

カーネル推定とクリギングの関係－理論的考察－

丸山 祐造¹, 岡部 篤行^{1,2}

¹ 東京大学 空間情報科学研究センター, ² 青山学院大学 総合文化政策学部

連絡先: <maruyama@csis.u-tokyo.ac.jp>

(1) **動機:** 空間予測手法として知られているクリギングと、回帰分析あるいはそれを拡張したカーネル回帰分析の類似点、相違点について理論的な観点から考察する。これらは主に空間データ解析を行う研究者や実務家から興味を持たれていたものの、少なくとも日本の統計コミュニティ及び GIS コミュニティにおいては、理論的な観点からきちんと整理されていなかった。

(2) **特徴と結果:** 両者の違いを理解するために、まず統計学の「予測」と「推定」という概念を説明する。これらはそれぞれ「(ランダムな)確率変数を言い当てる」、「(ランダムでない)定数の未知パラメータを言い当てる」ことである。パラメータとは平均や分散など分布の特徴を表すものである。クリギングにおいては、誤差項に空間相関を持つ重回帰線形モデルから、 n 地点において観測値を得ているときに未観測地点での同じ重回帰モデルからの(確率的な)目的変数を予測することが目的となる。予測という設定により、観測地点間の空間相関だけでなく、未観測地点と観測地点の空間相関をも反映した関数である予測量が構成される。一方カーネル回帰

は、目的変数の期待値が非線形関数であることを前提として、地点間の近さを考慮した観測値の重みつき平均によって、その期待値(非線形関数)を推定する。クリギングよりも簡便な手続きで未観測地点における目的変数の出方を言い当てる事が出来る事が利点である。しかし統計学的な枠組みや意味づけはやや弱く、特に未観測地点と観測地点の関係に着目して分析しようとしているにも関わらず、既に手元にある観測地点の観測値がどのような関係性のもとで生成されたかを考慮しないモデリングである点が弱点と考えられる。

(3) **意義:** 今回の成果によって、類似点及び相違点を含む両者の性質の全貌を見渡せるわけではない。しかし、1) 数理統計学的な視点からは問題設定に違いがあることを指摘し、2) 適切な線形変換を施すことによって得られるクリギング予測量とカーネル推定量の係数を比較することによりいくつかの考察を行った。これらの結果は理論的背景に興味を持つ GIS 関係者、実務家が、問題設定に応じて適切な手法を選ぶために非常に有益であると考えられる。

- ▶ 空間相関がある誤差項を持つクリギングモデル $Y(u_i) = \beta_0 + \epsilon(u_i)$
- ▶ 未観測地点 u_0 での $Y(u_0)$ の予測量 (ベスト線形予測量)

$$\left\{ \frac{1 - \tilde{\mathbf{r}}'_{Y0} \tilde{\mathbf{1}}_n}{\tilde{\mathbf{1}}'_n \tilde{\mathbf{1}}_n} \tilde{\mathbf{1}}'_n + \tilde{\mathbf{r}}'_{Y0} \right\} \tilde{\mathbf{y}}$$

- ▶ u_0 と相関が高い地点の観測値に重みをおいた重みつき平均
- ▶ カーネル推定 $\sum_{i=1}^n K_\lambda(u_0, u_i) y_i = \mathbf{k}(u_0, \mathbf{u})' \mathbf{y}$
- ▶ 係数ベクトルの成分は、地点間の距離が近いほど大
- ▶ 両者は実質的に同等である。つまりあるクリギング予測量を作るカーネルが存在するし、またあるカーネル推定量に対応する相関構造が存在する
- ▶ 記号の定義など、詳細はポスターで説明