しくなる μ

3.2 直方体分布における分布の集中度と平均自乗誤差 $MSE[S]$ の関係

次に、 $f(x)$ が直方体分布である場合について、分布の集中度と平均自乗誤差 $MSE[S]$ の関係を分析する。面積按分法は、点の分布として一様分布を暗に仮定した推定法であるのに対し、代表点内包法は代表点に点が集中していることを仮定している。従ってそれぞれの方法は、仮定が正しければ良い推定値を与え、仮定が誤っていればいるほど推定値の誤差が大きくなる。そこでここでは、一様分布から集中分布へ図9のように変化するにつれて、 $MSE[S]$ がどのように変化するかを調べ、それぞれの仮定からの乖離と $MSE[S]$ の関係を分析する。空間集計地区 T については、既に見たようにその大きさによって $MSE[S]$ が周期的に変化するため、一周期の平均値を求める。具体的には、 T の半径が0.62 ~ 1.12, 1.62 ~ 2.12, 3.62 ~ 4.12の3つの場合を取り上げ、各変域内での平均値を算出する。なお、異なる大きさの T での比較のため、 $MSE[S]$ については先ほどと同様に Q 及び R を用いて評価する。

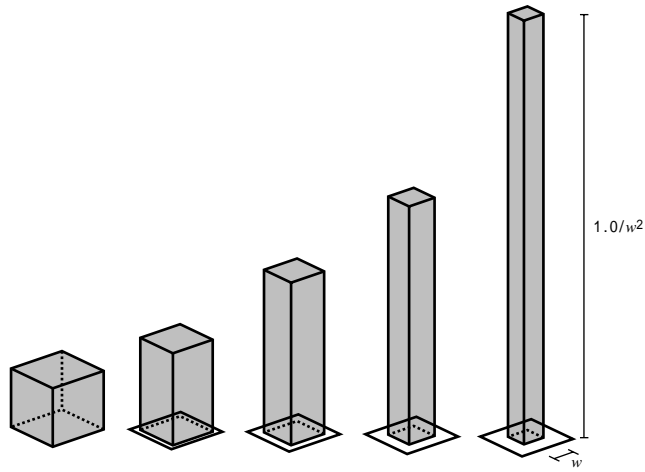
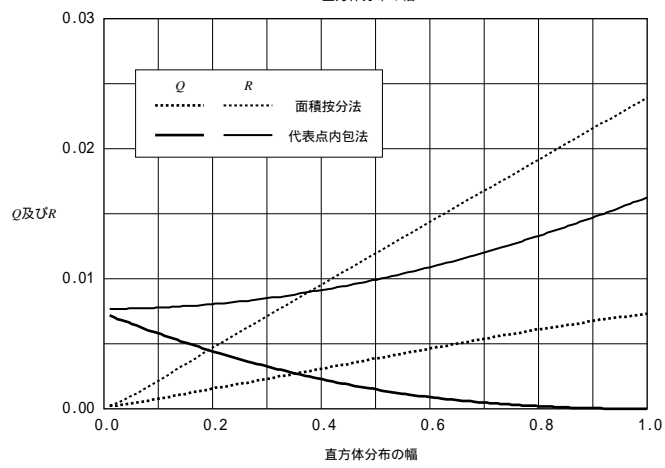
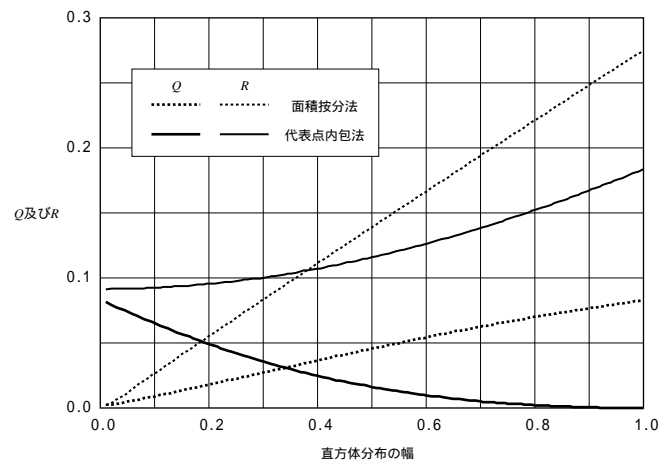


図9 一様分布から集中分布への変化



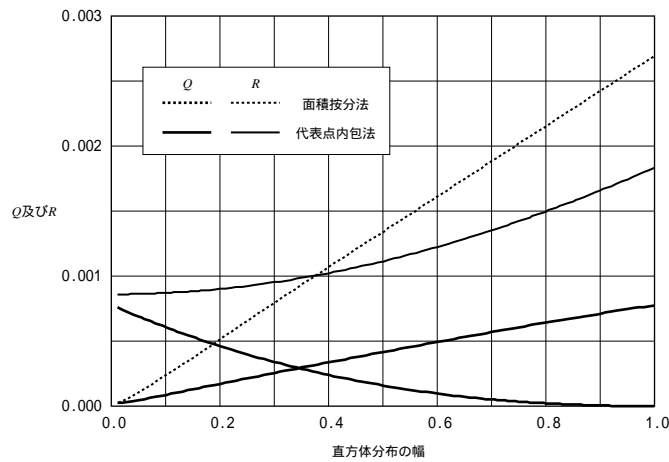


図10 直方体分布の集中度とMSE[S]の関係

a) T の半径=0.62 ~ 1.12, b) T の半径=1.62 ~ 2.12, c) T の半径=3.62 ~ 4.12

これらの図は、 T の大きさがそれぞれ異なるが、縦軸のスケールに差は見られるものの、全体的な形はほとんど同じと良い。代表点内包法の場合には、分布が集中するにつれて Q 値、 R 値が共に小さくなり、推定誤差が小さくなることを確認できる。理論的には、前出のQP1-QP2項が分布の集中に従ってQP3項に近づくこと、及び、RP2項が $2B$ に近づくことを示すことが可能であり、この計算結果はそれを裏付けていると言える。一方面積按分法の場合には、分布が集中するにつれて R 値はやはり小さくなるものの、 Q 値は大きくなっている。RA2項は分布の集中に従って大きくなり、極限では $2B$ になる。そのため、 R 値は緩やかに増加する。このことから、点の密度 μ がある程度以上であれば、分布が集中するにつれて推定誤差が大きくなるのがわかる。

一方、二つの面補間法の推定誤差が均衡する分布を考えてみると、 T の大きさや Q 値、 R 値によらず、直方体分布の幅が0.3 ~ 0.4あたりである。即ち、各セルの中央部約10%程度の領域に点が集中しているような状態が、代表点内包法でも面積按分法でも同じような推定誤差をもたらす。これよりも点が集中していれば、代表点内包法の方が良い推定値を与え、反対に点が分散していれば、面積按分法の方が良い推定値を与える、ということになる。

3.3 直方体分布におけるセル代表点と分布の位置関係

最後に、 $f(x)$ が直方体分布の場合について、分布の集中度と平均自乗誤差MSE[S]の関係を分析する。方法は先ほどと同様であり、直方体分布の幅としては0.1, 0.3, 0.5の3つの場合を考える。また T の大きさについては、先程と同様に半径が0.62 ~ 1.12, 1.62 ~ 2.12, 3.62 ~ 4.12の3つの場合を取り上げ、各変域内での平均値を算出する。

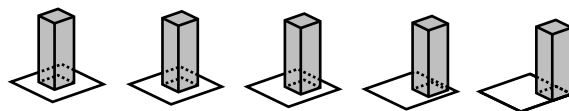


図11 四角錐分布とセル代表点の変化

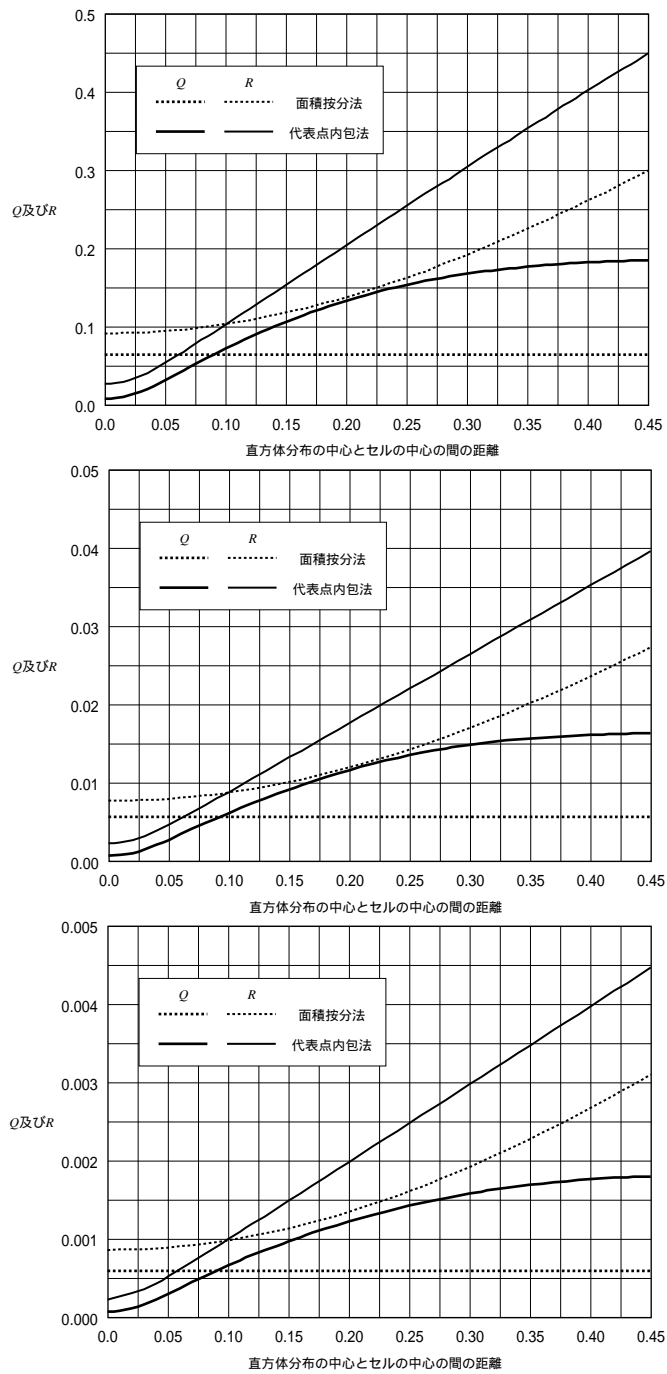


図12 直方体分布の中心からのずれとMSE[S]の関係 (直方体分布の幅=0.1)
 a) T の半径=0.62 ~ 1.12 , b) T の半径=1.62 ~ 2.12 , c) T の半径=3.62 ~ 4.12

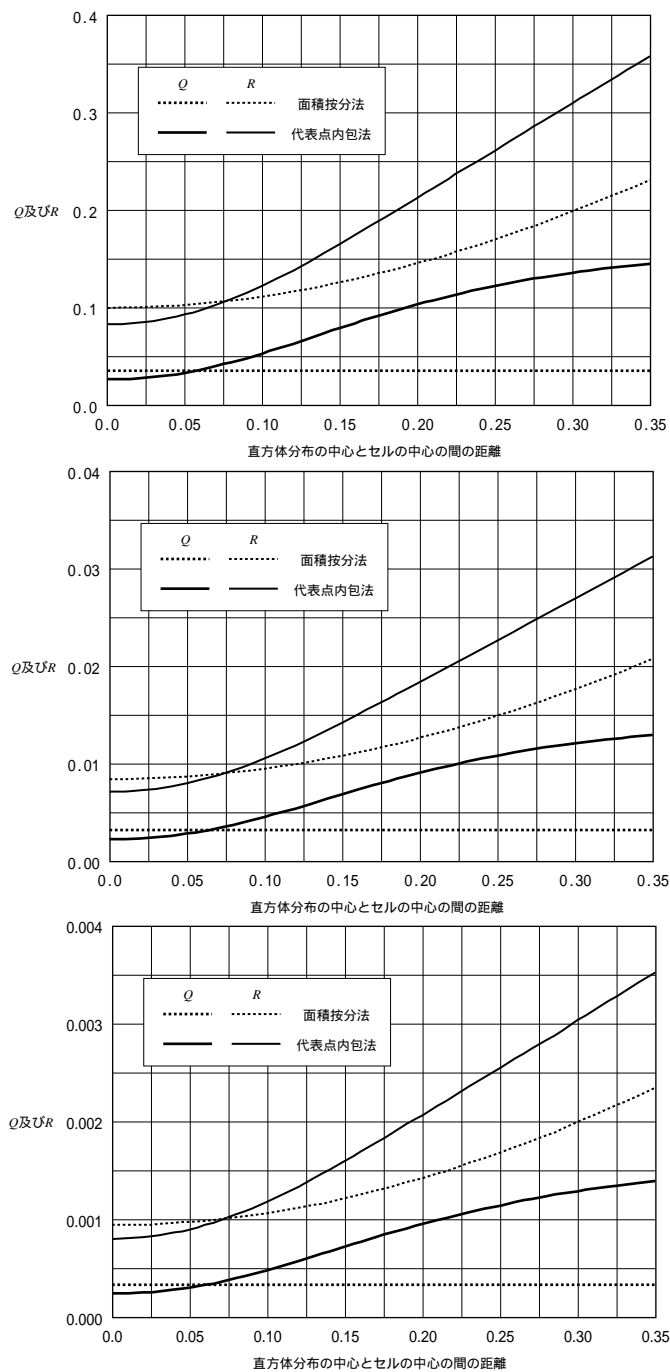


図13 直方体分布の中心からのずれとMSE[S]の関係 (直方体分布の幅=0.3)
 a) T の半径=0.62 ~ 1.12 , b) T の半径=1.62 ~ 2.12 , c) T の半径=3.62 ~ 4.12

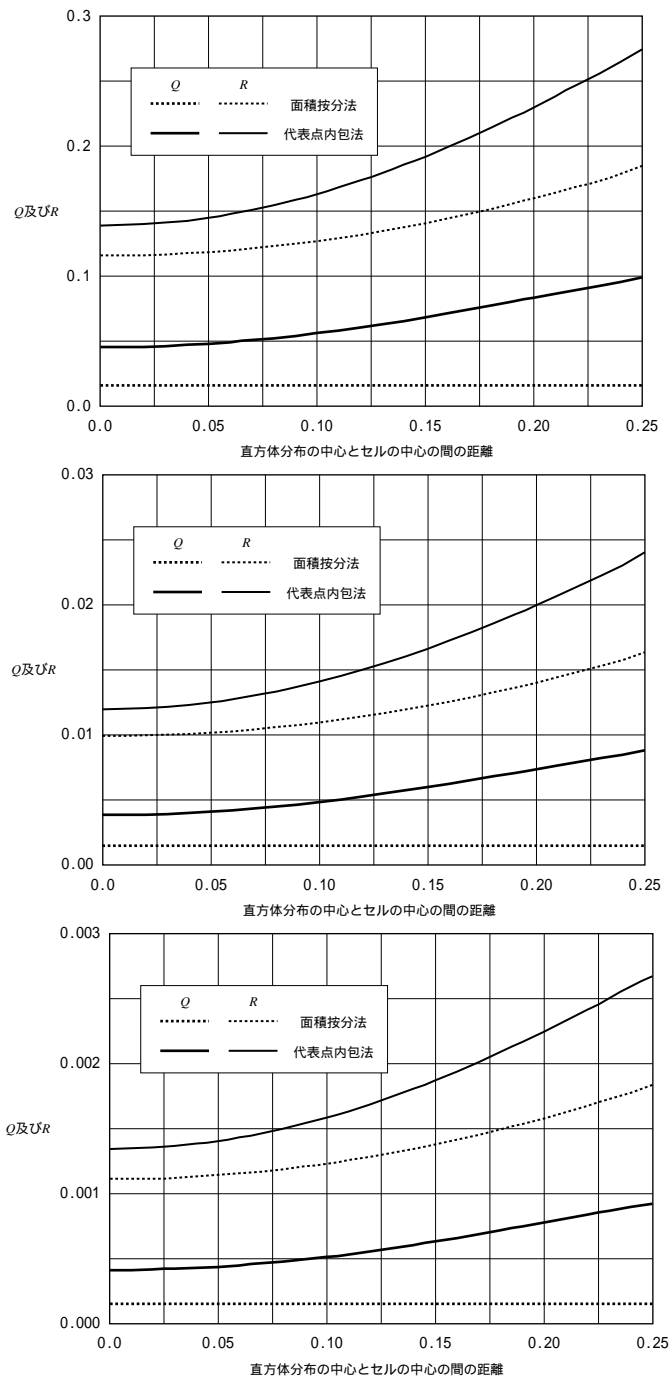


図14 直方体分布の中心からのずれと $MSE[S]$ の関係（直方体分布の幅=0.5）
 a) T の半径=0.62 ~ 1.12, b) T の半径=1.62 ~ 2.12, c) T の半径=3.62 ~ 4.12

図12~14を見ると、面積按分法における Q 値を除き、どのような場合でも分布がセルの中心からずれるほど推定誤差が大きくなることわかる。誤差の増加傾向は、空間集計地区 T の大きさにはあまり関係ないが、直方体分布の形については、集中度が高いほど、ずれによる誤差の増加が大きくなっている。図12及び13において、面積按分法のグラフと代表点内包法のグラフの交点に着目すると、いずれの場合でもずれが比較的小さな段階で、 $MSE[S]$ の大小が入れ替わっている。即ち、直方体分布のずれが非常に小さな場合には、代表点内包法の方が面積按分法よりも良い推定値を与えるが、直方体分布が多少、中心からずれると、面積按分法の方が代表点内包法よりも良い結果をもたらすことになる。このことは、面積按分法の方が代表点内包法よりも分布のずれに対して頑健であることを示している。

4. 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、代表点内包法を用いて別の空間集計単位に変換する際の変換誤差を分析した。理論的な考察に加え、正方形格子網を用いて点の分布形が推定誤差に与える影響や、代表点内包法と面積按分法との比較などを行い、実際的な知見を得ることができた。

今後の研究課題としては、変換先の空間集計地区が単独ではなく領域分割を構成する空間集計システムの場合の分析がまず挙げられる。メッシュデータから町丁目単位へのデータ変換など、このようなデータ変換はGISでは頻繁に行われ、その際の推計誤差の扱いはその後の分析結果へも大きな影響を与える。また、個数データだけではなく密度データに代表されるような連続データ、さらには、カテゴリカルデータの変換なども今後取り上げるべき問題であろう。

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. (1998): Accuracy of Count Data Transferred through the Areal Weighting Interpolation Method. *Discussion Paper Series*, **76E**, Department of Urban Engineering, University of Tokyo.
- 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 1 距離が l だけ離れている2点を内包する合同な図形の測度

距離が l だけ離れている2点を内包する図形 T と合同な図形の測度は、 T が半径 r の円の場合、

$$m(T;l) = \begin{cases} 4\pi r^2 \arccos\left(\frac{l}{2r}\right) - \pi l \sqrt{4r^2 - l^2} & (l \leq 2r), \\ 0 & (l > 2r). \end{cases} \quad (\text{A 01})$$

である。また、 T が一边 d の正方形の場合には、

$$m(T;l) = \begin{cases} 2\pi d^2 - 8dl + 2l^2 & (l \leq d), \\ 8d\sqrt{l^2 - d^2} - 2(2d^2 + l^2) + 4d^2 \left(\arcsin \frac{d}{l} - \arccos \frac{d}{l} \right) & (d < l \leq \sqrt{2}d), \\ 0 & (\sqrt{2}d < l). \end{cases} \quad (\text{A 02})$$

である。

Appendix 2 正方形の重心と、正方形内の任意の点との間の距離 l の確率分布

点間の距離 l の確率分布関数 $F(l)$ は、正方形の一边を d として、

$$F(l) = \begin{cases} \frac{\pi l^2}{d^2} & \left(l \leq \frac{d}{2} \right) \\ \frac{1}{d} \sqrt{4l^2 - d^2} + \frac{\pi l^2}{d^2} - \frac{4l^2}{d^2} \arccos \frac{d}{2l} & \left(\frac{d}{2} < l \leq \frac{\sqrt{2}d}{2} \right) \end{cases} \quad (\text{A 03})$$

であり、確率密度関数 $f(l)$ は、

$$f(l) = \begin{cases} \frac{2\pi l}{d^2} & \left(l \leq \frac{d}{2} \right) \\ \frac{2\pi l}{d^2} - \frac{8l}{d^2} \arccos \frac{d}{2l} & \left(\frac{d}{2} < l \leq \frac{\sqrt{2}d}{2} \right) \end{cases} \quad (\text{A 04})$$

である。