クリギング

三浦　英俊

クリギングは空間データを対象として任意の地点の値を予測する地理統計手法である．対象データに空間的な相関関係があり，それをコバリオグラムと呼ばれる共分散関数で適切に記述することができれば，予測値を容易に計算することができる．地質分野，環境科学，あるいは社会科学分野で広く適用されている．

1. はじめに

クリギングは空間データを対象として任意の地点の値を予測する地理統計手法の1つである．予測値は既知の観測データ値の加重平均によって表現され，値が未知である地点の空間補間を形成することができる．加重平均を計算する際の観測データ別の重みは，対象とする空間データが2次定常性（または本質的定常性）と呼ばれる空間的相関関係を有すると仮定したうえで，未知データの平均二乗予測誤差を最小にするものを求める．

クリギングはもともと鉱床の空間分布を予測するために開発された．現在は地質分野のみならず環境科学，水産学，社会科学分野など空間データを取り扱う分野で広く利用されている．

2. 定式化

クリギングを適用するには，予測対象となる地理的な事象に2次定常性を有する確率場を仮定する．すなわち地理的事象が地域全体の傾向，空間的な相関を持つ残差，およびランダムノイズという3つの成分によって構成されているとする．

予測対象値を含む空間を $D \subset \mathbb{R}^d$，$\mathbf{s} \in D$ を空間上の任意の地点，D 上の予測対象値を確率変数 $Z(\mathbf{s})$ とする．$Z(\mathbf{s})$ に次のような構造を仮定する．

$$Z(\mathbf{s}) = \mu + \delta(\mathbf{s})$$

μ は地域全体の傾向を表す平均，$\delta(\mathbf{s})$ は空間的な相関を持つ残差とランダムノイズの和である．

予測に使われる観測データの観測位置を $\mathbf{s}_1, \dots, \mathbf{s}_n \in D$ として，観測データ値 $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ はこれらの地点における実現値と見なす．$Z(\mathbf{s})$ が2次定常性を持つとは，任意の $\mathbf{s}, \mathbf{s}_i, \mathbf{s}_j \in D$ で以下の関係が成り立つことである．

$$\mathbf{E}\{Z(\mathbf{s})\} = \mu,$$
$$\mathbf{Cov}\{Z(\mathbf{s}_i), Z(\mathbf{s}_j)\} = C(\mathbf{s}_i - \mathbf{s}_j).$$

ここで $\mathbf{E}(\cdot)$ は期待値，μ は定数，$\mathbf{Cov}(\cdot)$ は共分散を表す．関数 $C(\cdot)$ をコバリオグラムと呼ぶ．これらの仮定は，$Z(\mathbf{s})$ の空間的な相関を2点間の距離によって記述できるとするものである．任意の地点 $\mathbf{s}_0 \in D$ における未知値 $Z(\mathbf{s}_0)$ の予測値は，クリギングによって加重平均

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{n} \lambda_i Z(\mathbf{s}_i) \tag{1}$$

と書ける．λ_i は重みパラメータである．ただし負値となることもある．予測値が不偏であるためには $\sum \lambda_i = 1$ でなければならない．

クリギングとは，観測データ $Z(\mathbf{s}_i)$ に対して適切なコバリオグラム関数 C を推定し，任意の地点 $\mathbf{s}_0$ の予測値 $\hat{Z}(\mathbf{s}_0)$ における重み λ_i を求めることに他ならない．

コバリオグラムには，対象とする数値が「距離が近いほどよく似た値となる」ことを記述する役割がある．ベクトル $\mathbf{h}$ を $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ としたとき，距離 $\|\mathbf{h}\|$ が大きいほどコバリオグラム $C(\mathbf{h})$ の値すなわち共分散が大きくなることが期待される．コバリオグラム関数として，球形モデル，指数型モデルなどいくつかのモデルが提案されている．また，$\gamma(\mathbf{h})$ を

$$\gamma(\mathbf{h}) = C(\mathbf{0}) - C(\mathbf{h})$$

とすると，$\gamma(\mathbf{h})$ はセミバリオグラムと呼ばれる．コバリオグラムの代わりにセミバリオグラムを用いて空間における $Z(\mathbf{s})$ のばらつきを記述することも可能である．

みうら　ひでとし
明海大学 不動産学部
〒279-8550 浦安市明海 1-2

クリギングを実行するためには，コバリオグラム（あるいはセミバリオグラム）関数を具体的に推定しなければならない．これには経験セミバリオグラムを用いた方法[1]や最尤推定法を用いる方法などがある[4]．

クリギングは，対象値の空間的な性質や目的とする予測値の違いによっていくつかの種類がある．$Z(\mathbf{s})$の期待値が対象空間において一定であると仮定できる場合は「通常クリギング」を適用することができる．対象値の期待値に空間的な一定の傾向がある場合（例えば気温データを対象としたとき，南に行くほど温度が高い場合など）に使用する普遍クリギング，地点でなく空間範囲内の対象値の積分量を予測するブロッククリギング，複数の観測データを利用する共クリギング，などがある．

予測値 $\hat{Z}(\mathbf{s}_0)$ を得るための重み λ_i は，次の平均二乗予測誤差 $\sigma^2(\mathbf{s}_0)$ を最小とする最適化問題を解くことによって得られる．

$$\min_{\lambda_i} \quad \sigma^2(\mathbf{s}_0)=\mathbf{E}(\hat{Z}(\mathbf{s}_0)-Z(\mathbf{s}_0))^2$$
$$=C(\mathbf{0})+\sum_{i,j}\lambda_i\lambda_j C(\mathbf{s}_i-\mathbf{s}_j)-2\sum_i \lambda_i C(\mathbf{s}_0-\mathbf{s}_i),$$

subject to

$$\sum_i \lambda_i=1.$$

上の最適化問題は例えばラグランジュ乗数法を用いて解くことができる．$Z(\mathbf{s})$ の期待値が対象空間において一定であるとしたとき，$\boldsymbol{\lambda}=(\lambda_1, \ldots, \lambda_n)'$, $\mathbf{c}=(C(\mathbf{s}_0-\mathbf{s}_1), \ldots, C(\mathbf{s}_0-\mathbf{s}_n))'$，および共分散行列 $\Sigma=\{C(\mathbf{s}_i-\mathbf{s}_j)\}_{ij}$ とすると，最適解は

$$\boldsymbol{\lambda}'=\left(\mathbf{c}+\mathbf{1}\frac{(1-\mathbf{1}'\Sigma^{-1}\mathbf{c})}{\mathbf{1}'\Sigma^{-1}\mathbf{1}}\right)'\Sigma^{-1}$$

となる．これは通常クリギングの場合の重みである．通常クリギングやブロッククリギングなど，ほかのクリギングでは $Z(\mathbf{s})$ に関する仮定や解くべき問題がそれぞれ異なり，したがって得られる $\boldsymbol{\lambda}$ も異なる．

3. 計算例

図1はある市の住宅地の地価公示の調査地点分布図である．観測データとして20箇所(n=20)の前年比増加率を使用し，3地点A, B, C（それぞれ $\mathbf{s}_0$ に相当する）の前年比増加率が未知であるとして推定を行う．

表1は観測データ $Z(s_i)$ と，クリギングによって得られた λ_i を示す．これらを(1)に代入してクリギング推定値 $\hat{Z}(s_0)$ を求める．推定地点に近い観測データほど λ_i の値が大きく，合計は1となることが観察できる．最下段に正解として実データを示す．Bの推定値は最も悪い．Bの正解実データは−3.3%であるが，

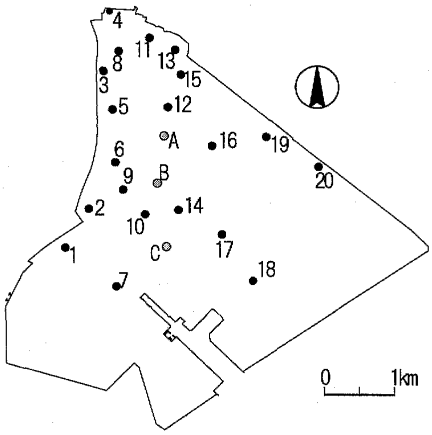

図1 観測データ 1, …, 20 と増加率推定地点 A, B, C

表1 観測データと推定地点別の重み（それぞれの λ_i のうち上位3つを**太字**で表示した）

i	前年比増加率 $Z(s_i)$[%]	A	B (λ_i)	C
1	−7.0	−0.01	−0.01	0.00
2	−8.1	−0.03	−0.06	0.02
3	−6.4	−0.02	−0.01	0.00
4	−4.0	0.00	0.00	0.00
5	−8.8	0.10	0.02	0.00
6	−7.1	**0.16**	0.14	−0.02
7	−7.0	−0.01	−0.02	**0.24**
8	−8.2	0.00	0.00	0.00
9	−7.9	0.06	**0.26**	−0.04
10	−6.8	0.02	**0.20**	**0.33**
11	−7.9	0.00	0.00	0.00
12	−6.1	**0.49**	0.08	−0.01
13	−8.1	−0.01	0.00	0.00
14	−6.9	0.08	**0.32**	**0.27**
15	−8.3	−0.02	−0.01	0.00
16	−7.5	**0.23**	0.13	−0.03
17	−8.8	−0.01	−0.01	0.19
18	−9.3	0.00	−0.01	0.07
19	−5.5	−0.02	−0.01	−0.01
20	−6.0	0.00	0.00	−0.01
合計		1.00	1.00	1.00
	推定値 $\hat{Z}(s_0)$	−6.8	−7.4	−7.1
	実データ	−5.6	−3.3	−7.0

20個の観測データはすべて−3.3%より小さいため，推定値が実データと大きく異なるのはやむを得ない．

参考文献

[1] 間瀬茂, 武田純,『空間データモデリング—空間統計学の応用』, 第6章, 共立出版, 2001.

[2] 張長平,『地理情報システムを用いた空間データ分析』, 140-144, 古今書院, 2001.

[3] Wackernagel, H. 原著,『地球統計学』, 森北出版, 2003.

[4] Cressie, N. A. C., *Statistics for Spatial Data Revised Edition*, Wiley-Interscience, 1993.

[5] Chiles, J.-P. and Delfiner, P., *Geostatistics : Modeling Spatial Uncertainty*, Wiley-Interscience, 1999.