【土木計画学研究・講演集　No.23(2)　2000年11月】

移動平均モデルに基づくKrigingを用いた空間内挿 *

Spatial Interpolation by Kriging based on Moving-Average Model

堤盛人** ・清水英範*** ・井出裕史****
Morito TSUTSUMI, Eihan SHIMIZU and Hiroshi IDE

1. はじめに

地理的・空間的に連続して分布するデータ（空間データ）を扱う分析においては、結果の表示や分析自体のために、何らかの内挿を必要とする場合が少なくない。空間データを合理的に内挿する手法の一つとして、鉱物学や地質学など自然科学と関わりの深い空間統計学の分野において、Kriging と呼ばれる手法が確立されつつある。Kriging は、ある条件下で最良線形不偏推定量を与えるという意味で統計学的に優れた手法である。

一方、社会経済分析と関わりの深い空間計量経済学と呼ばれる分野においては、空間データを用いた回帰分析において誤差項に何らかの系列相関が生じる場合に、誤差要素モデルと呼ばれるモデルが用いられる。ECM は、誤差項における未知の空間的相関に対し、パラメータを用いたモデル化によって対処しようとするものである。

本論文では、ECM に基づきながら Kriging を適用して空間内挿を行うことの理論的課題とその解決法について検討を行い、移動平均モデルに基づく Kriging による空間内挿法を提示する。

2. Kriging による空間内挿

(1) Kriging の基本的考え方

内挿手法の一つとして、鉱物学や地質学など自然科学と関わりの深い分野である geostatistics において、Kriging が確立されつつある（例えば、Bailey and Gatrell (1995)[1)]）。Kriging では、データが未知の地点 k における値 $\hat{\varepsilon}_k$ が、周囲のデータが既知の観測点における値 ε_i の線形和として、次式のように書けるものと仮定する。

$$\hat{\varepsilon}_k = \sum_{i=1}^{I} \lambda_i \varepsilon_i = \boldsymbol{\lambda}^t \boldsymbol{\varepsilon} \qquad (1)$$

t はベクトル及び行列の転置を表し、i は観測点の番号を表す（$i=1,2,\cdots,I$）。

Kriging を実際に適用する際には、任意の２地点 i, j における値 $\varepsilon_i, \varepsilon_j$ の共分散は、その２地点間の距離 d_{ij} のみによって決まるという定常性を仮定する。

$$Cov(\varepsilon_i, \varepsilon_j) = c(d_{ij}) \qquad (2)$$

式(2)の関数は covariogram と呼ばれる。このような条件のもと、Kriging は最良線形不偏推定量（Best Linear Unbiased Predictor : BLUP）を与えるという意味で統計学的に優れた手法である。具体的には、不偏性に関する制約条件のもとでの期待二乗誤差最小化から、式(1)における重み λ_i が次のように導かれる。

$$\boldsymbol{\lambda} = \boldsymbol{C}^{-1}\boldsymbol{c} \qquad (3)$$

ここで、$\boldsymbol{\lambda}$ は λ_i を各要素とするベクトル、$\boldsymbol{C}$ は観測点におけるデータ相互の共分散行列、$\boldsymbol{c}$ は値を求めたい地点のデータと観測点におけるデータとの間の共分散ベクトルである。ここで仮定した定常性の概念は、本論文において最も重要な部分であり、(2)で改めて説明を行う。

以下、本論文では、y を被説明変数、x_h（$h=1,2,\cdots,H$）を説明変数、β_h（$h=0,1,\cdots,H$）をパラメータとする線形回帰モデルを扱う。

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_H x_H \qquad (4)$$

そこで、次に、この回帰モデルの推定結果に Kriging を適用する方法について簡単に説明する。誤差項を含んだ統計モデルを以下のように行列表記する。

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \qquad (5)$$

$$\boldsymbol{y} = (y_1, \cdots, y_i, \cdots, y_I)^t,\quad \boldsymbol{\beta} = (\beta_0, \beta_1, \cdots, \beta_H)^t,\quad \boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_i, \cdots, \varepsilon_I)^t,\quad \boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{H1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1I} & \cdots & x_{HI} \end{pmatrix}$$

$\boldsymbol{\varepsilon}$ は誤差項であり、式(6),(7)を満たすものと仮定する。

$$E(\boldsymbol{\varepsilon}) = \boldsymbol{0} \qquad (6)$$

$$Cov(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I} \qquad (7)$$

誤差項に対する通常の仮定（式(7)）からの違背が生じる場合でも、誤差項の分散共分散行列 $Cov(\boldsymbol{\varepsilon})$ が既知であ

*キーワード：データ内挿
**正員、博士（工学）　***正員、工学博士
東京大学大学院　工学系研究科 社会基盤工学専攻
（〒113-8656 東京都文京区本郷７−３−１
E-mail: tsutsumi@planner.t.u-tokyo.ac.jp）
****正員、修士（工学）　日本政策投資銀行　本店都市開発部
（〒100-0004　東京都千代田区大手町１−９−１）

れば、一般化最小二乗法（GLS）などによりパラメータを推定すればよい。例えば、$Cov(\varepsilon)$ が既知の行列 Σ で表されるとすると、一般化最小二乗法によるパラメータの推定値（量）は次式により求められる。

$$\beta_{GLS}=\left(X'\Sigma^{-1}X\right)^{-1}X'\Sigma^{-1}y \qquad (8)$$

ここで、観測地点以外の地点の y の値を Kriging によって推定する場合、式(7)において使用した分散共分散行列 $\overline{C}$ と全く異なる covariogram を想定することは、技術的には何ら問題無く可能ではあるが、論理的には極めて奇異で不適当であろう。つまり、Σ と同じ構造を持つ covariogram を用いて C,c を求め、式(3)によって内挿に関わるパラメータ（重み）λ を求めるのがより合理的な方法と考えられる。

パラメータ β,λ が求められれば、ある地点 k における推定値（量）$\hat{y}_k$ は次式により求められる。

$$\hat{y}_k=x_k\hat{\beta}+\hat{\lambda}'\hat{\varepsilon} \qquad (9)$$

（$\hat{\beta},\hat{\lambda},\hat{\varepsilon}$ ：$\beta,\lambda,\varepsilon$ の推定値（量））

このような考え方に基づく内挿のモデル化は、General Linear Model （GLM）と呼ばれる。本論文では、以後、この GLM の適用を前提とする。

(2) covariogram における仮定

前節において説明したように、Kriging の適用に際しては、まず covariogarm の設定における定常性の仮定が重要となる。定常とは、観測データ ε_i の１次モーメント並びに２次モーメントに関して以下の性質が満足されることを言う。

$$E[\varepsilon_i]=0 \qquad (10)$$

$$E\left[(\varepsilon_i-E[\varepsilon_i])(\varepsilon_j-E[\varepsilon_j])\right]=c(d_{ij}) \qquad (11)$$

ここで、$c(\cdot)$ が covariogram であり、地点 i,j 間の距離 d_{ij} のみの関数として定義されている。（なお、ここでの定常は、厳密には弱定常（weak stationary）と呼ばれ、より強い仮定である定常（強定常）とは区別されるが、本論文では前者を単に定常と記す。）

内挿に関わる重みのパラメータ λ は、不偏性の制約条件付きの期待二乗誤差最小化から導出されるが、その際、この定常の仮定を用いることで式(8)のような解が導出される。逆に、定常性が満足されない場合に式(9)を用いて内挿を行うと、推定結果にバイアスが生じることとなる。一方、次章で説明する誤差要素モデルでは、この定常性（２次モーメントに関する仮定：式(11)）を仮定しない。そのため、結果として一般にはこれが満足されないこととなる。

3. 誤差要素モデルと空間内挿

(1) 誤差要素モデル

クロス・セクションデータに対して回帰分析などを行う場合、データサンプルの誤差項同士に相関が残ることがある。空間計量経済学や計量地理学の分野においては、時系列分析におけるアプローチのアナロジーから、誤差項の空間的自己相関に対する対処法として次のようなモデルが提案されている [2), 3)]。（誤差修正モデル error correction model *or* 誤差要素モデル error component model : ECM）

(a) 自己回帰モデル：Spatial Auto-Regressive Model

$$\varepsilon=\rho W\varepsilon+u \qquad (12)$$

(b) 移動平均モデル：Spatial Moving-Average Model

$$\varepsilon=\rho Wu+u \qquad (13)$$

(c) Spatial Correlation Model

$$\varepsilon=Wv+u \qquad (14)$$

ここで、u, v は以下の仮定を満たす確率変数であり、ρ はパラメータ、σ_u : u の分散、σ_v : v の分散である。

$$E[u]=0, E[v]=0,$$
$$Cov[u]=\sigma_u^2 I, Cov[v]=\sigma_v^2 I, E[uv']=0 \qquad (15)$$

行列 W は空間重み行列 spatial weight matrix（$I\times I$）または結合行列であり、その各要素 w_{ij} は通常次式のようにゾーン i とゾーン j の間の空間距離 d_{ij} の関数として与える（α, a : 定数）。

$$w_{ij}=1/d_{ij}{}^{\alpha}, \exp(-a\cdot d_{ij}), \cdots, w_{ii}=0 \quad (i\neq j) \qquad (16)$$

パラメータの推定に最尤法を用いる場合には、u, v に対して次のような正規分布が仮定される。

$$u\sim N(0,\sigma_u^2 I), v\sim N(0,\sigma_v^2 I) \qquad (17)$$

時系列相関に対しては、先行条件と現象との間の因果連鎖が時間の進行方向に副っており、ある時点における現象はその時点より過去へは影響を与えないという考えに基づきモデル化がなされる。これに対し、空間的自己相関に対しては、式(12)～(14)に示すような双方向に影響を与える構造としてモデル化されるのが特徴である。このことは、3.(3)で述べるように、Kriging と整合する形で ECM に基づく空間内挿を行う際の重要な論点となる。

(2) 誤差要素モデルと内挿

これまでも、空間計量経済学の分野において提案されたモデルを Kriging に拡張する手法に関する研究はいくつ

が存在する。しかし、著者らがレビューを行った範囲では、3.(1)に示したような空間を分析対象とした誤差要素モデルをもとに、Kriging と整合させる形で空間データの内挿を行った例は無い。

これは、一つには、ECM が主として適用される社会経済の分野では、内挿そのものを必要としない場合が多いという事実に起因するものと考えられる。別の理由としては、次節で述べるように、ECM から導かれる誤差項の分散共分散行列が、結果として Kriging が前提とする定常の仮定を満足しないという問題に起因するものと考えられる。

(3) 誤差要素モデルに基づく Kriging 内挿の理論的検討

本節では、誤差要素モデルの考え方をもとに Kriging を用いて空間内挿を行うことが可能であるか、それが可能でないとするとどのような問題によるものであるかついて考察を行う。

回帰モデルの誤差項に対し、3.(1)に示した(a)〜(c)のようなモデルを導入した場合、その分散共分散行列は次式のとおり与えられる。

$$\text{(a)}\quad Cov(\varepsilon)=\sigma_u{}^2\left[(I-\rho W)'(I-\rho W)\right]^{-1} \qquad (18)$$

$$\text{(b)}\quad Cov(\varepsilon)=\sigma_u{}^2\left[(I+\rho W)'(I+\rho W)\right] \qquad (19)$$

$$\text{(c)}\quad Cov(\varepsilon)=\sigma_v{}^2 W'W+\sigma_u{}^2 I \qquad (20)$$

時系列相関に対しては、先行条件と現象との間の因果連鎖が時間の進行方向に副っており、ある時点における現象はその時点より過去へは影響を与えないという考えに基づきモデル化がなされる。これに対し、空間的自己相関の場合には、(2)で述べたように、空間内において双方向に影響を与える構造としてモデル化されることが一般的であり、時系列とは根本的に異なる側面を持つ。この空間特有の双方向の影響、すなわちフィードバックが、時系列モデルと本質的に異なる結果をもたらす。このことは、誤差項の自己分散について考察すると明確になる。例えば、(b)の MA モデルにおいては、地点 i における自己分散は式(21)に示されているとおり、自分自身の分散の他に他のあらゆる地点を介した影響を受けている（図１）。

$$Var(\varepsilon_i)=\sigma_u{}^2\left[1+\rho^2\sum_{k=1}^{n}w_{ik}w_{kj}\right] \qquad (21)$$

従って、これらの地点がある特殊な配置に分布している場合を除いて、空間を対象とした AR あるいは MA モデルによって導かれる誤差項の分散共分散行列は、意図するか否かに関わらず、$\rho>0$ では結果として不均一分散を生じさせることとなる。一方、任意の異なる２地点 i,j 間における共分散、例えば(b)の MA モデルの誤差項の共分散行列の各非対角要素は次式のように表わされる。

$$Cov(\varepsilon_i,\varepsilon_j)=\sigma_u^2\left[\rho\left(w_{ij}+w_{ji}\right)+\rho^2\sum_{k=1}^{n}w_{ik}w_{kj}\right]\ (i\neq j) \qquad (22)$$

この式が示すとおり、異なる２地点 i,j 間における共分散は、それぞれの地点と別の地点 k'（$\neq i,j$） との相関によって影響されるため、これらの地点がある特殊な配置に分布している場合を除いて、やはり２地点 i, j 間の距離 d_{ij} のみの関数とはならない（図２）。この結果、地点 i から等距離にある地点 j と地点 k との間の共分散が異なるという結果が生じる。

このように、一般に、空間を対象とした ECM によってモデル化された誤差項の共分散は、Kriging で前提とされる定常性を満足しないことが分かる。これらの点は、時系列モデルにはない大きな特徴であると言える。実際、時系列モデルの内挿については、例えば時系列の AR モデルに関して Kriging と整合する内挿手法に関する研究が Arthur(1962)[4]によって行われている。

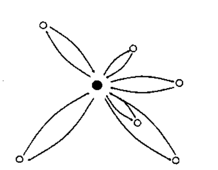

図１. 自己分散に影響を及ぼす空間相互作用

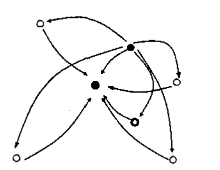

図２. 共分散に影響を及ぼす空間相互作用

●：地点 i　●：地点 j　O：地点 k
○：それ以外の地点
↗↘：モデル化により考慮される影響関係

4. 移動平均モデルに基づく Kriging による空間内挿法

3.で述べたように、ECM に基づき Kriging と整合する空間内挿を行うに際しては、結果的に誤差項の分散共分散行列が定常性の仮定を満足ないという問題が生じる。そこで本章では、ECM のフレームを維持したまま、誤差項の分散共分散が定常性を満足する方法を提示する。

まず、図３に示すように、実際の観測地点を含む地点が、無限に広がる格子点状に配置されており、観測されていない地点はそのうちのデータが欠損している地点であると想定する。このような仮定を置くと、(b)の MA モデルに

ついては、式(21), (22))から容易に確認できるように、分散の不均一性を回避すると同時に共分散も定常となり、Kriging の前提と整合する。

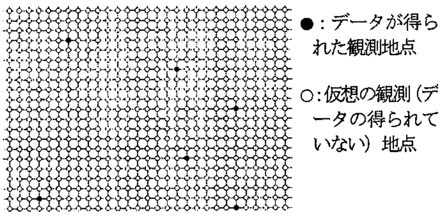

図3．仮想観測地点の設定

あとで述べる理由に加え、実際的にも内挿を行う点の数は有限個で良いため、無限に代えて十分多くの数の格子を想定する。格子点の数を $N\times N$ とすると、対応する空間重み行列結合 $W^{(N)}(N\times N)$は式(23)のように定義される。

$$W^{(N)}=\begin{pmatrix} w_{11} & w_{12} & \cdots & w_{1N} \\ w_{21} & w_{22} & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & w_{NN} \end{pmatrix} \tag{23}$$

ここで、観測データの入手が可能なのは図３において●で表された I 地点のみであり、観測データより導かれる回帰式は以下のとおり I 個となる。

$$y = X\beta + (I + \rho W^{(I)})u \tag{24}$$

ただし、$W^{(I)}$ は $I \times N$ の行列であり、$W^{(N)}$ から、仮想的に設定されたためにデータの得られていない格子点に対応する行を削除した行列である。

このとき、次式の誤差項に関する分散共分散行列は式(26)のとおり与えられる。

$$\varepsilon = (I + \rho W^{(I)})u \tag{25}$$

$$C = \sigma_u^2 (I + \rho W^{(I)})'(I + \rho W^{(I)}) \tag{26}$$

これより、通常の ECM 同様、最尤推定法により各パラメータが推定される。

ところが、無限格子点を想定することにより、別の問題が生じることとなる。何故なら、3.(3)で説明したように、地点 i の自己分散は他の点の影響を受け、また、地点 i と地点 j の共分散も同様にこれら以外の点の影響を受けるため、無限に多くの点を考慮することによって発散するからである。そこで range を設定する等の制約が必要となる。この際、各点に関しての自己分散の計算において考慮する地点数を同一にしなければ不均一分散をもたらすため、ある一定の範囲、具体的には半径 R の円内に含まれる地点のみを考慮することとする。すなわち、重み w_{ij} に対し、式(27)のような制約を置く。

$$w_{ij} = \begin{cases} f(d_{ij}) & (0 \le d_{ij} \le R) \\ 0 & (R < d_{ij}) \end{cases} \tag{27}$$

このような空間的な制約は、Kriging において range と呼ばれるものに相当するものである。同様に、共分散についても range を設定し、項が発散しないような構造とする。

このようにして、covariogram を定義することにより、空間を対象とした移動平均モデルに基づき、Kriging の基本的な前提と整合する内挿が可能となる。

5. おわりに

本論文では、空間計量経済学においてよく用いられてきた ECM に、空間統計学における covariogram の概念を取り込むことにより、ECM の概念をもとに空間内挿が行えることを、移動平均（MA）モデルについて示した。

今後の研究課題としては、実際のデータを用いて本研究で提案した手法と既存の Kriging を用いた内挿を行い、実証比較を行うことがあげられる。これについては、現在、降雨のデータを検討に着手したばかりであり、別の機会に改めて結果を発表したい。また、本研究では移動平均モデルについてのみ、拡張の方向性を提示した。自己回帰（AR）モデルでは、Kriging と整合する形でのモデルの拡張が不可能であると予想しているが、Spatial Correlation Model に関しては、まだ十分な検討を行っていない。これらについても、今後の課題としたい。

参考文献

1) Bailey, T. C. and Gatrell, A. C. (1995): *Interactive Spatial Data Analysis*, Longman.
2) Anselin, L. (1988): *Spatial Econometrics: Methods and Models*, Kluwer Academic.
3) 堤盛人・清水英範・福本潤也・井出裕史 (1998)：誤差項に空間的自己相関が存在する回帰モデルのパラメータ推定手法に関する考察, 土木計画学研究・論文集, No.15, pp.49-56.
4) Arthur, S.G. (1962): Best Linear Unbiased Prediction in the Generalized Linear Regression Model, *Journal of American Statistical Association*, Vol.57, pp.369-375.